FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

POPH
QUALIFICAR É CRESCER.

QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL
PORTUGAL 2007.2013

Governo da República Portuguesa

UNIÃO EUROPEIA
Fundo Social Europeu

Cofinanciado por:

COMPETE 2020

PORTUGAL 2020

UNIÃO EUROPEIA
Fundo Social Europeu

Deciphering the genetic basis of intellectual disability through unbiased genomic approaches
Fátima Daniela Teixeira Lopes
UMinho | 2017

Universidade do Minho
Escola de Medicina

Fátima Daniela Teixeira Lopes

Deciphering the genetic basis of intellectual disability through unbiased genomic approaches

maio de 2017

Universidade do Minho
Escola de Medicina

Fátima Daniela Teixeira Lopes

Deciphering the genetic basis of intellectual disability through unbiased genomic approaches

Tese de Doutoramento em Ciências da Saúde

Trabalho efetuado sob a orientação da
Professora Doutora Patrícia Espinheira de Sá Maciel
e do
Professor Doutor Jacques Michaud

maio de 2017

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração da presente tese. Confirmo que em todo o trabalho conducente à sua elaboração não recorri à prática de plágio ou qualquer forma de falsificação de resultados.

Mais declaro que tomei conhecimento integral do Código de Conduta Ética da Universidade do Minho.

Universidade do Minho, 05 / 05 / 2017

Nome: Fátima Daniela Teixeira Lopes

Assinatura: Fátima Daniela Teixeira Lopes

This work was supported by Foundation for Science and Technology (FCT) through a PhD studentship (SFRH/BD/90167/2012) and by the Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 262055.

This work was also supported by the FEDER through the Operational Programme Competitiveness Factors - COMPETE and the national funds through the FCT - Foundation for Science and Technology within the projects (POCI-01-0145-FEDER-007038), and by the project NORTE01-0145-FEDER-000013, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).

Aos meus pais e ao Albino

Agradecimentos/Acknowledgments

Este trabalho, assim como todo o percurso que a ele levou, não seria possível concretizar em solidão. Existem muitas pessoas a quem devo agradecimentos.

À minha orientadora, Professora Doutora Patricia Maciel, pela confiança que depositou em mim desde o meu estágio, pelo incentivo em seguir para doutoramento, pelas discussões quando estava sem conseguir andar para a frente e por nunca me deixar ficar sem uma correção de algo importante em todos estes anos! Muito obrigada por tudo!

To my supervisor, Professor Jacques Michaud, for welcoming me in his laboratory and giving me the opportunity to work fully integrated in the exome team.

À Fundação para a Ciência e Tecnologia pelo financiamento que possibilitou a realização deste trabalho, através da concessão da minha bolsa de doutoramento e do financiamento dos projectos de investigação.

À Escola de Medicina e ao Instituto de Investigação em Ciências da Saúde, em particular à professora Doutora Cecília Leão e ao professor Doutor Jorge Pedrosa pela oportunidade de realizar este trabalho na Escola de Medicina/ICVS.

Ao domínio de Neurociências por todas as reuniões e palestras científicas oferecidas aos alunos.

To the people from Michaud lab at CUHSJ and from the Exome meeting at CUHSJ for the warm welcoming in the group and all the discussions on Thursdays. In particular to Fadi and Julie for all that they teach me and help. Julie, thank you for all of our discussions before the meeting and for always letting me double check with you!

A todos os meus colegas do grupo grupo Maciel (meninas e meninos) pela ajuda que sempre me prestaram (que para a Sara, Andreias e Anabela já dura à muitos anos!). Muito obrigada por todas as discussões das minhas dúvidas, dos conselhos quando não sabia como havia de resolver algo, e pelo *input* nas reuniões. Muito obrigada a todos!

Aos médicos do Centro de Genética Médica Jacinto de Magalhães pela colaboração no projecto da genética do atraso mental, em particular à Drª Gabriela Soares pela sua sempre presente resposta e disponibilidade.

Às minhas colegas de grupo "da genética": Mafalda e Fátima Torres.

Mafalda, muito obrigada por todas as discussões do projecto do atraso e pelo teu trabalho no artigo do Rett-like.

Fatima Torres muito obrigada por toda a tua ajuda no artigo dos arrays, nas colheitas e nas experiências (principalmente nesta fase final pois não teria conseguido sem ti!). Gosto de saber que temos ainda muito trabalho juntas pela frente pois é um gosto trabalhar contigo!
As minhas colegas do piso 1 (Eduarda, Patricia, Ana, Mónica e Claúdia) por todas as vezes que me "distrairam" com as vossas brincadeiras. Obrigada por toda a ajuda e, acima de tudo, pelas gargalhadas!

Em especial:
À Diana, Raquel e Lígia pela amizade de já tantos anos, pelas conversas e pelo apoio quando estive longe de casa (Diana obrigada por me deixares ir chatear-te na cantina da UM com as minhas amarguras!). À Melinda, Silvia e Tânia que mesmo longe conseguem fazer sentir que estão sempre perto. À Filipa por todas as conversas existenciais ao longo destes anos e pelo teu apoio quando estive longe. À Sónia, minha companheira de viagem (e de papéis, e de requerimentos, e de formulários, e de deadlines, e tudo e tudo!), mas acima de tudo pela amizade ao longo destes 4 anos (que espero se mantenha para muitos mais). To Carla and Marguerite, my friends *made in* Montreal! Marguerite, thank you for all the work that we did together, for your friendship, for the conversations, for the dinners, for the beers after work. Thank you for all traveling around Quebec and for your friendship! Carla, obrigada por todos os almoços e cafés juntas, por todas as conversas, por todos os conselhos e por me teres feito sentir que tinha uma amiga em Montreal a quem podia pedir ajuda em Português! Obrigada pela tua amizade e apoio!

Aos meus pais sem os quais nada seria. Por sempre me terem apoiado em tudo e por me fazerem sentir que nunca estarei sozinha.

Ao Albino, meu amor, meu companheiro de vida nos bons e nos maus momento. Sem ti nada teria o mesmo sentido, nada teria o mesmo gosto. Obrigada por todo o teu apoio e ajuda, não só durante este trabalho, mas em tudo na minha vida!

Abstract

Neurodevelopmental disorders arise in childhood and are life-long condition that represents many challenges to the patients, their families and society, through public services. Among those, intellectual disability affects 1% of the population in developed countries, encompassing the most common group of neurodevelopmental disorders. Intellectual disability is characterized by cognitive impairment and limitation in functioning capacity and can be cause by exogenous factors (such as maternal alcohol abuse during pregnancy, infections and malnutrition) but it is well established that genetic factors play important roles in its pathophysiology. In ID, as in many others disorders, detecting the underlying genetic cause is a complex and time consuming process yet of great value to the patients and families because it allows the possibility of genetic counseling.

In the last years, advances in two major types of technologies allowed great advances in the discovery of new genomic anomalies causing intellectual disability: array comparative genomic hybridization and massive parallel sequencing. Array comparative genomic hybridization allowed a high resolution genome wide investigation of copy number variations leading to the discovery of many novel microdeletion/microduplication syndromes. Massive parallel sequencing evolved in a way that the sequencing of all the genome (or, at a lower cost, all the exons) is now possible to perform in any genetic laboratory in a time and cost-efficient manner, allowing the discovery of many novel variants in previously known and newly discovered intellectual disability genes. These two approaches have provided significantly new insights into the biological pathways associated with intellectual disability and tremendously improved the diagnostic process.

In this study we applied these two technologies to the study of the genetic basis of neurodevelopmental disorders. We studied a big group of patients with idiopathic intellectual disability by aCGH (which included two cohorts with different selection criteria - a research and a clinical cohort), a group of patients with a Rett syndrome-like clinical presentation by exome sequencing and re-analyzed exome data from a pediatric heterogeneous cohort.

Array comparative genomic hybridization allowed the detection of previously known microdeletion and microduplication syndromes in patients with until then unexplained intellectual disability, with yields of 13% in the research cohort and 18% in the clinical cohort. Importantly it also allowed the discovery of 12 new *loci* likely to cause neurodevelopmental disease as well as the gathering of additional patients with overlapping genomic imbalances and phenotypic features, allowing the definition of new (rare) syndromes.

Massive parallel sequencing – more specifically whole exome sequencing - was applied to a group of patients sharing similar clinical presentation (Rett syndrome-like) and proved to be very effective, leading to the identification of five new genes possibly involved in intellectual disability (*HTT*, *SMARCA1*, *GABBR2*, *RHOBTB2* and *EIF4G1*).

It is currently an accepted fact however, that the data generated by exome sequencing at a certain point in time, may not retrieve the genetic cause of the disease. This limitation is often related with the lack of information regarding the genes detected. Given the always increasing knowledge on genes and pathways involved in neurodevelopmental disorders, the need for reevaluation of older and previously unsolved cases emerges. This strategy was also applied in this work and proven to be extremely useful in the clinical context, adding new patients to help establish the relevance of candidate disease genes and raising new candidate genes (*DNAJC21*, *MYOD1* and *PAX7*).

In summary, this work helped to clarify the genetic basis of disease in several patients until then unsolved, as well as to bring forward new candidate loci and genes for intellectual disability and other neurodevelopmental disorders.

Resumo

As perturbações do neurodesenvolvimento surgem na infância e constituem doenças crónicas, criando inúmeras limitações para os doentes, famílias e sociedade – sob a forma de serviços públicos. Entre estas, o défice intelectual (previamente designado atraso mental) afecta 1% da população dos países desenvolvidos, sendo o tipo de doença do neurodesenvolvimento mais comum. O défice intelectual é caracterizado por uma limitação cognitiva e funcional e pode ser causado por factores exógenos (como o consumo materno de álcool durante a gravidez, infeções e malnutrição), sendo no entanto fatores genéticos reconhecidos como muito importantes para a patofisiologia do défice cognitivo. Neste grupo de doenças, assim como em muitas outras, a deteção da causa genética é um processo complexo e demorado mas de grande valor para os doentes e famílias, uma vez que abre portas á possibilidade de aconselhamento genético.

Nos últimos anos, avanços em duas grandes tecnologias de diagnóstico genético permitiram a descoberta de novas anomalias genéticas associadas e défice intelectual: array de hibridação genómica comparativa e sequenciação paralela massiva. Os arrays de hibridação genómica comparativa permitiram a análise de alta resolução de todo o genoma na busca de alterações do número de cópias, o que resultou na descoberta de várias novas síndromes associadas microdeleções/microduplicações. A sequenciação paralela massiva desenvolveu-se de uma forma em que a sequenciação de todo o genoma (ou, de forma mais económica, de todos os exões) é agora possível realizar em qualquer laboratório de genética em tempo útil e com um bom custo/benefício, permitindo a descoberta de variantes tanto em genes já conhecidos como em novos genes causadores de défice intelectual. Estas duas abordagens contribuíram significativamente para novas descobertas em vias moleculares associadas com défice intelectual e para o melhoramento do seu diagnóstico.

Neste estudo aplicamos estas duas tecnologias ao estudo da base genética de doenças do neurodesenvolvimento. Estudamos por aCGH um grande grupo de doentes com défice intelectual idiopático (o que incluiu dois coortes com diferentes critérios de seleção – um coorte de investigação e um coorte clinico), estudamos por sequenciação de exoma um grupo de doentes com sintomatologia semelhante à síndrome de Rett, e reanalisamos dados de exoma de um grupo heterogéneo de doentes pediátricos.

O array de hibridação genómica comparativa permitiu a deteção de microdeleções e microduplicações já conhecidas em doentes até à data com défice intelectual idiopático, com uma taxa de sucesso de 13% no coorte de investigação e de 18% no coorte clinico. Foi também possível a deteção de novos 12

novos *loci* passíveis de causar doença do neurodesenvolvimento assim como a recolha de doentes adicionais com desequilíbrios genómico e fenótipo sobreponíveis, contribuindo para a definição de novas síndromes raras.

A sequenciação paralela massiva – em particular a sequenciação do exoma - foi aplicada a um grupo de doentes com apresentação clinica semelhante (síndrome de Rett-like), revelando-se bem-sucedida na identificação de cinco novos genes possivelmente envolvidos no défice intelectual (*HTT*, *SMARCA1*, *GABBR2*, *RHOBTB2* e *EIF4G1*).

Atualmente é facto aceite que os dados gerados por sequenciação do exoma numa determinada altura poderão não levar á descoberta da causa genética da doença. Esta limitação está muitas vezes relacionada com a escassez de informação relativamente aos genes encontrados. Tendo em conta o sempre crescente conhecimento relativo a genes e vias moleculares envolvidas em doenças do neurodesenvolvimento, surge a necessidade de reavaliação de casos antigos não solucionados. Esta estratégia foi também aplicada neste trabalho e provando ser de grande utilidade no contexto clinico, levando à deteção de mais doentes que contribuem para a determinação dos genes candidatos relevantes assim como para a deteção de novos genes candidatos (*DNAJC21*, *MYOD1* and *PAX7*).

Em resumo, este trabalho contribuiu para a clarificação da causa genética de doença em vários doentes até à data não resolvidos, e propõe novos *loci* candidatos e genes que contribuem para o défice intelectual e outras doenças do neurodesenvolvimento.

Table of contents

Abbreviations

A:

aCGH – array comparative genomic hybridization

ACC - Agenesis of the corpus callosum

ACMG – American College of Medical Genetics

ADHD – Attention deficit hyperactive disorder

ADI-R – Autism Diagnostic Interview-Revised

ADOS - Autism Diagnostic Observation Schedule

ARFs – ADP Ribosylation Factors

ASD – Autism spectrum disorder

B:

bp – Base pair

BRPS – Bainbridge-Ropers syndrome

C:

CC – Clinical cohort

cDNA – Complementary deoxyribonucleic acid

CNPD – Comissão Nacional de Protecção de Dados

CNV – Copy number variation

CNS – Central nervous system

D:

DD – Developmental disabilities

DNA – Deoxyribonucleic acid

DQ – Development quotient

DGV - Toronto Database of Genomic Variants

E:

EBV - Epstein-Barr virus

EEG - Eletroencefalogram

ExAC - Exome Aggregation Consortium

F:

FISH – Fluorescent in situ hybridization

FSIQ – Full scale intelligence quotient

G:

GMDS - Griffiths Mental Developmental Scales

I:

ID – Intellectual disability

IQ – Intelligence quotient

iPSCs – Induced pluripotent stem cells

K:

Kb – Kilobases

KS - Kleefstra syndrome

M:

Mb – Megabase

MAC – macrocephaly

MAP - Microtubule associated protein

MIC – Microcephaly

MLPA – Multiplex ligation-dependent probe amplification

MPS – Massive parallel sequencing

mRNA – Messenger ribonucleic acid

MRI – Magnetic resonance imaging

N:

NDD – Neurodevelopmental disorder

NGS – Next generation sequencing

NKH – Non-ketotic hyperglycinemia

O:

OFC – Frontal-occipital circumference

P:

PCR – Polymerase chain reaction

PMS – Phelan-McDermid syndrome

Q:

QGD – Quotient global development

qPCR – Quantitative polymerase chain reaction

qRT-PCR – Real time quantitative polymerase chain reaction

R:

RC – Research cohort

RNA – Ribonucleic acid

RTT – Rett syndrome

RT – Reverse transcriptase

S:

SCF – Skp, Cullin, F-box containing complex

SD – Standard deviation

SNARE – soluble NSF attachment receptor

SNP – Single nucleotide polymorphism

SZR - Seizures

V:

VOUS – Variant of unknown significance

W:

WBS – Williams-Beuren Syndrome

WES – Whole exome sequencing

X:

XLID – X-linked intellectual disability

Thesis Outline

This dissertation includes both published and unpublished data and is divided in four chapters. The work presented relates to the identification of genetic anomalies leading to disease in patients with neurodevelopmental disorders.

Chapter 1 is the general introduction to the theme of this dissertation. This chapter provides an overview of (I) neurodevelopmental disorders and related pathways, (II) copy number variations and their contribution for human diseases and (III) single nucleotide variations in ID. A brief description of array comparative genomic hybridization (aCGH) and massive parallel sequencing (MPS) techniques and their contribution for the identification of new disease–associated genes is also provided.

Chapter 2 describes the results of an exploratory project in which patients with idiopathic ID were studied by aCGH, and is composed of five sub-chapters.

Sub-chapter 2.1 regards the work performed in a large cohort of Portuguese ID patients studied by aCGH. A clinical description of the cohort and a presentation of the main CNVs found are provided. We also present a detailed discussion of the genes possibly contributing for the patients' phenotype. The work presented here is in preparation for publication.

Sub-chapter 2.2 presents a small report of patients carrying CNVs at the 1q43-q44 region. In here we compare the clinical features and genomic imbalances of five patients and explore the contribution of copy number losses and gains in *AKT3* for occipital-frontal circumference alterations. The work presented here is in preparation for publication.

Sub-chapter 2.3 reports clinical and genetic findings in two patients with a microdeletion of 2p13.2 leading to haploinsufficiency of two genes involved in exocytosis/Notch signaling and retinoic acid metabolism. This work is published in the *Orphanet Journal of Rare Diseases* (Wen J*, Lopes F*, *et al*. *Phenotypic and functional consequences of haploinsufficiency of genes from exocyst and retinoic acid pathway due to a recurrent microdeletion of 2p13.2*. Orphanet J Rare Dis. 2013 Jul 10;8:100. doi: 10.1186/1750-1172-8-100 *both authors contributed equally to this work).

Sub-chapter 2.4 reports the clinical and genetic features of seven patients with ID, dysmorphisms and behavioral anomalies that carry CNVs at 7q33 citoband. A review of the available data for 7q33 genomic imbalances is provided, as well an attempt to determine the key common genes among all the cases. The work presented here is in preparation for publication.

Sub-chapter 2.5 presents a previously published case report of a Rett-syndrome *like* patient with a pericentric X-chromosome inversion that leads to altered methylation of the *MECP2* gene and to expression changes, in the absence of structural genomic dosage changes in this gene. This work is published in the International Journal of Developmental Neuroscience (Vieira JP*, Lopes F*, *et al*. *Variant Rett syndrome in a girl with a pericentric X-chromosome inversion leading to epigenetic changes and overexpression of the MECP2 gene*. Int J Dev Neurosci. 2015 Nov;46:82-7. doi: 10.1016/j.ijdevneu.2015.07.010. *both authors contributed equally to this work).

Chapter 3 includes the results of the application of massive parallel sequencing techniques (particularly whole exome sequencing) to two groups of patients using different approaches and with different purposes.

Sub-chapter 3.1 presents the results of trio exome sequencing in a group of 19 patients with a Rett-syndrome like clinical presentation. Several previously known disease-associated genes and new candidates for Rett-like clinical presentations have been identified in this work and are discussed. This work is published in the Journal of Medical Genetics (Lopes F*, Barbosa M*, *et al*. *Identification of novel genetic causes of Rett syndrome-like phenotypes.* J Med Genet. 2016 Mar;53(3):190-9. doi: 10.1136/jmedgenet-2015-103568. *both authors contributed equally to this work).

Sub-chapter 3.2 focused on the re-analyses of 64 patients previously studied in a clinical whole exome sequencing context and reinforces the importance of revisiting unsolved cases. For a large percentage of the cases was possible to find new candidate genes; the factors contributing for these findings are discussed in detail in this sub-chapter. The work presented here is in preparation for publication.

Chapter 4 is the general discussion of the dissertation, in which an integrated view of the findings is provided and in which the main conclusions and future perspectives are discussed.

Disclaimers

Sub-chapter 2.1

Genomic imbalances in known and novel candidate loci in a group of Portuguese patients with idiopathic intellectual disability

Fátima Lopes*, Fátima Torres*, Gabriela Soares, Mafalda Barbosa, João Silva, Frederico Duque, Miguel Rocha, Joaquim de Sá, Guiomar Oliveira, Maria João Sá, Teresa Temudo, Susana Sousa, Carla Marques, Ana Sofia Lopes, Catarina Gomes, Gisela Barros, Felisbela Rocha, Arminda Jorge, Cecília Martins, Sandra Mesquita, Susana Loureiro, Elisa Maria Cardoso, Maria José Cálix, Andreia Dias, Cristina Martins, Céu R. Mota, Diana Antunes, Juliette Dupont, Sara Figueiredo, Sónia Figueiroa, Susana Gama Sousa, Sara Cruz, Adriana Sampaio, Erik A Sistermans, Bauke Yklstra, Paula Rendeiro, Purificação Tavares, Margarida Reis-Lima, Jorge Pinto-Basto, Ana Maria Fortuna, Patrícia Maciel

In preparation; *both authors contributed equally for this work

The results presented throughout this sub-chapter are integrated in a larger publication featuring the comparison of two Portuguese cohorts of patients with ID studied by aCGH: a research cohort (RC) and a clinical cohort (CC). These results will be published in an international peer reviewed research article. The author of this thesis contributed for all the laboratory work, genomic data acquisition, analysis and interpretation of results for the RC, as well as for the manuscript preparation and discussion of all the cases. For the CC the laboratory work and genomic data acquisition was performed at Centro de Genética Clínica with particular contribution of Torres F (co-forst author). Lopes F also participated in the interpretation of results for the CC.

Sub-chapter 2.2

The effect of CNVs at 1q43-q44 in neurodevelopment phenotypes and head circumference alterations

Fátima Lopes, Fátima Torres, Gabriela Soares, Clara D. van Karnebeek, Cecília Martins, Diana Antunes, João Silva, Luís Filipe Botelho, Susana Sousa, Paula Rendeiro, Purificação Tavares, Hilde Van Esch, Evica Rajcan-Separovic, Patricia Maciel

In preparation

The results presented in this chapter refer to the collection of five patients from different laboratories. This work is integrated in a future publication being prepared by the author of this dissertation and she declares that she was involved in all the work presented in the publication except:

- Patients' clinical data collection: for all the cases this was performed by the referring physician;
- Molecular studies: aCGH for patient for patient 2, 3 and 4 was not performed by Lopes F, array painting and mRNA expression studies for patient 5 was not performed by Lopes F.

Sub-chapter 2.3

Phenotypic and functional consequences of haploinsufficiency of genes from exocyst and retinoic acid pathway due to a recurrent microdeletion of 2p13.2

Wen J*, Lopes F*, *et al*. *Phenotypic and functional consequences of haploinsufficiency of genes from exocyst and retinoic acid pathway due to a recurrent microdeletion of 2p13.2*. Orphanet J Rare Dis. 2013 Jul 10;8:100. doi: 10.1186/1750-1172-8-100. *authors contributed equally

The results presented in this chapter refer to publication of two patients from different laboratories. The author of this dissertation declares that she was involved in all the work presented in the publication except:

- Patients' clinical data collection: for both the cases this was performed by the referring physician;
- Molecular study by aCGH and functional studies for patient 1 were not performed by Lopes F.

Sub-chapter 2.4

The contribution of 7q33 interstitial deletions for intellectual disability

Fátima Lopes*, Fátima Torres*, Sally Ann Lynch, Arminda Jorge, Susana Sousa, João Silva, Paula Rendeiro, Purificação Tavares, Ana Maria Fortuna, Patricia Maciel

*both authors contributed equally for this work

The results presented in this chapter will be integrated in a publication of seven patients with CNVs in the 7q33 region. These results will be published in an international peer reviewed article. The author of the thesis declares that she was involved in all the work presented in this sub-chapter.

Sub-chapter 2.5

Variant Rett syndrome in a girl with a pericentric X-chromosome inversion leading to epigenetic changes and overexpression of the *MECP2* gene

Vieira JP*, Lopes F*, *et al*. *Variant Rett syndrome in a girl with a pericentric X-chromosome inversion leading to epigenetic changes and overexpression of the MECP2 gene*. Int J Dev Neurosci. 2015 Nov;46:82-7. doi: 10.1016/j.ijdevneu.2015.07.010. *authors contributed equally

The results presented in this chapter refer to publication of one patient. The author of this dissertation declares that she was involved in all the work presented in the publication except:

- Patient clinical data collection: this was performed by the referring physician (Vieira JP);
- Molecular study by Sanger sequencing, X chromosome inactivation, conventional karyotyping and FISH were not performed by Lopes F.

Sub-chapter 3.1

Identification of novel genetic causes of Rett syndrome-like phenotypes

Fátima Lopes*, Mafalda Barbosa*, Adam Ameur, Gabriela Soares, Joaquim de Sá, Ana Isabel Dias, Guiomar Oliveira, Pedro Cabral, Teresa Temudo, Eulália Calado, Isabel Fineza Cruz, José Pedro Vieira, Renata Oliveira, Sofia Esteves, Sascha Sauer, Inger Jonasson

*authors contributed equally

The results presented throughout this chapter are integrated in a publication performed in collaboration with other authors and laboratories, in which the two first authors (Lopes F. and Barbosa M.) contributed equally for the work. Because unless complete with the data from the collaborators the work performed by Lopes F. wouldn't make sense, the complete publication is presented in this sub-chapter. The author of the thesis declares that she was involved in all the work presented in the publication except:

- Clinical data collection: work performed by Mafalda Barbosa, Gabriela Soares, Joaquim de Sá, Ana Isabel Dias, Guiomar Oliveira, Pedro Cabral, Teresa Temudo, Eulália Calado, Isabel Fineza Cruz, José Pedro Vieira and Renata Oliveira;
- Clinical data summary and interpretation (in particular the individual cases clinical description, table II and figure 1): work performed by Barbosa M. and Maciel P;

- CNV screening in the cohort (except patient 7): samples were processed with Illumina Infinium OmniExpress array at the Genomics Core of Icahn School of Medicine at Mount Sinai (NY, USA) by Barbosa M. and Pinto D;
- Exome data generation for the cohort: library preparation, SOLID sequencing, mapping and variant calling was performed in the Department of Immunology, Genetics and Pathology of the Uppsala University (Uppsala, Sweden) by Ameur A, Inger Jonasson, Ann-Christine Syvänen and Ulf Gyllensten.

CHAPTER 1

General Introduction

Intellectual disability

Intellectual disability (ID) is a very frequent and disabling type of neurological condition affecting school-aged children and has an estimated prevalence of approximately 1% in the general population in developed countries (Maulik *et al.*, 2011). The diagnosis of ID is based on three criteria: (I) significant sub-average general intellectual functioning (IQ<70), (II) limitation in at least two skills of adaptive behavior (communication, self-care, home living, social/interpersonal skills, use of community resources, self-direction, functional academic skills, work, leisure, health and safety) and (III) onset of symptoms before 18 years of age (Salvador-Carulla *et al.*, 2008). In the general population the mean intelligence quotient (IQ) score is distributed around 100. When an individual scores an IQ below 70, that person is classified as intellectually handicapped. The ID levels according with the IQ score, proportion of cases within each class, equivalent mental age and predicted functional level in adulthood are presented in table I.

Table I – Classification of ID level according with the IQ range, percentage of individuals affected, expected mental age and functional level in adulthood.

Severity	IQ	Proportion of diagnosis	Equivalent mental age in adulthood	Functional level
Borderline	70-84	+NA	NA	Equivalent to the general population or requiring minimal social support
Mild	50-69	85%	9-12y	Might be able to live independently but with social support
Moderate	35-49	10%	6-9y	Has some communication and self-help skills; requires moderate supervision
Severe	20-34	4%	3-6y	Has only basic self-help and communication skills; requires supervision
Profound	<20	1%	<3y	No self-help or communication skills; requires constant and structured living conditions

In addition to the severity level, ID can also be classified into syndromic (if associated with clinical, radiological or metabolic co-morbidities) or non-syndromic ID (when ID is the main clinical feature, only accompanied by other minor features - which corresponds to a majority of the patients) (Frints *et al.*, 2002). In the clinical practice however, this division is not always straightforward and easy to make (Ropers, 2006). Although ID itself is not an isolated disease, in this work we will use the term ID to designate a group of pathologies in which the dominant characteristic is ID.

Genetics of intellectual disability

ID can be classified in two large groups based on its etiology: genetic and non-genetic. The non-genetic contributors of ID are not fully known but it is suggested that these can be prenatal (for example maternal drug use, fetal infections), perinatal (for example trauma, asphyxia, infections) or postnatal (mostly caused by infections but also by nutritional deficits) (Huang *et al.*, 2016). The genetic causes of ID account for 30%-50% of all ID cases (Curry *et al.*, 1997) and can occur through (I) chromosomal rearrangements leading to deleterious alterations of gene dosage, (II) deregulation of imprinted regions or genes, (II) dysfunction of single genes required for cognitive development (these can occur via small mutations, regulatory effects or single gene copy number alterations) (Chelly *et al.*, 2006).

Neurodevelopmental disorders (NDDs) can also be caused by inborn errors of metabolism. For this specific sub-type of disorders developed countries have implemented population-scale newborn screenings that allow the detection and treatment of more than 40 different metabolic disorders (Sharer, 2016). The neonatal metabolic screening comprises the first point in post-natal life in which children carrying a genetic disease can be identified (even without showing any symptoms). After this point only when manifesting symptoms of disease do the children undergo an advanced set of tests that includes specific genetic and metabolic workups and neuroimaging studies.

For decades the workflow for the study of children with NDD started with the search for genomic imbalances by conventional karyotyping, followed (when the results were normal) by the study of specific microdeletion/microduplication syndromes and subtelomeric analysis (either by MLPA or FISH). Only after these studies, was testing for the most frequent single gene disorders (such as Fragile X syndrome) performed (Shaffer and Bejjani, 2004). Currently, this paradigm changed and genome-wide approaches are being used as the first-tier test for NDD (Miller *et al.*, 2010; Ales *et al.*, 2016).

Copy number variations role in intellectual disability

A structural variation is the term used to describe genomic rearrangements, which can be translocations, insertions, inversions, deletions, duplications, heterozygosity losses or even more complex rearrangements originating from the combination of some of these anomalies (figure 1) (Alkan *et al.*, 2011; Bessa *et al.*, 2012).

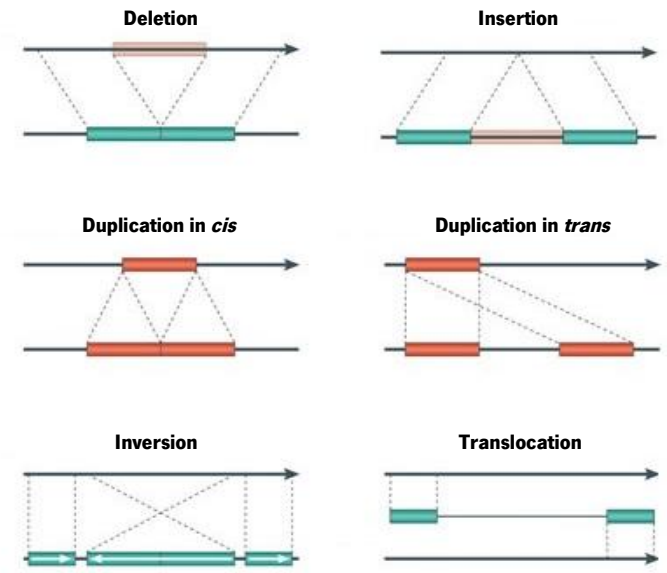


Figure 1 – Examples of possible structural variations in the genome. Adapted from Alkan C. *et al* 2011.

Copy number variations (CNVs) (including indels) are estimated to account for up to 1.2% of human genome variability, being the most important source of genomic variation. Single nucleotide polymorphisms (SNPs) and inversions are estimated to explain 0.1% and 0.3% of the difference from the reference human genome, respectively (Pang *et al.*, 2010). In some populations, up to 20% of differences in gene activity are associated with copy-number variations (Pennisi, 2007). This wide range of genomic imbalances can lead to different clinical outcomes, with large chromosomal abnormalities and large structural variations often resulting in classic syndromes (e.g. Down syndrome, Cri-du-chat syndrome) associated with additional phenotypic features to ID and clinically recognizable often by dysmorphology analysis. These are often larger than 5Mb-10Mb and can be detected by conventional karyotyping (microscopical analysis of all the chromosomes by G-banding) or by a targeted analysis (only looking at specific regions, by fluorescent in situ hybridization (FISH) or multiplex ligation-dependent probe amplification (MLPA)) (Vissers *et al.*, 2010; Martin *et al.*, 2015). Objectively speaking, G-banded karyotyping is a genome-wide assay (although one with very low resolution) and for many years was the available method for research and diagnosis. However, only a very low percentage of ID cases were solved due to the low resolution of this methodology, which even in its high-resolution form can miss variants within the 3Mb-10Mb range, and the fact that it is subject to inter-observer variations, leading to many unsolved cases (Shaffer and Bejjani, 2004). Techniques as FISH and MLPA were also widely used, together with conventional or high-resolution karyotyping, in order

to make a more specific, targeted analysis of a selected region. The problem with this approach is that the suspicion of a syndrome would have to be made based upon a clinical presentation which, consequently, would have to be consistent with a previously defined one (Riggs *et al.*, 2014). In a situation in which the suspected diagnosis was incorrect, the patients presented a variable phenotype or if the molecular basis of that presentation was unknown, this type of approach would likely lead to a negative result.

Array comparative genomic hybridization

The array comparative genomic hybridization (aCGH) technique makes the comparison of the DNA of a patient with the DNA of a healthy control (more often a pool of controls) through the labeling of both samples with different fluorochromes so that they can be distinguished. Both samples are competitively co-hybridized in a slide containing the immobilized reference DNA fragments (probes) that represent the genome (figure 2) (Shinawi and Cheung, 2008).

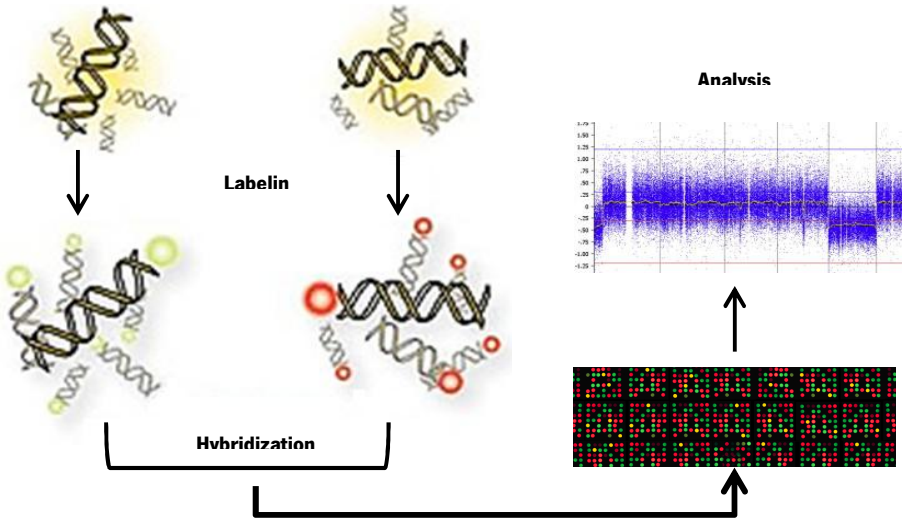


Figure 2 – Schematic representation of the aCGH technique.

The resolution of an aCGH is determined by the number of probes used and the genomic distance between each of the probe. Currently, most of the laboratories around the world make use of two types of aCGH: targeted and whole genome. In a targeted array, the entire genome is usually covered but the genomic regions containing known disease-associated deletions or duplications are analyzed at a higher resolution. The whole-genome array provides an equally high resolution throughout the entire genome (Vermeesch *et al.*, 2012). The availability of a "truly" genome-wide technique such as aCGH was what made possible for those patients with

extremely rare variations and clinically ambiguous syndromes to receive a diagnosis (reviewed by Riggs *et al.*, 2014).

The discovery of new microdeletion and microduplication syndromes

The contribution of contiguous gene syndromes (deletions or duplications of several adjacent genes) has for many years been known as contributors for ID and components of recognizable syndromes (Schmickel, 1986). The advent of aCGH use in the clinics has increased the rate of discovery of novel (previously undetectable due to their reduced size) microdeletion/microduplication syndromes. Until then, recurrent deletions or duplications in the same genomic region used to be found when a group of patients with overlapping phenotype was gathered, the definition of new syndromes being based on an accurate phenotype-genotype correlation – the phenotype-first approach to discovery (Vissers *et al.*, 2010). An important aspect is that the common affected region needed to be very similar in size and genes affected (so called recurrent) in order to result in a similar and clinically identifiable phenotype, which still implied an adequate clinical training and experience. On the other hand, in the case of nonrecurrent CNVs, the individuals are unlikely to share the same breakpoints, which could originate a broader clinical presentation (Capalbo *et al.*, 2017). The widespread use of aCGH by laboratories around the world allowed the shift to a genotype-first approach, in which patients with overlapping genomic imbalances are gathered and the phenotype revisited on a second step, leading to the definition of novel syndromes which have the advantage of having a confirmed common etiology (Carvill and Mefford, 2013). This approach was extremely successful in the course of the last years, resulting in the current list of approximately 70 novel microdeletion and microduplication syndromes associated with developmental disorders in the Decipher database ('DECIPHER v8.1', n.d.), of which 51 are associated with NDD/ID (table II).

Table II – List of 51 microdeletion and microduplication syndromes listed in Decipher as associated with neurodevelopmental disorders ('DECIPHER v8.1', n.d.).

Chr	Syndrome	Associated disorder	Minimal alteration size	Grade classification
1	1p36 microdeletion syndrome	NDD/ID	12.83 Mb	I - Pathogenic
1	1q21.1 recurrent microduplication (possible susceptibility locus for neurodevelopmental disorders)	NDD/ID	1.35 Mb	III - Susceptibility Locus
1	1q21.1 recurrent microdeletion (susceptibility locus for neurodevelopmental disorders)	NDD/ID	1.35 Mb	III - Susceptibility Locus
2	2p21 microdeletion syndrome	NDD/ID	179.13 kb	NC
2	2p15-16.1 microdeletion syndrome	NDD/ID	2.53 Mb	NC
2	2q33.1 deletion syndrome	NDD/ID	8.28 Mb	I - Pathogenic
2	2q37 monosomy	NDD/ID	352.78 kb	I - Pathogenic
3	3q29 microduplication syndrome	NDD/ID	1.62 Mb	NC
3	3q29 microdeletion syndrome	NDD/ID	1.62 Mb	NC
4	Wolf-Hirschhorn syndrome (4p16.3 microdeletion)	NDD/ID	541.04 kb	I - Pathogenic
5	Cri du Chat syndrome (5p deletion)	NDD/ID	12.52 Mb	I - Pathogenic
5	Sotos syndrome (5q35.3 microdeletion)	NDD/ID	1.33 Mb	I - Pathogenic
7	7q11.23 duplication syndrome	NDD/ID	1.40 Mb	
7	Williams-Beuren Syndrome (7q11.23 microdeletion)	NDD/ID	1.40 Mb	I - Pathogenic
8	8p23.1 duplication syndrome	NDD/ID	3.66 Mb	NC
8	8p23.1 deletion syndrome	NDD/ID	3.66 Mb	NC
8	8q21.11 microdeletion syndrome	NDD/ID	539.78 kb	NC
9	9q subtelomeric deletion syndrome	NDD/ID	217.14 kb	I - Pathogenic
11	WAGR 11p13 deletion syndrome	NDD/ID	650.75 kb	NC
11	Potocki-Shaffer syndrome (11p11.2 microdeletion)	NDD/ID	2.06 Mb	I - Pathogenic
12	12p13.33 microdeletion syndrome	NDD/Speech delay	266.47 kb	NC
12	12q14 microdeletion syndrome	NDD/ID	3.57 Mb	NC
15	Prader-Willi syndrome (Type 1) (15q11.2, BP1-BP3)	NDD/ID	5.69 Mb	I - Pathogenic
15	Angelman syndrome (Type 1) (15q11.2, BP1-BP3)	NDD/ID	5.69 Mb	I - Pathogenic

(Cont.)

15	Prader-Willi syndrome (Type 2) (15q11.2, BP2-BP3)	NDD/ID	4.82 Mb	I - Pathogenic
15	Angelman syndrome (Type 2) (15q11.2, BP2-BP3)	NDD/ID	4.82 Mb	I - Pathogenic
15	15q13.3 microdeletion syndrome	NDD/ID	1.54 Mb	NC
15	15q24 recurrent microdeletion syndrome	NDD/ID	1.56 Mb	NC
15	15q26 overgrowth syndrome	NDD/ID	3.16 Mb	NC
16	ATR-16 syndrome (16 deletion)	NDD/ID	774.37 kb	I - Pathogenic
16	Rubinstein-Taybi Syndrome (16p13.3 microdeletion)	NDD/ID	155.07 kb	I - Pathogenic
16	16p13.11 recurrent microduplication (neurocognitive disorder susceptibility locus)	NDD/Autism	1.50 Mb	NC
16	16p13.11 recurrent microdeletion (neurocognitive disorder susceptibility locus)	NDD/ID	1.50 Mb	NC
16	16p11.2-p12.2 microduplication syndrome	NDD/ID	7.81 Mb	NC
16	16p11.2-p12.2 microdeletion syndrome	NDD/ID	8.69 Mb	NC
16	Recurrent 16p12.1 microdeletion (neurodevelopmental susceptibility locus)	NDD/ID	520.76 kb	NC
16	16p11.2 microduplication syndrome	NDD/Autism	593.00 kb	NC
17	Miller-Dieker syndrome (MDS) (17p13.3 microdeletion)	NDD/ID	2.59 Mb	I - Pathogenic
17	Potocki-Lupski syndrome (17p11.2 duplication syndrome)	NDD/Autism	3.45 Mb	NC
17	Smith-Magenis syndrome (17p11.2 microdeletion)	NDD/ID	3.45 Mb	I - Pathogenic
17	NF1-microdeletion syndrome (17q11.2 microdeletion)	NDD/ID	1.16 Mb	I - Pathogenic
17	17q21.31 recurrent microdeletion syndrome (Koolen de Vries syndrome)	NDD/ID	589.24 kb	I - Pathogenic
22	Cat-Eye syndrome (Type I) (22p deletion)	NDD/ID	16.97 Mb	NC
22	22q11 deletion syndrome (Velocardiofacial / DiGeorge syndrome)	NDD/ID	2.44 Mb	I - Pathogenic
22	22q11 duplication syndrome	NDD/ID	2.44 Mb	III - Susceptibility Locus
22	22q11.2 distal deletion syndrome	NDD/ID	1.81 Mb	NC
22	22q13 deletion syndrome (Phelan-Mcdermid syndrome)	NDD/ID	142.33 kb	I - Pathogenic
X	Xp11.22-p11.23 microduplication	NDD/ID	3.78 Mb	NC
X	Xp11.22-linked intellectual disability	NDD/ID	282.21 kb	NC
X	Xq28 (*MECP2*) duplication	NDD/ID	75.93 kb	NC

(Cont.)

X	Xq28 microduplication	NDD/ID	257.29 kb	NC

NDD – Neurodevelopmental disorder; ID – Intellectual disability; NC – Not classified by Decipher

The interpretation challenge of copy number variations

The molecular findings of an aCGH analysis might not always be amenable to a straightforward interpretation. The results might include (I) the identification of well-characterized pathologic genomic imbalances in patients that fit the clinical presentation of the syndrome or (II) in patients with an atypical presentation. Also, (III) the identification of new genetic abnormalities can be a common finding, the study of the parents and control populations being essential for conclusions. In the latter situation, searching for other patients with similar CNVs is also necessary in order to make a better genotype-phenotype correlation, as is the investigation of the genes involved in the imbalance and their known functional relevance, particularly in the nervous system (Edelmann and Hirschhorn, 2009).

Variable expressivity of microdeletion and microduplication syndromes

In general, the absence of a given genomic portion often leads to more severe consequences than its excess, and variable expressivity and incomplete penetrance are more prevalent for duplications than for deletions, although they may also occur in deletion syndromes (Edelmann and Hirschhorn, 2009). A good example of this situation is the 22q11.2 microdeletion syndrome (also known as DiGeorge or velocardiofacial syndrome) and the 22q11.2 microduplication, both which are associated with increased risk of developing cognitive and psychiatry diseases. The 22q11.2 microdeletion is the most frequent microdeletion syndrome in the general population (Harris, 2017). Patients have multiple anomalies (cleft palate, cardiac defects, immune dysfunction and hypocalcaemia) and variable psychiatric and cognitive symptoms (ID, autism, ADHD, anxiety and schizophrenia) (Biswas and Furniss, 2016). On the other hand, the microduplication is quite rarer and the phenotype can range from asymptomatic carriers to patients with ID, speech delay, growth retardation and/or hypotonia. Additionally, even within patients, the ID level can be very heterogeneous (Van Campenhout *et al.*, 2012; Torres *et al.*, 2016). The factors responsible for the clinical variability in 22q11.2 CNVs and others are still poorly understood. However, for each individual case that we face in the clinic the following factor should be taken in consideration: the specific genetic background of each person, the individual epigenetic differences and the possibility of a second molecular change that is still undetermined (the second hit hypothesis) (Watson *et al.*, 2014).

Classification of variants

The American College of Medical Genetics recommends, for a systematic evaluation and clinical interpretation of CNVs: the familiarization with well-established contiguous gene syndromes, the consideration of CNV size, the consideration of the individual genes involved in the CNV interval and the comparison with internal and external databases (Kearney *et al.*, 2011).

Currently there are four databases that are widely used in the interpretation of CNVs in the context of ID (Martin *et al.*, 2015) (table III).

Table III – Copy number variants databases for control population and patients. Adapted from Martin C. *et al* 2015.

Database	Controls database	Patients database
Database of Genomic Variants (DGV)	Yes	No
dbVar	Yes	Yes
Decipher	No	Yes
ClinVar	No	Yes

One of the most useful tools currently available for finding patients with CNVs (either with very or recurrent CNVs) is the Decipher database (https://decipher.sanger.ac.uk). The website collects information of more than 18.000 patients, from more than 250 different laboratories ('DECIPHER v8.1', n.d.). In order to facilitate the interpretation and genotype-phenotype correlations, the website divides the already known CNVs into three different grades:

Pathogenic (grade I): this category of variants corresponds to highly penetrant CNVs whose pathogenicity gathers agreement among geneticists. All the cases in the literature with these variants have a clinical phenotype, there are no heathy carriers with it, nor it is present in healthy control databases, and even though some variability might be observed there are core common features to all the cases. Often the key gene(s) is known and functional data confirming it is available, as well as consensus between multiple sources of information (OMIM, GeneReviews, etc.). Genetic counseling is available and clinical management implications are often defined.

Likely pathogenic (grade II): a likely pathogenic CNV is a highly penetrant variant but with variable phenotypic features other than a main one (usually DD or NDD/ID related). Although all the reported cases usually have a clinical phenotype, the number of patients with this category of CNVs is (still) not very high, which also limits the phenotype-genotype correlations. Also, there is usually no available

functional data for the genes involved or confirmed pathogenic genes. Nevertheless, these can be used for diagnostic purposes and reproductive counselling.

Susceptibility locus (grade III): a CNV in a susceptibility locus can be present in affected patients, unaffected parents, other healthy individuals and even control populations. Its penetrance is incomplete, and the phenotype mild or highly variable, resulting in inconsistent features. For these CNVs the possibility of a second contributing diagnosis should always be considered and caution should be taken in the diagnostic and genetic counseling context. The recent developments and challenges concerning these *loci* have recently been reviewed by Torres et al (Torres *et al.*, 2016).

In the clinical practice, an aCGH analysis can retrieve a CNV that fits into one of these three categories present in Decipher, or in two others: benign variant or variant of unknown significance (VOUS). Benign variants are always found in any aCGH analysis and represent common polymorphisms and/or CNVs present at a significant number in controls (usually more than three controls) (Edelmann and Hirschhorn, 2009; Kearney *et al.*, 2011). VOUS represent a broader category of variants, that can later be classified as benign or pathogenic, but for which at the time of analysis there is not enough evidence for unequivocal classification (Kearney *et al.*, 2011).

Single nucleotide variations role in intellectual disability

A shift in paradigm

The advances in the cost-effectiveness ratio of aCGH technology has led to a replacement of conventional karyotyping in the first-tier genetic testing (Miller *et al.*, 2010). During the last ten years large technical advances were made in both the array and massive parallel sequencing (MPS) techniques, making these two technologies leading strategies in the genetic studies of many disorders including ID (Watson *et al.*, 2014). Although aCGH can be considered the "star" of the last ten years of ID genetics, with laboratories throughout the world making the shift from conventional to molecular karyotyping, MPS had an exponential increase when applied to unraveling the molecular cause of ID. This shift in paradigm is possible to observe by the number of publications in Pubmed between the beginning of 2006 and the end of 2016 for ID together with aCGH or MPS (figure 3A and 3B).

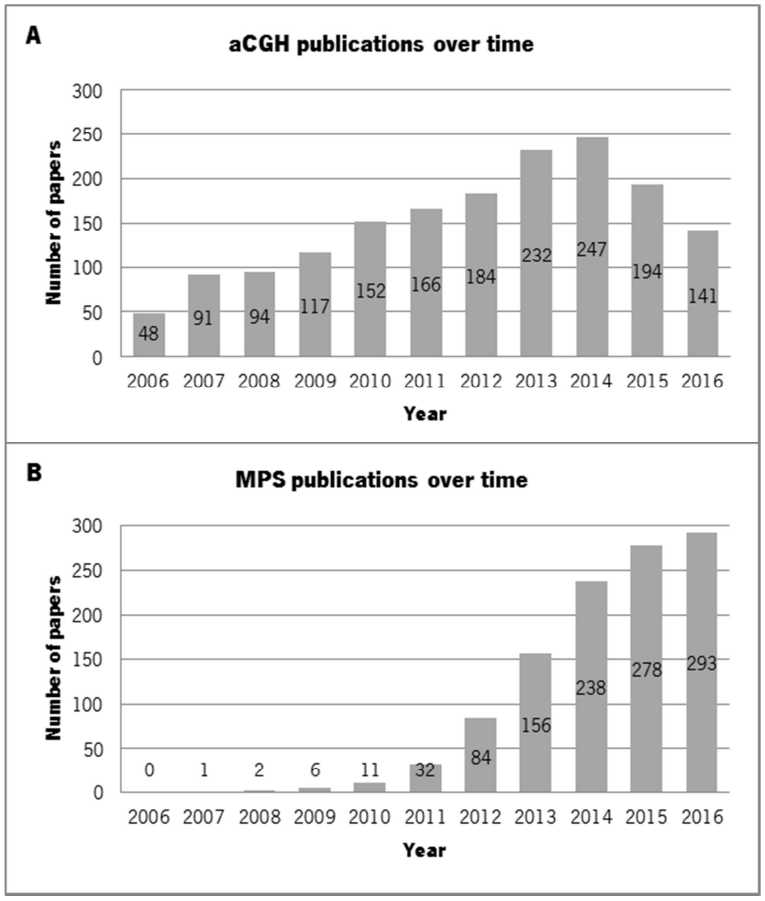


Figure 3 – Number of publication in PubMed per year in the last ten years for ID using aCGH and MPS based technologies. (A) Number of PubMed publications per year using the search terms "(intellectual disability OR mental retardation) AND (array comparative genomic hybridization OR chromosomal microarray)" where it is possible to observe a progressive increase followed by a decrease in the last two years, and (B) using the terms "(intellectual disability OR mental retardation) AND (next-generation sequencing OR exome sequencing)" where the exponential increase in the publication number since 2011 is visible.

In PubMed, it is possible to observe that in spite of the constant increase for many years, since 2015 there is a decrease in the number of identified publications combining the search terms "ID and aCGH" (figure 3A). On the other hand, since 2011 the search terms combining "ID and NGS" (next generation sequencing) have increased very significantly (figure 3B) and it is likely that this tendency is maintained in the upcoming years. With the advances in technology and the variety of platforms available in the market the cost of MPS has dramatically decreased in the last decade, and more developments are expected in the near future. MPS technologies are having a major impact on the medical practice and have popularized the whole exome sequencing approach in research (de Ligt *et al.*, 2012; Rauch *et al.*, 2012). In the near future, whole-genome sequencing will be on the basis of personalized genomic medicine, integrating phenotypic data and familial history into the patient genome, providing personalized and unique individual-directed healthcare (Gullapalli *et al.*, 2012).

However, we should keep in mind that even though aCGH has limited resolution, the currently most used MPS approach in clinical practice is exome sequencing, which assesses only a fraction of the genome (Watson *et al.*, 2014). For this reason, we can anticipate that in the next years a similar shift will happen between exome and genome sequencing.

Next-generation or massive parallel sequencing – principles of the technique

The decoding of the human genome together with the need for a faster and cheaper sequencing method opened the door for the development of high-throughput sequencing technologies (Xuan *et al.*, 2012). In MPS technologies the DNA is fragmented into small pieces that will originate DNA templates that are afterwards randomly read along the entire genome in parallel (figure 4). This is why this technology is also called MPS (Zhang *et al.*, 2011).

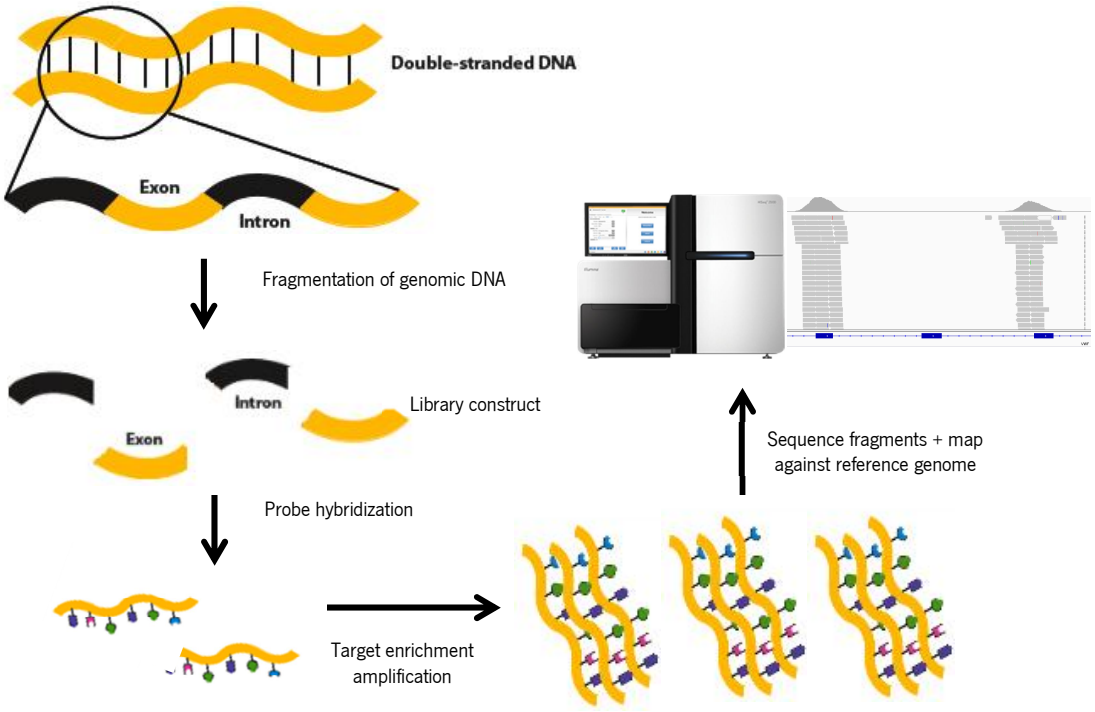


Figure 4 - Schematic representation of the MPS technique

Sequencing technologies have evolved rapidly in the last decade, with very significant technical improvements and decrease in the cost-per-base. In the last years, a multitude of commercial companies offering NGS services (at research and clinical level) has evolved. The strong competition among these companies has proven to be advantageous for scientists, since it lead to the significant decrease of the cost per Mb of sequence (Gullapalli *et al.*, 2012). Together with this trend, there was a fast developing interest in academia and private diagnostic laboratories to establish NGS services for

clinical diagnosis. In the past decade two main types of instruments were introduced in the market: second generation platforms and third generation platforms (Mardis, 2017). The second generation platforms require the amplification of the DNA molecule in order for it to be represented in a series of clones that after will be read in the equipment. Examples of these technologies are the 454 Roche, Illumina and ABI SOLiD systems. Significant advances throughout the years has been made in these methodologies allowing the development of small bench top sequencers that are less expensive and have easier preparation protocols (Xuan *et al.*, 2012). These machines allowed the access of many small laboratories to NGS technologies. The third generation sequencing platforms are based on the reading of a single molecule (such as Pacific Biosiences and Oxford Nanopore). These technologies have limitations such as higher error rates but retrieve very long reads, an essential feature for the correct assembly of large genomes (Mardis, 2017). The generation, analysis and storage of data generated by MPS require sophisticated informatics tools and specialized bioinformatics staff, something that is not always easy for small laboratories (Xuan *et al.*, 2012).

Another very important advance in the MPS filed in the last years was introduction in the market of benchtop MPS instruments. These equipments present a lower cost (when compared with the corresponding larger capacity equipment from the respective companies) and were created to sequence targeted areas of the genome (usually gene sets) or small genomes, makes them the first choice for many diagnostic laboratories when it comes to change the approach from Sanger based methods to NGS (Marian, 2012). Additionally, with the advent of benchtop systems there was also an effort to create commercial software solutions for NGS analysis that will help in the standardization of data analysis within and inter laboratories.

Exome sequencing in ID

In the last years, exome and targeted sequencing approaches proved to be valuable for the detection of genetic variants responsible for clinical phenotypes, previously without a known cause (Rizzo and Buck, 2012; Beaulieu *et al.*, 2013; Gregor *et al.*, 2013). Exome sequencing has allowed the discovery of new genes causing classical syndromes that had long been described but didn't have their etiology identified due to the limited number of patients, namely: Schinzel-Giedion syndrome, rare forms of Charcot-Marie-Tooth neuropathy, Freeman-Sheldon syndrome, Kabuki syndrome and hyperphosphatasia-mental retardation syndrome, Miller syndrome, among others (Hoischen *et al.*, 2010; Krawitz *et al.*, 2010; Lupski *et al.*, 2010; Ng *et al.*, 2010, p. 2, 2010; Hu *et al.*, 2011; Najmabadi *et al.*, 2011). These studies also allowed the unbiased discovery of disease genes in patients with and without specific

syndromes such as intellectual disability and glycosylation disorders (de Ligt *et al.*, 2012; Timal *et al.*, 2012). In these cases, the approach should be based on each individual patient using a trio based sequencing approach (in which the patients and both parents are sequenced).

With the widespread application of MPS to the clinics, genes that were previously associated with one type of disease started to present themselves as candidates in different disease spectra. Moreover, distinct clinical features started to be observed in association with mutations in the same gene and even with the same mutation type, highlighting the wide variability associated with NDD syndromes (Harripaul *et al.*, 2017).

Interpretation and classification challenges

The use of MPS technologies generates a very large amount of data that brings challenges in handling and interpretation (van El *et al.*, 2013). The filtering of false negative and false positive results has to be controlled when using MSP technologies. Pilot studies using previously analyzed samples should be performed in order to optimize the sequencing pipeline and for quality thresholds to be implemented (Costa *et al.*, 2013; van El *et al.*, 2013). When searching for a diagnostic it is important to be aware of the extent to which a variant can be interpreted or not as causative of a certain phenotype. If the alteration occurs in a gene with proven relation to disease the interpretation is facilitated. However, when the variant occurs in genes with an unknown relation to the phenotype more caution is needed (de Ligt *et al.*, 2012; van El *et al.*, 2013).

In 2015 the American College of Medical Genetics (ACMG) released a statement of the standards and guidelines to be followed in the interpretation of SNVs detected by MPS. This document provides a series of criteria and combination rules in order to ultimately classify a variant as pathogenic, likely pathogenic, benign, likely benign or of uncertain significance (Richards *et al.*, 2015). In a similar fashion to what happen with the interpretation of CNVs significance, SNVs are also subjected to doubt and can result in a range of variable expressivity and therefore face similar challenges, as mentioned before.

Incidental findings

When performing a genome-wide analysis there is always the risk of identifying genetic alterations recognizable to be related to additional phenotypes that were not initially expected or considered (for example: observing a pathogenic variant in BRCA1 gene in a young girl tested due to ID). These findings are now classified as “unsolicited (or incidental) findings” (van El *et al.*, 2013). Currently the ACMG

defines a "incidental finding" as the result obtained of a deliberate search for pathogenic or likely pathogenic variants in genes that are not relevant for the disease for which the patient is being studied (for example, looking for a susceptibility to breast cancer gene in a patient in which the exome is being done because of NDD). The ACMG also released a curated list of carefully chosen 56 genes that could lead to 24 disorders for which early intervention can impact in outcome, and thus for which communication of findings to the patient should be considered according to the ethical principle of beneficence (Green *et al.*, 2013). This is challenged by other groups of scientists but is an important aspect of the implementation of these approaches (Quinlan-Jones *et al.*, 2016).

When using MPS it is absolutely necessary for the patient (or responsible adult, such as the parents of a child or person with ID) to sign an informed consent in which he/she is informed of the possible implications of the study, including the possibility of unsolicited findings and what will happen if they are detected. However, the content of the informed consent should be defined by the health professionals involved, and this should be revised/discussed with the patient/person in charge when facing a specific situation (for instance, when detecting a susceptibility factor for a certain type of cancer that can be prevented by a tighter preventive screening) (Green *et al.*, 2013; van El *et al.*, 2013). Concerning this subject, clinicians should keep sharing their experiences regarding genomic studies in order to keep improving the best practices for counseling and informed consent procedures (van El *et al.*, 2013; Amendola *et al.*, 2015).

Biological processes underlying ID

Whenever a gene is found to be mutated or suspected to be a key gene in a CNV in a patient, the strategy is to see if its biological role was determined before and whether it is connected to a pathway/network previously associated with ID. In order to do this, knowledge about the main mechanistic themes involved in NDD/ID related disorders is an important factor for the proper evaluation of a gene's contribution in the patient. Functional links are possible to identify among the currently known ID causative genes and several of these operate in two main functional themes: neurodevelopment (namely neurogenesis, cell proliferation or differentiation and neuronal migration) and synapse function (namely synapse formation and plasticity) (Kramer and van Bokhoven, 2009) (schematic summary in figure 5).

Neurogenesis

Defects in neurogenesis-related genes are often found in patient with ID associated with alteration in head circumference (microcephaly or macrocephaly). These alterations can be heterogeneous, arise

from the abnormal number of neurons and can be de novo or familial and are very often associated with ID (Francis *et al.*, 2006). Microcephaly (MIC) is determined by reduced frontal-occipital head circumference (more than 3 SD below the mean) and can be primary (present since birth) or secondary (acquired during childhood). Multiple genes have been detected as having a role in controlling head size, often associated with cell division and cell cycle regulation processes. *MCPH1*, *ASPM*, *CDK5RAP2* and *CENPJ* are examples of genes in which recessive mutations cause ID and MIC (figure 2). Often these genes are also involved in mitotic spindle dynamics and DNA repair (Pulvers *et al.*, 2015).

Neuronal migration

After neurogenesis, neurons have to migrate from the ventricular to cortex in an organized and timely process. Neuronal migration disorders can arise from the disruption of several stages of the migration process (Vaillend *et al.*, 2008). Alterations in genes related with this process often result in anomalies in the structural organization of the neuronal layers in the cortex and often results in cortical malformations including lissencephaly (smooth brain without the proper gyri and sulci), polymicrogyria (excessive number of small and partly fused gyri separated by shallow sulci resulting in an irregular brain surface), schizencephaly (formation of abnormal unilateral or bilateral clefts in the cerebral hemispheres) or heterotopia (presence of grey matter in incorrect places). A distinctive feature of patients with mutations in genes causing anomalies of neuronal migration is that ID is often associated with epilepsy (Guerrini and Parrini, 2010). As an example, the *LIS1* and *TUBA1A* genes encode microtubule associated proteins (MAPs) for which recessive mutations are associated with ID and lissencephaly (figure 2) (Parrini *et al.*, 2016).

Synaptic function

Synapses are specialized points of contact between neurons that allow neurotransmission. Their functional and structural modification by experience is thought to be the basis of memory and learning. The inappropriate function of these structures leads to the disruption of neuronal circuits and to brain disease, which is the case of ID. In childhood, the brain suffers intense reorganization of the synaptic connections reaching the plasticity peak in the first two years of life (Zukin *et al.*, 2009). For these reasons, is not surprising that inumerous genes involved in several pathways related to synaptic function have been described in patients with ID.

Synaptogenesis is the process of creating news synapses. For this, the activation of genes encoding synaptic proteins is required as well as the formation of vesicles trafficking pre- and post-synaptic protein complexes, in order to lead to the formation of functional pre- and post-synaptic compartments (Vaillend *et al.*, 2008). An excitatory synapse is usually formed from small dendritic protrusions called

dendritic spines and includes a presynaptic and a postsynaptic site, separated by the synaptic cleft and held together by cell adhesion molecules (Tallafuss *et al.*, 2010).

A recurrent finding in neurons harboring mutations in patients with ID is the abnormal structure of synapses, namely the increase in length of the dendritic spines. In fact, the length of the dendritic spine is correlated with the ability and efficacy of the spine to be plastic. This structure and its dynamics is regulated by the underlying cytoskeleton (Svitkina *et al.*, 2010). Throughout the years many genes related with cytoskeleton regulation have been discovered to be mutated in patients with ID, such as *ARHGEF6* and *PAK3* that belong to Rho GTPase pathway and were the first cytoskeleton-related genes to be associated with ID (von Bohlen Und Halbach, 2010). Other genes such as *TRIO*, *AUTS2* and *OPHN1* also encodes for members of the Rho-GTPase pathway and are known to cause ID (Barresi *et al.*, 2014, p. 1; Hori *et al.*, 2014; Pengelly *et al.*, 2016). Cytoskeleton alterations also might impair axon growth and guidance, also important for the correct formation of neuronal network and synapses.

Another equality important component of the synapse (of the chemical type) is vesicle trafficking machinery since the neurotransmitters are not able to cross the synaptic cleft alone. The docking and fusion of the vesicles to the membrane is regulated by the SNARE (soluble N-ethylmalei-mide-sensitive factor attachment protein receptor) complex (Lin and Scheller, 2000). Mutations that lead to the alteration of proteins involved in vesicle targeting, docking or fusion are good causal candidates for ID. *STXBP1* is one of the genes belonging to this category, for which mutations cause ID and epilepsy (Lopes *et al.*, 2016; Vlaskamp *et al.*, 2016). Also, *STX1A*, *STX3* and *SNAP25B* are other examples of genes belonging to the SNARE complex that when mutated are associated with ID related disorders (Gao *et al.*, 2010; Shen *et al.*, 2014; Chograni *et al.*, 2015, p. 3)

Cell adhesion molecules such as integrins and cadherins are prevalent in synapses and play a crucial role in their correct positioning, structure and function. Neurexins (pre-synaptic) and neuroligins (post-synaptic) are examples of these molecules that were found mutated in patients with ID and autism (*NLGN3*, *NLGN4* and *CNTNAP2*).

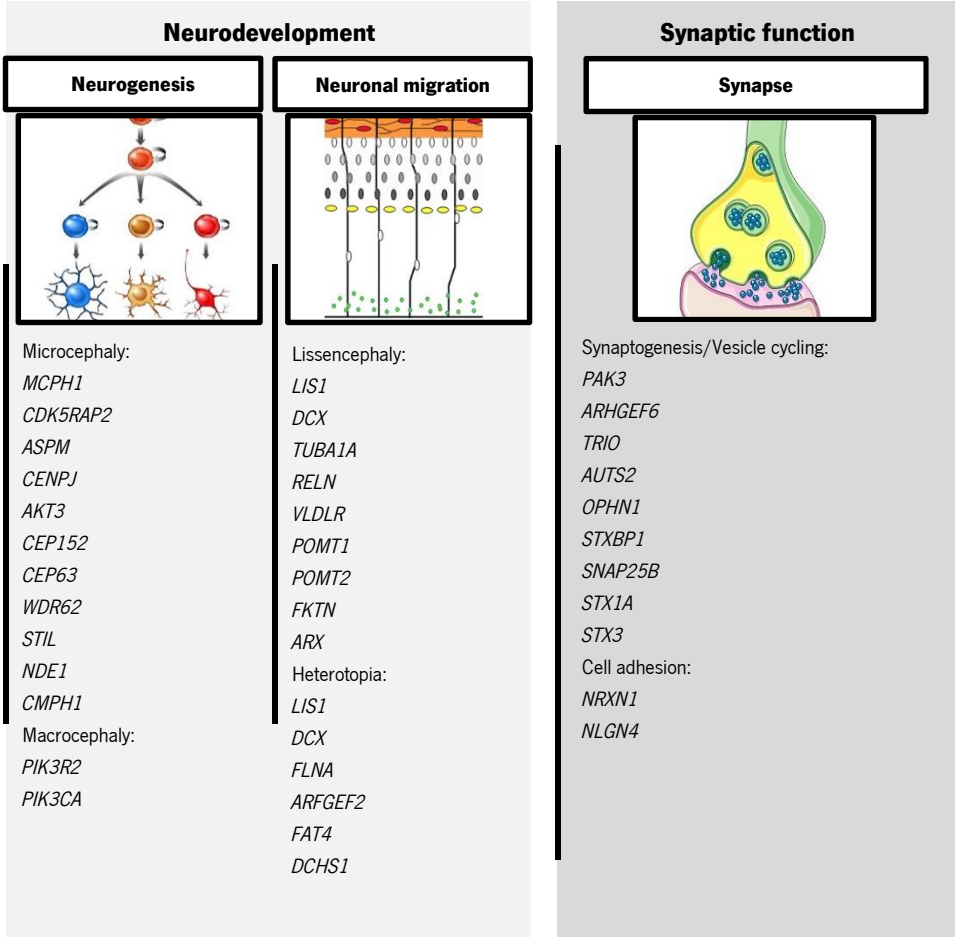


Figure 5 - Some examples of genes found to be mutated in NDD and ID and the functional process where they belong. On the left (light grey) two steps of neurodevelopment (neurogenesis and neuronal migration) are represented; on the right (dark grey) synaptic function is represented.

Transcription regulation

Gene expression is a strictly regulated process that contributes tremendously for neuronal differentiation and synaptic plasticity. In fact, chromatin-remodeling factors are always present in pre- and post-synaptic compartments (Penzes *et al.*, 2013). Several NDD and ID related disorders are caused by mutations in genes encoding regulators of signal transduction pathways and transcription factors involved in chromatin remodeling, gene expression and protein maturation (Izumi, 2016). Alterations in any of these mechanisms can result in a big impact in the regulation of genes in critical developmental stages. Several genes involved in transcription regulation have been described in ID patients such as *NF1* (that causes neurofibromatosis type 1 and is responsible for the ID presentation in the patients), *MECP2* (that causes Rett syndrome) and *JARID1C* (that causes non-syndromic ID) (other examples are presented in figure 6).

mRNA and translation regulation

RNA metabolism and protein synthesis play a crucial role in axonal and dendritic function in neurons. After synapse formation the constant transport of mRNA to dendrites is required for neuronal growth and functioning (Linder *et al.*, 2015). One of the most widely known proteins is FMRP (encoded by the *FMR1* gene), altered in fragile X syndrome and that is responsible for the dendritic localization and translation of mRNAs in response to activation of mGluRs. Other genes involved in RNA processing and metabolism, such as *ZC3H14*, *EXOSC2* and *THOC2* were also more recently described mutated in patients with NDD (Kumar *et al.*, 2015; Di Donato *et al.*, 2016; Fasken and Corbett, 2016). Dysregulation of mRNA translation is also implicated in the pathogenesis of ID, including Fragile X syndrome, in which the transcriptional silencing of the *FMR1* gene leads to an increase in general translation at the dendrites (Muzar *et al.*, 2016). Other genes such as *EIF2B2*, *EIF2B4*, and *EIF2B5* (all encoding translation initiation factors GTP exchanging proteins) are described as mutated in leukoencephalopathies with vanishing white matter, also recently associated with ID (Sartor *et al.*, 2015).

Ubiquitin-proteasome system

The proper degradation of unwanted or excessive proteins at the synapse is as important as the synthesis. In synapse, the ability for it to change dynamically (plasticity) relies on its ability to synthesize and degrade proteins in a fast and efficient manner (Louros and Osterweil, 2016). The targeting of proteins to be degraded occurs through the attachment of ubiquitin tags to the protein to be eliminated. This ubiquitin tag acts as a tag for the proteins that are no longer useful or are harmful for the cell and should be eliminated (Klein *et al.*, 2016). This tagging is regulated by a group of proteins called ubiquitin ligases, which include cullins. Impairment of ubiquitin-proteasome system activity involving ubiquitin ligase genes such as *UBE3A*, *UBE3B*, *CUL4B* and *HUWE1*, has been reported in patients with ID (Sell and Margolis, 2015). More recently mutations in *TRIP12*, *PSMD12* and *HERC1* genes (all with biological functions related with proteostasis) were also described in patients with ASD (autism spectrum disorder) and ID (Küry *et al.*, 2017; Utine *et al.*, 2017; Zhang *et al.*, 2017).

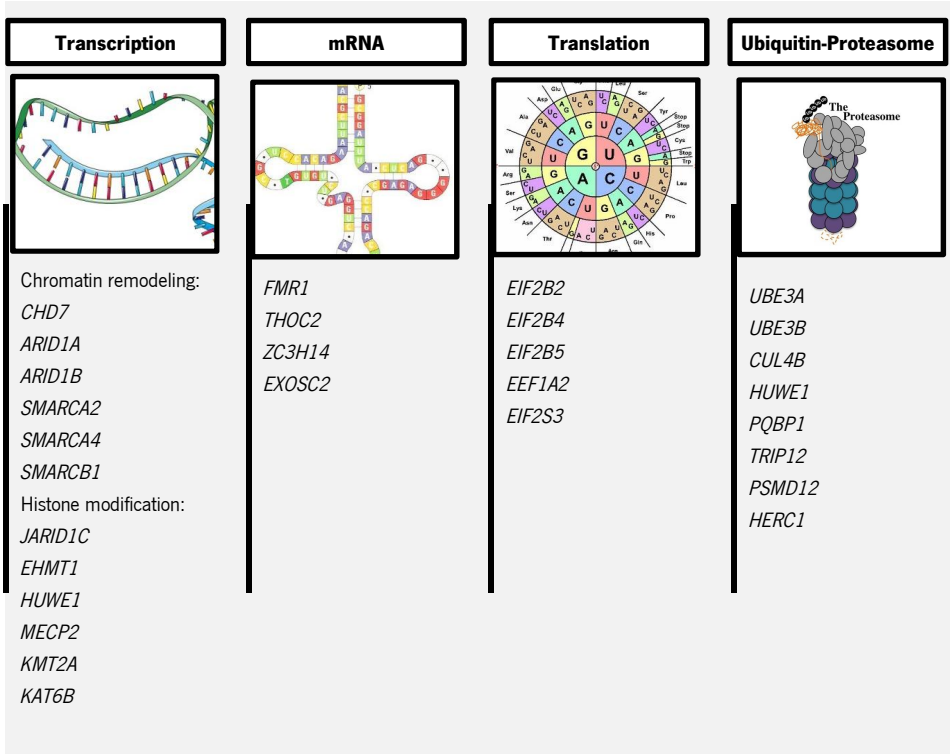


Figure 6 - Some examples of genes found to be mutated in NDD and ID four biological pathways.

Although the knowledge of the neurodevelopmental mechanisms and molecular pathways involved in NDD is essential for the correct interpretation of aCGH and MPS data, we should keep in mind that there is always room for error. In vitro analysis and bioinformatics prediction are almost always capable of linking one gene/protein to another, which can lead to false assumptions. Additionally, not all the genes are studied from the functional point of view and just because there isn't at a given moment an information about a gene´s function, particularly in the nervous system, it does not mean that there it is not contributing for ID in the patient. In this sense, mutation identification in ID patients can also contribute to increase our knowledge on the function of genes in the nervous system, in a virtuous cycle of knowledge.

Aims

In the last years several studies have revealed the contribution of genome-wide genetic approaches to the detection of variant underlying NDD.

In this work our goal was to study patients with idiopathic NDD by aCGH and MPS techniques, with the global aims of:

- To discover more insight into the genetics of ID and NDD;
- To contribute for the diagnosis of the patients studied and for the possibility of genetic counseling.

The specific aims of this work were:

- To explore the contribution of CNVs to ID in two cohorts of Portuguese patients (a clinical cohort and a research cohort);
- To define new *loci* and candidate genes for NDD;
- To explore the contribution of new (rare) CNVs and the affected genes to ID/NDD;
- To contribute for a better definition of the clinical spectrum of already known CNVs and disease associated genes.

References

Ales M, Luca L, Marija V, Gorazd R, Karin W, Ana B, et al. Phenotype-driven gene target definition in clinical genome-wide sequencing data interpretation. Genet. Med. Off. J. Am. Coll. Med. Genet. 2016; 18: 1102–1110.

Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. Nat. Rev. Genet. 2011; 12: 363–376.

Amendola LM, Dorschner MO, Robertson PD, Salama JS, Hart R, Shirts BH, et al. Actionable exomic incidental findings in 6503 participants: challenges of variant classification. Genome Res. 2015; 25: 305–315.

Barresi S, Tomaselli S, Athanasiadis A, Galeano F, Locatelli F, Bertini E, et al. Oligophrenin-1 (OPHN1), a gene involved in X-linked intellectual disability, undergoes RNA editing and alternative splicing during human brain development. PloS One 2014; 9: e91351.

Beaulieu CL, Huang L, Innes AM, Akimenko M-A, Puffenberger EG, Schwartz C, et al. Intellectual disability associated with a homozygous missense mutation in THOC6. Orphanet J. Rare Dis. 2013; 8: 62.

Bessa C, Lopes F, Maciel P. Molecular Genetics of Intellectual Disability [Internet]. 2012[cited 2017 Feb 28] Available from: http://www.intechopen.com/books/latest-findings-in-intellectual-and-developmental-disabilities-research/molecular-genetics-of-intellectual-disability

Biswas AB, Furniss F. Cognitive phenotype and psychiatric disorder in 22q11.2 deletion syndrome: A review. Res. Dev. Disabil. 2016; 53–54: 242–257.

von Bohlen Und Halbach O. Dendritic spine abnormalities in mental retardation. Cell Tissue Res. 2010; 342: 317–323.

Capalbo A, Rienzi L, Ubaldi FM. Diagnosis and clinical management of duplications and deletions. Fertil. Steril. 2017; 107: 12–18.

Carvill GL, Mefford HC. Microdeletion syndromes. Curr. Opin. Genet. Dev. 2013; 23: 232–239.

Chelly J, Khelfaoui M, Francis F, Chérif B, Bienvenu T. Genetics and pathophysiology of mental retardation. Eur. J. Hum. Genet. EJHG 2006; 14: 701–713.

Chograni M, Alkuraya FS, Ourteni I, Maazoul F, Lariani I, Chaabouni HB. Autosomal recessive congenital cataract, intellectual disability phenotype linked to STX3 in a consanguineous Tunisian family. Clin. Genet. 2015; 88: 283–287.

Costa JL, Sousa S, Justino A, Kay T, Fernandes S, Cirnes L, et al. Nonoptical massive parallel DNA sequencing of BRCA1 and BRCA2 genes in a diagnostic setting. Hum. Mutat. 2013; 34: 629–635.

Curry CJ, Stevenson RE, Aughton D, Byrne J, Carey JC, Cassidy S, et al. Evaluation of mental retardation: recommendations of a Consensus Conference: American College of Medical Genetics. Am. J. Med. Genet. 1997; 72: 468–477.

Di Donato N, Neuhann T, Kahlert A-K, Klink B, Hackmann K, Neuhann I, et al. Mutations in EXOSC2 are associated with a novel syndrome characterised by retinitis pigmentosa, progressive hearing loss, premature ageing, short stature, mild intellectual disability and distinctive gestalt. J. Med. Genet. 2016; 53: 419–425.

Edelmann L, Hirschhorn K. Clinical utility of array CGH for the detection of chromosomal imbalances associated with mental retardation and multiple congenital anomalies. Ann. N. Y. Acad. Sci. 2009; 1151: 157–166.

van El CG, Cornel MC, Borry P, Hastings RJ, Fellmann F, Hodgson SV, et al. Whole-genome sequencing in health care. Recommendations of the European Society of Human Genetics. Eur. J. Hum. Genet. EJHG 2013; 21 Suppl 1: S1-5.

Fasken MB, Corbett AH. Links between mRNA splicing, mRNA quality control, and intellectual disability. RNA Dis. Houst. Tex 2016; 3

Francis F, Meyer G, Fallet-Bianco C, Moreno S, Kappeler C, Socorro AC, et al. Human disorders of cortical development: from past to present. Eur. J. Neurosci. 2006; 23: 877–893.

Frints SGM, Froyen G, Marynen P, Fryns J-P. X-linked mental retardation: vanishing boundaries between non-specific (MRX) and syndromic (MRXS) forms. Clin. Genet. 2002; 62: 423–432.

Gao MC, Bellugi U, Dai L, Mills DL, Sobel EM, Lange K, et al. Intelligence in Williams Syndrome is related to STX1A, which encodes a component of the presynaptic SNARE complex. PloS One 2010; 5: e10292.

Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet. Med. Off. J. Am. Coll. Med. Genet. 2013; 15: 565–574.

Gregor A, Oti M, Kouwenhoven EN, Hoyer J, Sticht H, Ekici AB, et al. De Novo Mutations in the Genome Organizer CTCF Cause Intellectual Disability. Am. J. Hum. Genet. 2013; 93: 124–131.

Guerrini R, Parrini E. Neuronal migration disorders. Neurobiol. Dis. 2010; 38: 154–166.

Gullapalli RR, Desai KV, Santana-Santos L, Kant JA, Becich MJ. Next generation sequencing in clinical medicine: Challenges and lessons for pathology and biomedical informatics. J. Pathol. Inform. 2012; 3: 40.

Harripaul R, Noor A, Ayub M, Vincent JB. The Use of Next-Generation Sequencing for Research and Diagnostics for Intellectual Disability. Cold Spring Harb. Perspect. Med. 2017; 7

Harris JC. Understanding Psychiatric Disorders in People With 22q11.2 Deletion and Duplication. JAMA Psychiatry 2017

Hoischen A, van Bon BWM, Gilissen C, Arts P, van Lier B, Steehouwer M, et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nat. Genet. 2010; 42: 483–485.

Hori K, Nagai T, Shan W, Sakamoto A, Taya S, Hashimoto R, et al. Cytoskeletal regulation by AUTS2 in neuronal migration and neuritogenesis. Cell Rep. 2014; 9: 2166–2179.

Hu H, Eggers K, Chen W, Garshasbi M, Motazacker MM, Wrogemann K, et al. ST3GAL3 mutations impair the development of higher cognitive functions. Am. J. Hum. Genet. 2011; 89: 407–414.

Huang J, Zhu T, Qu Y, Mu D. Prenatal, Perinatal and Neonatal Risk Factors for Intellectual Disability: A Systemic Review and Meta-Analysis. PloS One 2016; 11: e0153655.

Izumi K. Disorders of Transcriptional Regulation: An Emerging Category of Multiple Malformation Syndromes. Mol. Syndromol. 2016; 7: 262–273.

Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants: Genet. Med. 2011; 13: 680–685.

Klein ME, Monday H, Jordan BA. Proteostasis and RNA Binding Proteins in Synaptic Plasticity and in the Pathogenesis of Neuropsychiatric Disorders. Neural Plast. 2016; 2016: 3857934.

Kramer JM, van Bokhoven H. Genetic and epigenetic defects in mental retardation. Int. J. Biochem. Cell Biol. 2009; 41: 96–107.

Krawitz PM, Schweiger MR, Rödelsperger C, Marcelis C, Kölsch U, Meisel C, et al. Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome. Nat. Genet. 2010; 42: 827–829.

Kumar R, Corbett MA, van Bon BWM, Woenig JA, Weir L, Douglas E, et al. THOC2 Mutations Implicate mRNA-Export Pathway in X-Linked Intellectual Disability. Am. J. Hum. Genet. 2015; 97: 302–310.

Küry S, Besnard T, Ebstein F, Khan TN, Gambin T, Douglas J, et al. De Novo Disruption of the Proteasome Regulatory Subunit PSMD12 Causes a Syndromic Neurodevelopmental Disorder. Am. J. Hum. Genet. 2017; 100: 352–363.

de Ligt J, Willemsen MH, van Bon BWM, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic exome sequencing in persons with severe intellectual disability. N. Engl. J. Med. 2012; 367: 1921–1929.

Lin RC, Scheller RH. Mechanisms of synaptic vesicle exocytosis. Annu. Rev. Cell Dev. Biol. 2000; 16: 19–49.

Linder B, Fischer U, Gehring NH. mRNA metabolism and neuronal disease. FEBS Lett. 2015; 589: 1598–1606.

Lopes F, Barbosa M, Ameur A, Soares G, de Sá J, Dias AI, et al. Identification of novel genetic causes of Rett syndrome-like phenotypes. J. Med. Genet. 2016; 53: 190–199.

Louros SR, Osterweil EK. Perturbed proteostasis in autism spectrum disorders. J. Neurochem. 2016; 139: 1081–1092.

Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DCY, Nazareth L, et al. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. N. Engl. J. Med. 2010; 362: 1181–1191.

Mardis ER. DNA sequencing technologies: 2006-2016. Nat. Protoc. 2017; 12: 213–218.

Marian AJ. Challenges in medical applications of whole exome/genome sequencing discoveries. Trends Cardiovasc. Med. 2012; 22: 219–223.

Martin CL, Kirkpatrick BE, Ledbetter DH. Copy number variants, aneuploidies, and human disease. Clin. Perinatol. 2015; 42: 227–242, vii.

Maulik PK, Mascarenhas MN, Mathers CD, Dua T, Saxena S. Prevalence of intellectual disability: a meta-analysis of population-based studies. Res. Dev. Disabil. 2011; 32: 419–436.

Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am. J. Hum. Genet. 2010; 86: 749–764.

Muzar Z, Lozano R, Kolevzon A, Hagerman RJ. The neurobiology of the Prader-Willi phenotype of fragile X syndrome. Intractable Rare Dis. Res. 2016; 5: 255–261.

Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, et al. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature 2011; 478: 57–63.

Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat. Genet. 2010; 42: 790–793.

Pang AW, MacDonald JR, Pinto D, Wei J, Rafiq MA, Conrad DF, et al. Towards a comprehensive structural variation map of an individual human genome. Genome Biol. 2010; 11: R52.

Parrini E, Conti V, Dobyns WB, Guerrini R. Genetic Basis of Brain Malformations. Mol. Syndromol. 2016; 7: 220–233.

Pengelly RJ, Greville-Heygate S, Schmidt S, Seaby EG, Jabalameli MR, Mehta SG, et al. Mutations specific to the Rac-GEF domain of TRIO cause intellectual disability and microcephaly. J. Med. Genet. 2016

Pennisi E. Breakthrough of the year. Human genetic variation. Science 2007; 318: 1842–1843.

Penzes P, Buonanno A, Passafaro M, Sala C, Sweet RA. Developmental vulnerability of synapses and circuits associated with neuropsychiatric disorders. J. Neurochem. 2013; 126: 165–182.

Pulvers JN, Journiac N, Arai Y, Nardelli J. MCPH1: a window into brain development and evolution. Front. Cell. Neurosci. 2015; 9: 92.

Quinlan-Jones E, Kilby MD, Greenfield S, Parker M, McMullan D, Hurles ME, et al. Prenatal whole exome sequencing: the views of clinicians, scientists, genetic counsellors and patient representatives. Prenat. Diagn. 2016; 36: 935–941.

Rauch A, Wieczorek D, Graf E, Wieland T, Endele S, Schwarzmayr T, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet 2012; 380: 1674–1682.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. Off. J. Am. Coll. Med. Genet. 2015; 17: 405–424.

Riggs ER, Ledbetter DH, Martin CL. Genomic Variation: Lessons Learned from Whole-Genome CNV Analysis. Curr. Genet. Med. Rep. 2014; 2: 146–150.

Rizzo JM, Buck MJ. Key principles and clinical applications of 'next-generation' DNA sequencing. Cancer Prev. Res. Phila. Pa 2012; 5: 887–900.

Ropers H-H. X-linked mental retardation: many genes for a complex disorder. Curr. Opin. Genet. Dev. 2006; 16: 260–269.

Salvador-Carulla L, Rodríguez-Blázquez C, Martorell A. Intellectual disability: an approach from the health sciences perspective. Salud Publica Mex. 2008; 50 Suppl 2: s142-150.

Sartor F, Anderson J, McCaig C, Miedzybrodzka Z, Müller B. Mutation of genes controlling mRNA metabolism and protein synthesis predisposes to neurodevelopmental disorders. Biochem. Soc. Trans. 2015; 43: 1259–1265.

Schmickel RD. Contiguous gene syndromes: a component of recognizable syndromes. J. Pediatr. 1986; 109: 231–241.

Sell GL, Margolis SS. From UBE3A to Angelman syndrome: a substrate perspective. Front. Neurosci. 2015; 9: 322.

Shaffer LG, Bejjani BA. A cytogeneticist's perspective on genomic microarrays. Hum. Reprod. Update 2004; 10: 221–226.

Sharer JD. An Overview of Biochemical Genetics. Curr. Protoc. Hum. Genet. 2016; 89: 17.1.1-17.1.16.

Shen X-M, Selcen D, Brengman J, Engel AG. Mutant SNAP25B causes myasthenia, cortical hyperexcitability, ataxia, and intellectual disability. Neurology 2014; 83: 2247–2255.

Shinawi M, Cheung SW. The array CGH and its clinical applications. Drug Discov. Today 2008; 13: 760–770.

Svitkina T, Lin W-H, Webb DJ, Yasuda R, Wayman GA, Van Aelst L, et al. Regulation of the postsynaptic cytoskeleton: roles in development, plasticity, and disorders. J. Neurosci. Off. J. Soc. Neurosci. 2010; 30: 14937–14942.

Tallafuss A, Constable JRL, Washbourne P. Organization of central synapses by adhesion molecules. Eur. J. Neurosci. 2010; 32: 198–206.

Timal S, Hoischen A, Lehle L, Adamowicz M, Huijben K, Sykut-Cegielska J, et al. Gene identification in the congenital disorders of glycosylation type I by whole-exome sequencing. Hum. Mol. Genet. 2012; 21: 4151–4161.

Torres F, Barbosa M, Maciel P. Recurrent copy number variations as risk factors for neurodevelopmental disorders: critical overview and analysis of clinical implications. J. Med. Genet. 2016; 53: 73–90.

Utine GE, Taşkıran EZ, Koşukcu C, Karaosmanoğlu B, Güleray N, Doğan ÖA, et al. HERC1 mutations in idiopathic intellectual disability. Eur. J. Med. Genet. 2017; 60: 279–283.

Vaillend C, Poirier R, Laroche S. Genes, plasticity and mental retardation. Behav. Brain Res. 2008; 192: 88–105.

Van Campenhout S, Devriendt K, Breckpot J, Frijns J-P, Peeters H, Van Buggenhout G, et al. Microduplication 22q11.2: a description of the clinical, developmental and behavioral characteristics during childhood. Genet. Couns. Geneva Switz. 2012; 23: 135–148.

Vermeesch JR, Brady PD, Sanlaville D, Kok K, Hastings RJ. Genome-wide arrays: quality criteria and platforms to be used in routine diagnostics. Hum. Mutat. 2012; 33: 906–915.

Vissers LELM, de Vries BBA, Veltman JA. Genomic microarrays in mental retardation: from copy number variation to gene, from research to diagnosis. J. Med. Genet. 2010; 47: 289–297.

Vlaskamp DRM, Rump P, Callenbach PMC, Vos YJ, Sikkema-Raddatz B, van Ravenswaaij-Arts CMA, et al. Haploinsufficiency of the STX1B gene is associated with myoclonic astatic epilepsy. Eur. J. Paediatr. Neurol. EJPN Off. J. Eur. Paediatr. Neurol. Soc. 2016; 20: 489–492.

Watson CT, Marques-Bonet T, Sharp AJ, Mefford HC. The genetics of microdeletion and microduplication syndromes: an update. Annu. Rev. Genomics Hum. Genet. 2014; 15: 215–244.

Xuan J, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: Promises and challenges. Cancer Lett. 2012

Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. J. Genet. Genomics Yi Chuan Xue Bao 2011; 38: 95–109.

Zhang J, Gambin T, Yuan B, Szafranski P, Rosenfeld JA, Balwi MA, et al. Haploinsufficiency of the E3 ubiquitin-protein ligase gene TRIP12 causes intellectual disability with or without autism spectrum disorders, speech delay, and dysmorphic features. Hum. Genet. 2017; 136: 377–386.

Zukin RS, Richter JD, Bagni C. Signals, synapses, and synthesis: how new proteins control plasticity. Front. Neural Circuits 2009; 3: 14.

DECIPHER v8.1 [Internet]. [cited 2014 Aug 5] Available from: https://decipher.sanger.ac.uk/

CHAPTER 2

Copy number variations as the cause of intellectual disability

Genomic imbalances in known and novel candidate *loci* in a group of Portuguese patients with idiopathic intellectual disability

Genomic imbalances in known and novel candidate *loci* in a group of Portuguese patients with idiopathic intellectual disability

Fátima Lopes*, Fátima Torres*, Gabriela Soares, Mafalda Barbosa, João Silva, Frederico Duque, Miguel Rocha, Joaquim de Sá, Guiomar Oliveira, Maria João Sá, Teresa Temudo, Susana Sousa, Carla Marques, Ana Sofia Lopes, Catarina Gomes, Gisela Barros, Felisbela Rocha, Arminda Jorge, Cecília Martins, Sandra Mesquita, Susana Loureiro, Elisa Maria Cardoso, Maria José Cálix, Andreia Dias, Cristina Martins, Céu R. Mota, Diana Antunes, Juliette Dupont, Sara Figueiredo, Sónia Figueiroa, Susana Gama Sousa, Sara Cruz, Adriana Sampaio, Erik A Sistermans, Bauke Yklstra, Paula Rendeiro, Purificação Tavares, Margarida Reis-Lima, Jorge Pinto-Basto, Ana Maria Fortuna, Patrícia Maciel

In preparation

*both authors contributed equally for this work

Abstract

Microarray comparative genomic hybridization is currently the standard method for screening for copy number variations in patients with intellectual disability. The aim of this work was to identify genetic causes of disease using this technique in a Portuguese cohort of 325 patients with intellectual disability studied in two different contexts: clinical and research. The yield obtained was 8.5% in the research and 11.7% in the clinical cohort for pathogenic alterations. Likely pathogenic variants were detected in 5.3% of the research cases and in 5.8% of the clinical ones. As for variants of unclear significance, 14.9% were detected in the research cohort while 14.6% were present in the clinical cases. These numbers are in accordance with those described for similar cohorts. Relevant likely pathogenic structural variations were present for a total of 18 cases (5.5%), namely in 1p22.1-p21.3, 2p15, 2q11.2-q12.2, 7q33, 9q33.2-q33.3, 10q26.3, 12p13.33, 17p11.2, 20q13.12-q13.13, Xq24 and Xq26.3. Additionally, for several cases carrying copy number gains the expression of genes located inside, at the breakpoint or close to the breakpoint region was analyzed. For *EHMT1*, *TSC1*, *INF2*, *FBXW2*, *NEK6*, *PSMB7* and CUL4B genes their expression in increased. Only for *TECPR2*, *TSPAN9* and *LAMP2* genes there was no change in expression. None of the genes evaluated presented reduction in expression. Globally we identified 13 new candidate *loci* for ID, pinpointed the most likely relevant genes within these *loci*, and established genotype-phenotype correlations.

Introduction

Intellectual disability (ID) is one of the most common neurodevelopmental disorders (NDs) in all populations worldwide. It affects nearly 1-2% of the general population, being characterized by a decrease in the cognitive abilities, language and social skills (American Psychiatric Association., 2013).

A substantial number of ID patients are predicted to have cytogenetic imbalances, resulting from deletions or duplications in very small segments of chromosomes. Most of these Copy Number Variations (CNVs) are large (>400 kb), typically involving dozens of genes, are individually rare (frequency <0.1%) (Morrow, 2010), and are usually associated with a broad phenotypic spectrum, ranging from normal development, to cognitive impairment (Coe *et al.*, 2012).

Relatively recent approaches to the identification of CNVs have highlighted the relevance of rare *de novo* and essentially private mutations that contribute to a significant proportion of risk of developing NDs, and these types of molecular changes are presently an unavoidable element in the field of both Neuropsychiatry and Neuropediatrics (Coe *et al.*, 2012; Malhotra and Sebat, 2012).

The frequency of detection of chromosome abnormalities and/or genomic rearrangements in patients with NDDs using aCGH mainly depends on the clinical patient inclusion criteria and on the microarray design (Miller *et al.*, 2010). Nevertheless, detection rates are usually much higher in patients with ID/developmental delay (DD) that also present malformations or dysmorphic features and more severe cognitive impairment (Coe *et al.*, 2012; Kirov *et al.*, 2014). The characterization of these genomic variants in different patient cohorts as well as in the general population is necessary to clarify their clinical consequence and establish adequate genotype-phenotype correlations (Mannik *et al.*, 2015).

Herein, we present the results obtained by studying 325 Portuguese patients with ID using aCGH, in whom we found known pathogenic variants and new candidate pathogenic CNVs. As expected, the great majority of the detected variants were rare and restricted to one patient/family; nevertheless, the efforts towards their characterization represent a step forward in order to clarify their clinical significance.

Methods

Patients, inclusion criteria, and clinical characterization

This work included the analysis of 325 ID patients of Portuguese origin, of which 188 were included in a research cohort (RC) and 137 were studied in the context of clinical genetics diagnosis (clinical cohort, CC). For the inclusion of patients in the RC, the patients needed to have (I) documented developmental delay/ ID (IQ test equal/below 70 for patients with more than 3 years or on basis of clinical evaluation by a pediatrician); (II) dysfunction/impairment in more than 2 areas of communication, self-care, home living, social/interpersonal skills, use of community resources, self-direction, functional academic skills, work, leisure and safety; (III) unknown aetiology, in spite of standard aetiological investigation; (IV) onset of ID during childhood; (V) previous normal investigations including biochemistry workup, high-resolution karyotype, fragile X testing and FISH studies when clinically indicated, *ATRX* analysis, and pregnancy TORCH serologies if available. Patients with large genomic imbalances detectable by G-banded karyotyping, common environmental etiologies and common genetic etiologies were generally excluded. All the patients in this group were evaluated by a multidisciplinary team, which included a pediatrician and/or a neuropediatrician, a medical geneticist and a psychologist.

Collection of parents/relatives' blood samples and Ethics approval and informed consent

The enrollment of the patients and families was done by the referring doctor, clinical information was gathered in an anonymous database and written informed consent was obtained for all participants according to the Portuguese Data Protection Authority (CNPD). This study was approved by the ethics committee of the Center for Medical Genetics Dr Jacinto Magalhães, National Health Institute Dr Ricardo Jorge.

Genetic analysis

aCGH was performed using the following platforms: Agilent 180K (AMADID:023363; 180.000 in situ synthesized 60-mer oligonucleotide probes, mean resolution of 17Kb); KaryoArray®v3.0 (Agilent 8x60k) (probes distributed throughout the genome with an average resolution of 9Kb in 357 regions associated with microdeletion/microduplication syndromes, telomeres and centromeres, and with an average resolution of 175Kb in the backbone); Agilent Whole Genome 244K (240.000 markers distributed throughout the genome, with an average resolution of 9Kb);

Affymetrix CytoScan HD (probes distributed throughout the genome, with an average resolution of 20Kb) or CytoScan 750K (750.000 markers distributed throughout the genome, with a medium resolution of 8-20Kb). A diploid DNA without variations was used as a reference: for the Agilent 180K (Kreatech´s MegaPoll Reference DNA, Kreatech Diagnostics, Amsterdam); for the Affymetrix platforms: diploid genomic DNA provided with the CytoScan® Array Kit ; genomic coordinates are according to Human Genome Build hg19; analysis was performed using the appropriate software of each platform: Agilent 180K, Nexus Copy Number 6.0 software with FASST2 Segmentation algorithm (BioDiscovery Inc, El Segundo, CA); KaryoArray®v3.0 (Agilent 8x60k) and Agilent Whole Genome 244K, Aberration Detection Method 2 (ADM-2); Affymetrix CytoScan HD and CytoScan 750k, Analysis Suite (ChAS 3.0) software (Affymetrix).

Classification of the variants

The genomic variants detected were classified using criteria described elsewhere (Lee *et al.*, 2007; Leung *et al.*, 2010; Miller *et al.*, 2010) as: pathogenic (CNVs that overlap critical regions of known microdeletion or microduplication syndromes or overlap other genomic regions that are defined as clinically significant, such as subtelomeric regions); very likely pathogenic (gene rich large CNVs, containing morbid OMIM genes, that overlap a genomic imbalance in a CNV database for affected individuals, for example, DECIPHER); variants of unknown clinical significance (VOUS), (expanded or altered CNV inherited from a parent but not within a known region of CNVs or CNVs that carry genes that exhibit haploinsufficiency); or benign (CNV in a database of healthy individuals). For terminology simplification throughout the text and in the tables the term CNV is used for pathogenic, likely pathogenic variants and VOUS, whereas polymorphic CNVs (present in more than 3 controls in public databases, such as Human Genome Database Variants) were generally not considered for further analysis.

Quantitative PCR (qPCR) confirmations

For CNV confirmation we performed qPCR, using *SDC4* (ENSG00000124145, Chr.20) and *ZNF80* (ENSG00000174255, Chr.3) as references genes (detailed description in Supplementary data). Quantitative PCR reactions were carried out in a 7500-FAST Real Time PCR machine (Applied Biosystems, Foster City, CA, USA) using Power SYBR Green® (Applied Biosystems, Foster City, CA, USA), as described elsewhere and following the general recommendations for qPCR (Hoebeeck *et al.*, 2005; D'haene *et al.*, 2010; Svec *et al.*, 2015). The specificity of each of

the reactions was verified by the generation of a melting curve for each of the amplified fragments. The primer efficiency was calculated by the generation of a standard curve fitting the accepted normal efficiency percentage (primers used listed in supplementary data). Ct values obtained for each test were analyzed in DataAssist™ software (Applied Biosystems, Foster City, CA, USA).

mRNA expression analysis

Total RNA was isolated from leucocytes of patients and controls using QIAsymphony RNA Kit (QIAGEN GmbH, Germany), according to the manufacturer's protocol. First-strand cDNA, synthesized using SuperScript® III Reverse Transcriptase (RT) (Thermo Scientific, Waltham, MA USA). The genes selected to study within each CNVs were selected based on their localization in the alteration (either in breakpoint or close by), their functional relevance and their predicted expression values in the periphery [data retrieved from GeneCards database (www.genecards.org) and GTExportal (http://www.gtexportal.org)]. The comparisons were carried out using a pool of four adult male controls and a pool of four adult females, making a total of eight healthy individuals used for comparison with the patients. Primers used for all genes are listed on supplementary data to this chapter. Quantitative PCR reactions were carried out in a 7500-FAST Real Time PCR machine (Applied Biosystems, Foster City, CA, USA) using Power SYBR Green® (Applied Biosystems, Foster City, CA, USA). The expression levels of the genes were normalized to the *B2M* and *PPIB* genes and relative quantification was used to determine the fold change difference between each gene and each reference gene, using the DDCT method, as described elsewhere and following the general recommendations (Pfaffl *et al.*, 2001; Taylor *et al.*, 2010).

Results

Clinical overview

Global data

The total sample studied here included 325 Portuguese patients with idiopathic ID/DD studied by aCGH, of which 188 were from the RC. Due to the different origin of the two cohorts (clinical and research based) a more extensive clinical description was only possible to obtain for the RC. Considering the total group sample of 188 Portuguese patients with idiopathic ID/DD studied by

aCGH. The RC included a total of 188 patients (66 females and 122 males). 106 (56%) of the patients were globally classified by the medical geneticist or following practitioner as having syndromic ID and 82 (44%) were classified as having non-syndromic ID. 108 patients had mild ID, 39 moderate, 17 severe, 9 profound and 9 borderline ID. The majority of the patients had a familial history of ID (n=112, 60%) (figure 1).

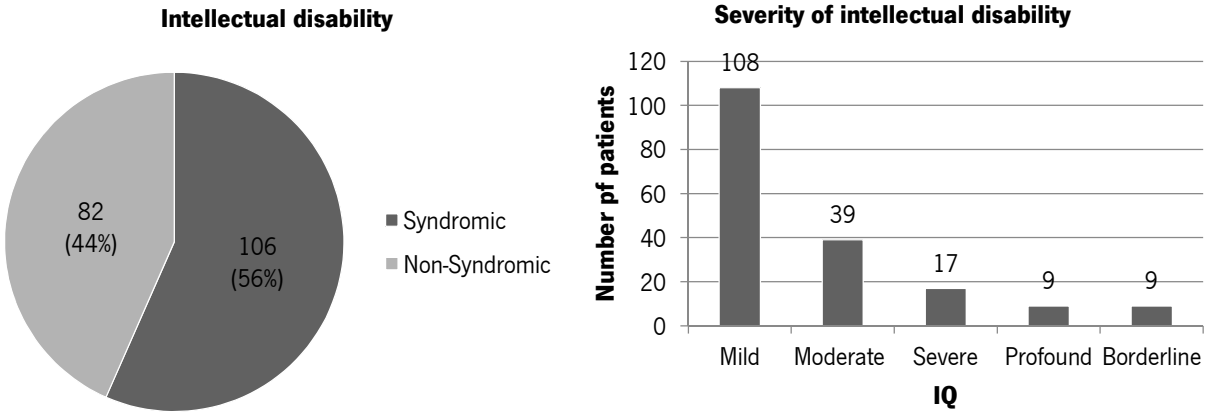


Figure 7 – Clinical characteristics of the RC patients studied by aCGH. On the left the ratio of syndromic vs non-syndromic cases and on the right the patients' distribution based on the IQ level.

A total of 90 patients presented ID together with some form of congenital anomay (in detail, 24 with cardiac malformation, 62 with a brain malformation, 34 with urogenital malformations and 12 with other type of malformation) (figure 2). The majority of the patients had facial dysmorphisms (64%) (figure 2).

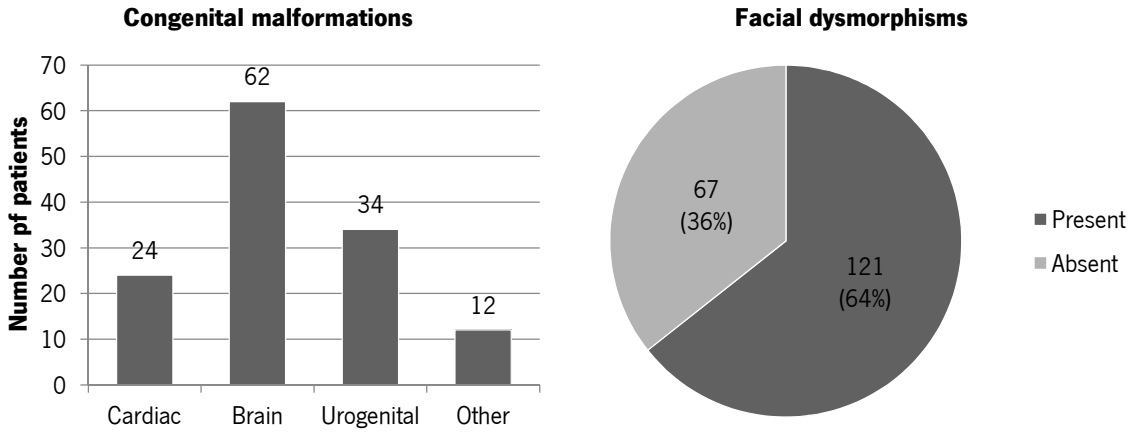


Figure 8 – Distribution of congenital anomalies present in the RC patients (left) and facial dysmorphisms (right).

25 patients had epilepsy, 14 had macrocephaly and 34 had microcephaly (figure 3).

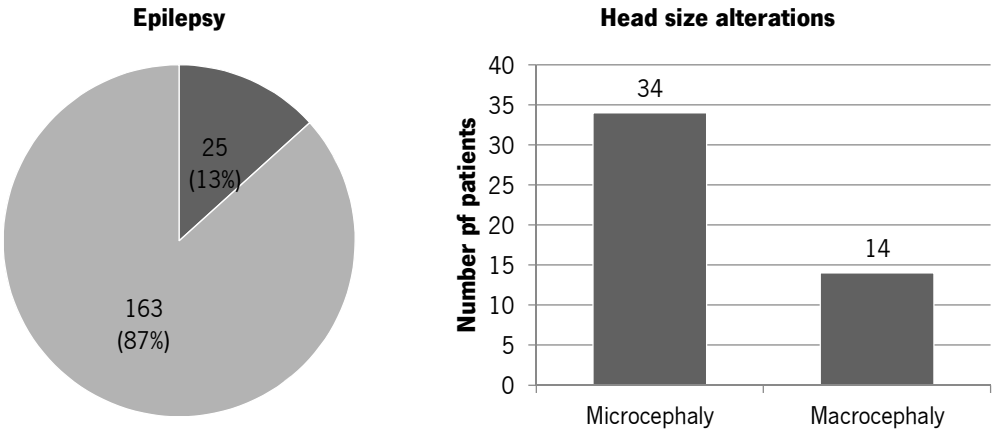


Figure 9 – Number of RC patients with epilepsy (left) and with brain size alterations (microcephaly and macrocephaly, on the right).

The majority of the patients showed behavioral alterations (n=116, 62%), the most common being attention deficit and hyperactivity disorder (ADHD) (n=75). Aggressiveness and autism were also present features in the cohort (n=33 and n=14 respectively). A significant number of patients presented other types of behavioral abnormalities such as stereotypies, sleep disturbances, reduced activity, etc (n=49) (figure 4).

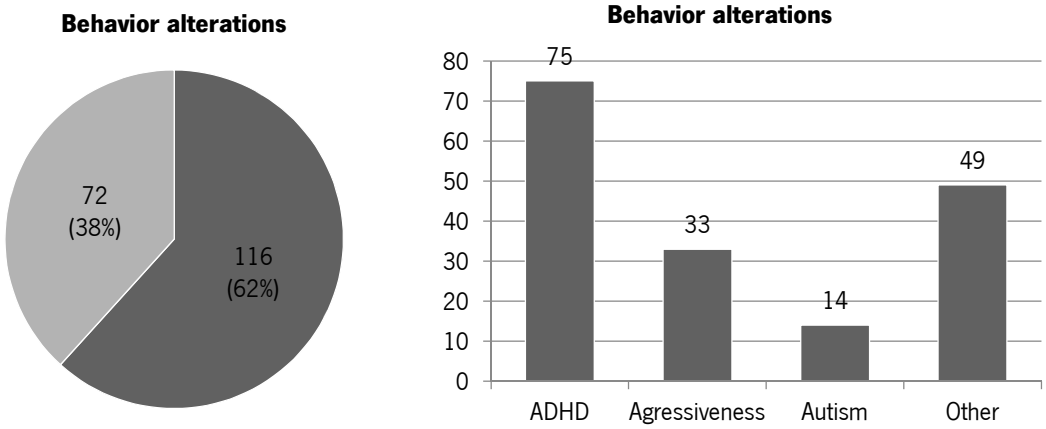


Figure 10 – Behavioral alterations observed in RC patients (total percentage on the left and sorted by category on the right).

Patients with non-polymorphic CNVs

The RC included a total of 16 females and 38 males in which a non-polymorphic alteration was found. Within the 54 patients with CNVs, 24 had non-syndromic ID while 30 were syndromic, 37 patients had mild ID, 9 moderate, 4 severe, 3 profound and 1 with borderline IQ (figure 5).

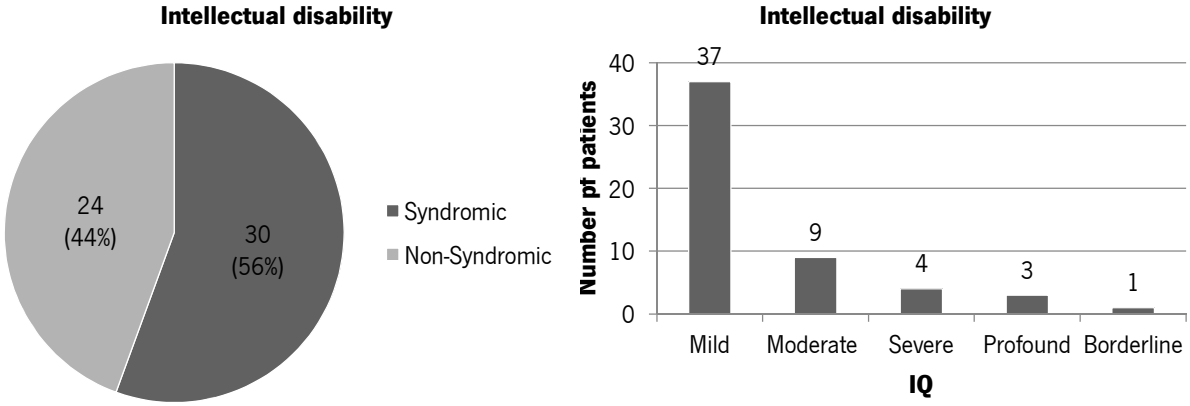


Figure 11 - Clinical characteristics of the RC patients with non-polymorphic CNVs detected (from the total of 54 patients). The majority of the cases with non-polymorphic CNVs had syndromic ID (left), the milder form of ID being the most common among the cases (right).

A total of 34 patients had a known familial history of ID; congenital malformations were present in 23 patients, which included brain malformations in 17 patients, cardiac malformations in 6 and urogenital malformations in 4. 57% (n=31) of the cases showed facial dysmorphisms of some type (figure 6).

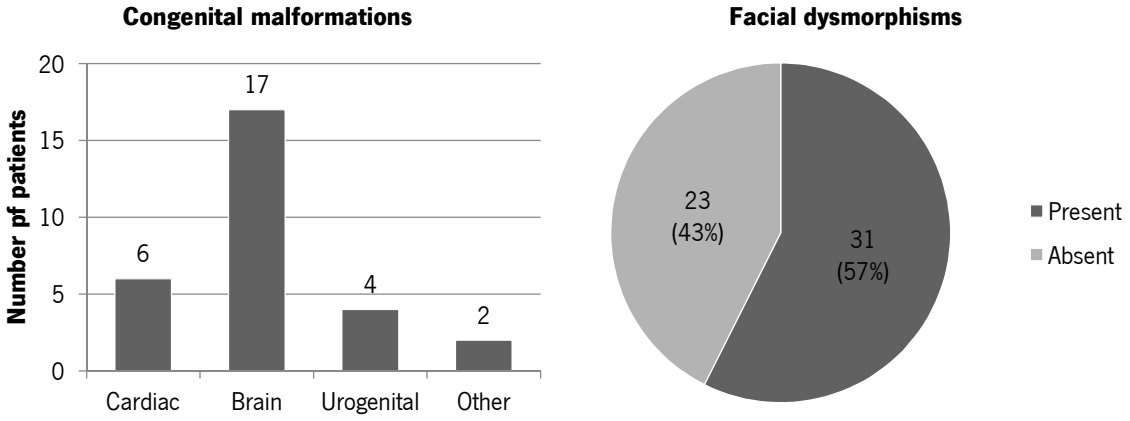


Figure 12 - Distribution of congenital anomalies (left) and facial dysmorphisms (right) in the cases with non-polymorphic CNVs.

Seven of the 54 RC patients with non-polymorphic CNVs had epilepsy, 11 had microcephaly and only 1 had macrocephaly (figure 7).

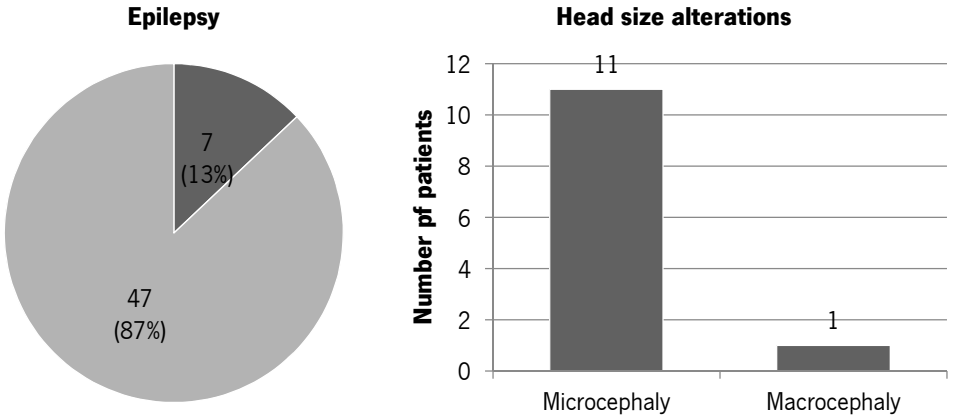


Figure 13 – Occurrence of epilepsy (left) and head size alterations (microcephaly and macrocephaly, on the right) in the RC cases with non-polymorphic CNVs.

Behavioral alterations were a recurrent finding in these patients, being present in 35 patients (65%), ADHD being the most common (figure 8).

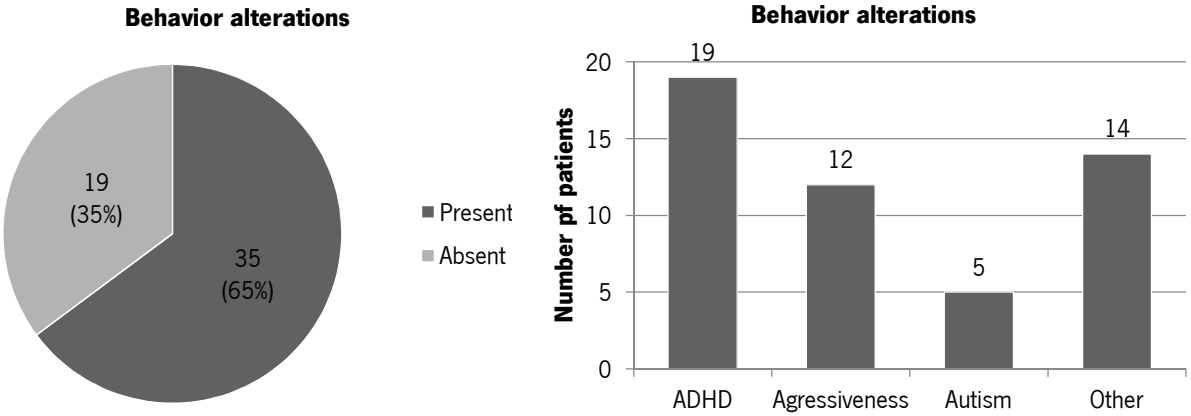


Figure 14 - Behavioral alterations present in the RC cases with non-polymorphic (total percentage on the left and sorted by category on the right).

A detailed comparison of clinical features present in the entire cohort versus only the patients with non-polymorphic CNVs is presented in table I.

Table IV – Side by side comparison of the clinical features present in the entire RC vs only the patients with non-polymorphic CNVs

Entire research cohort (n=188)	**Patient with non-polymorphic CNVs (n=54)**
Gender	**Gender**
Males **122 (65%)**	Males **38 (70%)**
Females **66 (35%)**	Females **16 (30%)**
Intellectual disability	**Intellectual disability**
Syndromic **106 (56%)**	Syndromic **30 (56%)**
Non-syndromic **82 (44%)**	Non-syndromic **24 (44%)**
Mild **113 (60%)**	Mild **37 (68%)**
Moderate **39 (21%)**	Moderate **9 (17%)**
Severe **17 (9%)**	Severe **4 (7%)**
Profound **10 (5%)**	Profound **3 (6%)**
Borderline **9 (5%)**	Borderline **1 (2%)**
History of ID	**History of ID**
Sporadic **76 (40%)**	Sporadic **20 (37%)**
Co-morbidities	**Co-morbidities**
Congenital anomalies **90 (48%)**	Congenital anomalies **23 (43%)**
Epilepsy **25 (13%)**	Epilepsy **7 (13%)**
Microcephaly **34 (18%)**	Microcephaly **11 (20%)**
Macrocephaly **14 (7%)**	Macrocephaly **1 (2%)**

Genomic imbalances: overview

From the total group of patients 30.2% had CNVs detected by aCGH (table II).

For the research cohort, 28.7% (n=54) of patients had CNVs detected by aCGH (table II): 8.5% (n=16) had known pathogenic variants, 5.3% (n=10) had likely pathogenic variants and 14.9% (n=28) had variants of unknown clinical significance (VOUS). The remaining 71.3% of patients (n=134) had normal aCGH results (including known polymorphic variants).

For the clinical cohort, 32.1% (n=44) of patients had CNVs detected by aCGH: 11.7% (n=16) had known pathogenic variants, 5.8% (n=8) had likely pathogenic variants and 14.6% (n=20) had

variants of unknown clinical significance (VOUS). The remaining 69.8% of patients (n=93) had normal aCGH results (including known polymorphic variants).

Table V – Number of non-polymorphic CNVs sorted by category.

	Research cohort (RC)	Clinical cohort (CC)	Both cohorts
Number of patients	188	137	325
Known pathogenic	16 (8.5%)	16 (11.7%)	32 (9.8%)
Likely pathogenic	10 (5.3%)	8 (5.8%)	18 (5.5%)
VOUS	28 (14.9%)	20 (14.6%)	48 (14.7%)
Total with CNVs	54 (28.7%)	44 (32.1%)	98 (30.2%)

Known pathogenic variants

The known pathogenic variants detected were mainly *de novo* CNVs associated with the following: a) deletion syndromes: 1p36.23-p36.21, 1q43-q44, Coffin-Siris, Cri-du-Chat, 7q11.23, 8p23.1, 9p13.1-p11.2, Jacobsen, 12q24.21-q24.22, 16p11.2, 17q21.31, 22q11 and 22q13.3, Xp22.33-p22.31; b) duplication syndromes: 9q34.13-q34.3, 12q24.21, 13q12.12-q34, 15q11.2-q13.1, 21q11.2-q22.11 and *MECP2* duplication syndrome. For all these syndromes there are reports in the literature describing the phenotypic and genetic findings for similar patients, therefore only some particular cases will be discussed here. We also included in this category alterations occurring in risk associated *loci*: 1q21.1 and 16p13.3.

Table VI – List of pathogenic CNVs detected in the patients.

Cohort	ISCN description (Hg19)	Type	Size (Mb)	Genes	Key gene(s) involved	Associated syndrome	Phenotype overlap	Inheritance	Array platform
R1 (♂)	1p36.23-p36.21(8,593,674-15,396,672)X1dn	del	6.8	86	*ANGPTL7, CASZ1, MAD2L2, RER*	-	-	*de novo*	1
R2 (♀)	1q43-q44(243,552,007-243,738,675)x1dn	del	0.19	2	*AKT3*	1q43-q44 deletion syndrome	Yes	*de novo*	1
C1 (♂)	1q43q44(243,592,147-243,749,968)x1	del	0.16	2	*AKT3*	1q43-q44 deletion syndrome	Yes	*ND*	2
C2 (♂)	1q43-q44(240,043,427-249,233,096)x1dn	del	3.7	18	*AKT3*	1q43-q44 deletion syndrome	Yes	*de novo*	2
R3 (♂)	2p13.1-p13.3(70,894,906-74,986,518)x1dn[v]	del	4	62	*CYP26B1, EXOC6B*	-	-	*de novo*	1
R4 (♂)	3q22.1-q23(131,415,639-141,618,552)x1dn	del	1	65	*FOXL2*	BPES	Yes (eye features)	*de novo*	1
C3 (♂)	5p15.33-p15.32(204,849-5,014,883)x1	del	4.81	30	*TERT [CTNND2 not involved]*	Cri-du-Chat (atypical)	No	ND	3
R5 (♂)	6q25.3(156,012,754-158,804,494)X1dn	del	2.6	14	*ARID1B*	Coffin-Siris syndrome	Yes	*de novo*	1
R6 (♂)	7q11.23(72,741,861-73,145,916)x1mat[b]	del	0.4	11	*BAZ1B, STX1A, WBSCR22*	-	No	maternal	1
C4 (♂)	7q11.23(72,721,760-74,140,846)x1	del	1.419	28	*BAZ1B, STX1A, WBSCR22, ELN*	-	No	ND	3
R7 (♀)	8p23.1(7,039,276-12,485,558)x1dn	del	5.5	70	SOX7, *GATA4*	8p23.1 deletion syndrome	Yes (cardiac)	*de novo*	1
C5 (♂)	11q24.2-q25(125,232,584-134,446,160)x1dn	del	9.214	54	*KIRREL3, ETS1, FLI1, KCNJ1, KCNJ5, RICS*	Jacobsen syndrome	Yes	*de novo*	3
R8 (♀)	12q24.21-q24.22(115,505,500-117,441,683)X1dn[v]	del	0.2	10	*MED13L*	-	Yes	*de novo*	1
C6 (♂)	16p11.2(29,674,336-30198,123)x1dn	del	0.524	29	*KCTD13*	16p11.2 deletion syndrome	ND	*de novo*	3
C7 (♂)	17q21.31(43,710,371-44,215,352)x1	del	0.505	8	*CRHR1, MAPT, STH, and part of the KIAA1267 (KANSL1)*	17q21.31 deletion syndrome (Koolen-De Vries syndrome)	ND	ND	4
C8 (♂)	22q11.21(18,894,835-21,505,417)x1	del	2.611	59	*TBX1*	22q11 deletion syndrome	ND	ND	3
C9 (♂)	22q13.3(49,513,903-51,178,264)x1	del	1.664	39	*SHANK3*	22q13.3 deletion syndrome (Phelan-McDermid syndrome)	ND	ND	3

(cont.)

Cohort	ISCN description (Hg19)	Type	Size (Mb)	Genes	Key gene(s) involved	Associated syndrome	Phenotype overlap	Inheritance	Array platform
C10 (♂)	1q21.1q21.2(146,106,723-147,830,830)x3	dup	1.7	17	*HYDIN2, PRKAB2*	susceptibility locus for neurodevelop disorders	Yes	*de novo*	2
R9 (♂)	1q21.1(145,883,119-148,828,690)X3[v]	dup	2.5	23	*HYDIN2, PRKAB2, GJA5*	susceptibility locus for neurodevelop disorders	Yes	paternal	1
C11 (♀)	arr 9q34.3(140540819-140659057)x3	dup	0.118	2	*EHMT1*	-	Partially	maternal	2
R10 (♂)	12q24.21(116,408,736-116,704,303)X3 dn[v]	dup	0.3	2	*MED13L*	-	Yes	*de novo*	1
C12 (♂)	13q12.12-q34(23,749,431-115,083,342)x2.15*	dup	91.33	##	-	Trisomy 13 (mosaicism)	ND	ND	3
C13 (♀)	15q11.2-q13.1(22880274-29331964)x3mat	dup	6.45	111	*CYF1P1, NIPA2, NIPA1, MKRN3, NDN, MAGEL2, SNURF/SNRPN, UBE3A GABRB3*	15q11-q13 duplication syndrome	Yes	maternal	3
C14 (♀)	16p13.11(15,034,010-16,199,882)x3	dup	1.166	11	*NDE1*	16p13.11 duplication syndrome	ND	ND	5
R11 (♂)	16p13.11(15,421,671-16,443,968)x3	dup	1	19	*NDE1*	16p13.11 duplication syndrome	Yes	maternal	1
R12 (♂)	16p13.11(15,484,180-16,308,344)x3	dup	0.8	9	*NDE1*	16p13.11 duplication syndrome	Yes	maternal	1
C15 (♂)	21q11.2-q22.11(14,417,523-34,894,625)x3	dup	20.47	110	*DSCR1, DSCR2, DSCR3, DSCR4, APP*	-	No	ND	3
R13 (♂)	Xp11.22(53,569,653-53,769,748)X2mat	dup	0.2	3	*HUWE1*	-	Yes	maternal	1
C16 (♀)	Xp22.31(6,440,776-8,135,568)x3	dup	1.695	7	*VCX3A, HDHD1, STS, VCX, PNPLA4*	Xp22.31 duplication	ND	ND	2
R14 (♂)	Xq28(152,348,378-155,228,013)x2dn	dup	2.8	78	*MECP2*	MECP2 duplication syndrome	Yes	*de novo*	1
R15 (♂)	Xq28(153,130,545-153,602,293)x2mat	dup	0.5	16	*MECP2*	MECP2 duplication syndrome	Yes	maternal	1

(cont.)

Cohort	ISCN description (Hg19)	Type	Size (Mb)	Genes	Key gene(s) involved	Associated syndrome	Phenotype overlap	Inheritance	Array platform
R16 (♂)	9q34.13-q34.3(135,767,911-141,153,431)X3dn	dup	5.516	135	*EHMT1, RXRA, GRIN1, UAP1L1*	9q34 duplication syndrome	Partially	*de novo*	1
	14q32.31-q32.33(102,959,110-104,578,612)X3dn	dup	1.620	22	*MARK3, KLC1, EIF5*	-	-	*de novo*	1
	14q32.33(105,104,831-106,531,339)X3dn	dup	1.427	24		-	-	*de novo*	1

Legend: Patients R1 to R14: from research cohort; Patients C1 to C17: from clinical cohort; NP: not performed; ND: not determined; (*): mosaicism; (‡): inherited from an affected parent; (¥): Published in detail elsewhere. Array platform 1: Agilent 180K; 2: Affymetrix Cystoscan 750K; 3: KaryoArray®v3.0 (Agilent 8x60k); 4: Affymetrix CytoScan HD array; 5: Agilent Whole Genome 244K

1p36.23-p36.21 deletions

Patient R1 (figure 9) has a 6.7 Mb *de novo* deletion at 1p36.23-p36.21. This patient is an adult male (30 years old) with moderate ID (IQ= 49), microcephaly, broad nasal bridge, hypoplastic nares, microretrognathia, kyphosis, hypertelorism and telecanthus. Large terminal deletions in 1p36 are known to cause the 1p36 deletion syndrome (Jordan *et al.*, 2015) however, interstitial deletions in this region are quite rarely described associated with NDDs.

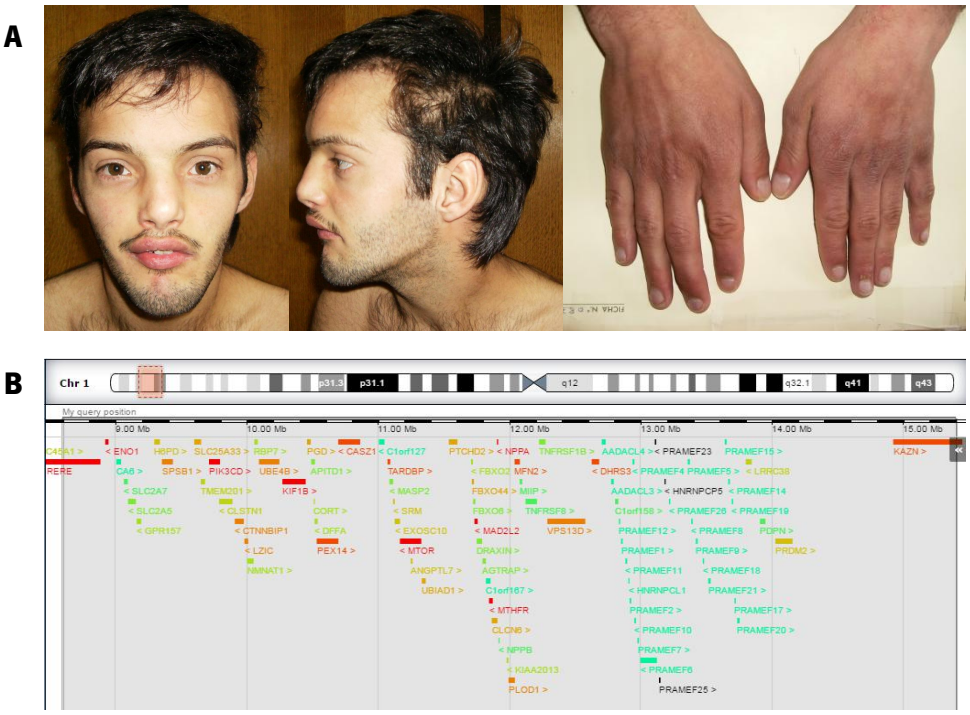


Figure 15 - Facial appearance and hands (A) and deleted 1p36.23-p36.21 region (B).

The deletion present in this patient affects 126 genes among which four (*PEX14, PLOD1, NMNAT1 and MTHFR)* previously been associated with disease ('The Deciphering Developmental Disorders (DDD) Study - DECIPHER v9.12', n.d.) and eight *MTOR*, *ENO1*, *PIK3CD, RERE, NPPA, MAD2L2, MTHFR, KIF1B* have a high haploinsufficiency score (Huang *et al.*, 2010), are possibly contributing for the phenotype. In the Decipher database there are patients with similar overlapping deletions, of which three cases (Decipher cases 266689, 248448 and 251601) also carry a *de novo* deletion and share similar clinical features, namely ID, MIC (for Decipher 251601) and facial dysmorphisms.

MTHFR (Methylenetetrahydrofolate Reductase (NAD(P)H)) encodes an enzyme of the folate metabolism responsible for the conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate which is essential for homocysteine remethylation to methionine and synthesis of S-adenosylmethionine. (Chen *et al.*, 2005) Recessive mutations in *MTHFR* cause

homocystinuria due to methylenetetrahydrofolate reductase deficiency, an error of folate metabolism, that can present itself as a severe neurologic deterioration with early death to asymptomatic in adulthood (Rosenblatt *et al.*, 1992). Additionally, compound heterozygous mutations in *MTHFR* were also described in a patient with infantile epileptic encephalopathy (Prasad *et al.*, 2011). Because the patient doesn't present epilepsy and only recessive mutations in *MTHFR* gene are associated with disease is possible that this is not contributing very strongly for the phenotype. However, during the years many studies have implicated *MTHFR* polymorphisms as a susceptibility factor for neurodevelopmental disorders, schizophrenia, neural tube defects and Down syndrome, although many times associated with controversy and contradictory results (Kuzman and Muller, 2012; Wu *et al.*, 2013; Victorino *et al.*, 2014; Tsang *et al.*, 2015).

MTOR (Mechanistic Target Of Rapamycin (Serine/Threonine Kinase)) was described as mutated in patients with Smith-Kingsmore (SK) syndrome, characterized by ID, macrocephaly, dysmorphisms and small thoraces (Baynam *et al.*, 2015). *MTOR* mutations were also described in a patient with epileptic encephalopathy without brain malformation (Allen *et al.*, 2013). Besides ID the patient doesn't present any major similarities with SK patients (apart from the alterations in head size which seems to be opposite between them). Nevertheless, *MTOR* is a key gene in a pathway in which several upstream and downstream components are known to play important roles in synaptic plasticity, and are associated with several NDD disorders (for instance fragile X syndrome, Down syndrome, Tuberous sclerosis and autism) (Hoeffer and Klann, 2010; Costa-Mattioli and Monteggia, 2013), hence it is a plausible contributor to the neurodevelopment phenotype of this patient).

RERE (Arginine-Glutamic Acid Dipeptide (RE) Repeats) encodes a protein that positively regulates retinoic acid signaling and that, when mutated in mice, leads to somite asymmetry (Vilhais-Neto *et al.*, 2010). Also, in a mouse model with both a null and a hypomorphic *Rere* allele several anomalies were described: microphtalmia, postnatal growth retardation, brain hypoplasia, decreased number of NeuN-positive neurons in the hippocampus, cardiovascular malformations, hearing loss and renal agenesis (Kim *et al.*, 2013). More recently, 10 patients with *de novo* dominant mutations in *RERE* gene with NDD and ID were described for the first time, being suggestive that haploinsufficiency of *RERE* is enough to cause ID related phenotypes (Fregeau *et al.*, 2016). Taking this into account, and similarly to what is thought to happen with the 1p36 terminal deletion syndrome patients, the haploinsufficiency of *RERE* is a good candidate to

contribute (alone or in conjunction with the haploinsufficiency of other surrounding genes) for the growth retardation, microcephaly and ID observed in the patient.

3q22.1-q23 deletions (BPES)

Patient R4 (figure 10) is a 15 years old boy with mild ID (IQ= 54), brachycephaly, cortical atrophy, low weight and hypoplastic genitalia with hypospadias. He also presents a peculiar eye dysmorphy characterized by microphtalmia, epicanthus and ptosis. Interestingly, he carries a 10.2 Mb *de novo* deletion at 3q22.1-q23 affecting the *FOXL2* (Forkhead Box L2) gene, for which mutations are causative of blepharophimosis syndrome (BPES, MIM 110100) (Verdin and Baere, 2012). In the recent years it has been shown that large deletions affecting *FOXL2* and its surroundings are often associated with ID combined with BPES (D'haene *et al.*, 2009; Zahanova *et al.*, 2012) and this patients reinforces the association of deletions in this region with that phenotype.

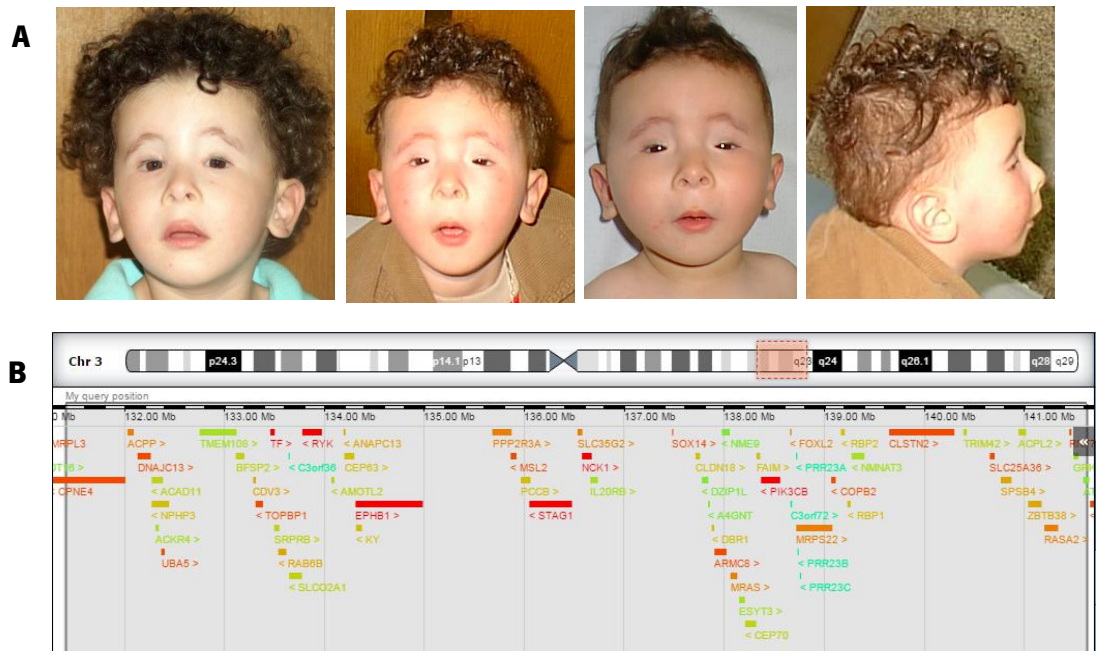


Figure 16 - Facial appearance at different ages (A) and deleted 3q22.1-q23 region (B).

7q11.23 deletions

In this work we detected two non-related patients with 7q11.23 deletions.

One of these is patient R6 (figure 11A), a 19 years old boy with severe ID (IQ=21) whose mother is suspected to have mild ID (carrier). He has cerebral atrophy, microcephaly, epilepsy and recurrent respiratory infections. Behavioral alterations include motor stereotypies, increased

activity, sleep disturbances and aggressiveness. He carries a 0.4 Mb maternal deletion at 7q11.23 (inside the Williams-Beuren syndrome (WBS, MIM 612547) critical region) affecting the *BAZ1B*, *STX1A*, *WBSCR22*, genes among others. CNVs at 7q11.23 are described in patients with ID/DD associated with the involvement of several systems and have variable expressivity (Merla *et al.*, 2010). Overall, this patient seems to have a more severe ID than the one expected in a classical WBS patient, does not present hypersociability (in fact he is aggressive), has microcephaly and cortical atrophy. Additionally, he doesn´t have congenital heart problems, which can be due to the fact that the *ELN* (elastin) gene is not affected in the patient (Delio *et al.*, 2013). A neurocognitive assessment of the patient was performed as described previous for WBS patients (Wechsler, 1991; Simões *et al.*, n.d.). Overall, in spite of several attempts, the patient was unable to perform any of the subtests.

The overall neurocognitive assessment documents severe ID with concurrent deficits in adaptive functioning, matching a profound intellectual disability classification (Full Scale Intelligence Quotient – FSIQ – <20), with a severe impairment observed in the Verbal comprehension, Performance and Working memory. In fact, no proficiency of the verbal over non-verbal measures was observed; specifically, a severe impairment in language production, comprehension and verbal concept formation, were observed in parallel with deficits in the ability to organize and visually perceive information, visual-motor coordination, learning and non-verbal concept formation. Furthermore, severe difficulties in attention, auditory and visual short term memory abilities were documented. Behaviorally, the patient R6 displayed motor stereotypies (fondle or asking to be fondled) and frequent verbal preservative repetitions (saying “hello” and “caress”). He was unable to understand the questions or requirement of several tasks, which compromised the ability to successfully perform in the neuropsychological assessment.

The other finding at 7q11.23 region is in patient C4 (figure 11B). He is a 6 years old boy referred to the consultation due to DD, with previous normal investigations including high-resolution karyotype and MLPA analysis of subtelomeric regions. He was evaluated at 5 years and 7 months with Griffiths Mental Developmental Scales (GMDS) as having a global development quotient (DQ) of 58. In addition to ataxia, motor and language delay, he presents dysmorphic features that include flat nose, thin upper lip, narrower central incisors, big ears and hyperacusis, narrow girdles, mainly scapular, and gross fingers. An echocardiography did not reveal any anomaly; presently, he is underweight and does not present behavioral problems. Patient C4 has a 1.419Mb deletion at 7q11.23 that encompasses 27 genes, including the *ELN, GTF2IRD1*

(GTF2I repeat domain containing 1), *GTF2I* (general transcription factor II-I), *LIMK1* (LIM domain kinase 1) and *CLIP2* (CAP-Gly domain containing linker protein 2) genes. Importantly, this patient also carries a 457.8Kb duplication at 15q13.3 that encompasses the *CHRNA7*, *OTUD7A* and *CHRFAM7A* genes. Duplications at 15q13.3 have been described as a risk factor for NDs, including ADHD/ID/ASD (Williams *et al.*, 2012), and also congenital anomalies reviewed in Torres *et al.*, 2016), therefore this CNV can also be contributing for the phenotype.

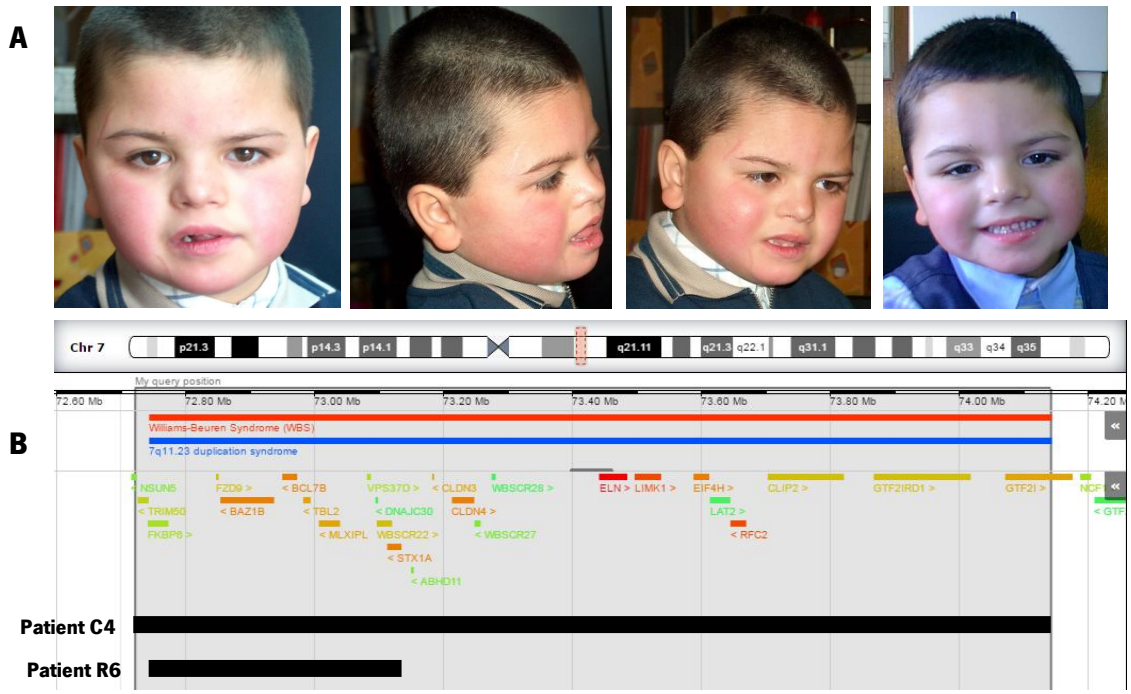


Figure 17 - Facial appearance of patient R4 (A) and WBS region shaded in grey (B). Each patient affected region is represented by individual black bars.

From the clinical point of view neither of the patients has a classic WBS presentation, which in the case of patient R6 might be due to the fact that the deleted region is relatively smaller that th more frequently affected region in WBS. With exception of *ELN* gene (deleted in patient C4 but not in patient R6), associated with cardiac problems, some the genes altered were suggested to be linked to the specific cognitive profile and craniofacial features presented by the WBS patients (Ferrero *et al.*, 2010).

22q13.3 deletions (Phelan-McDermid syndrome)

Patient C9 is a 7 years old boy who was referred to consultation at the age of 3, due to global DD, particularly in language and fine motor skills, but without behavior alterations and

dysmorphic features, and with regular growth. He has cognitive deficit, scoring below the average in all GMDS sub-scales (locomotor, personal-social, language, eye and hand co-ordination, performance, practical reasoning). Presently, he has also mild hypotonia and minor dysmorphic features. He has a 1.66Mb deletion at 22q13.3 that encompasses 39 genes (figure 12), including the S*H3 and multiple ankyrin repeat domains 3* (*SHANK3*) gene. This gene encodes for a Shank protein, which belongs to the group of multidomain scaffold proteins of the postsynaptic density that connect neurotransmitter receptors, ion channels, and other membrane proteins to the actin cytoskeleton and G-protein-coupled signaling pathways (Sheng and Kim, 2000). Shank proteins also play a role during axonal outgrowth and presynaptic development and function(Halbedl *et al.*, 2016). *SHANK3* is located within the minimum critical region of the 22q13.3 deletion syndrome, or Phelan-McDermid syndrome (PMS). Mutations in this gene are associated with several NDs, including autism, presented by more than 50% of the PMS patients (Phelan and Mcdermid, 2012), schizophrenia and bipolar disease (Choi *et al.*, 2015).

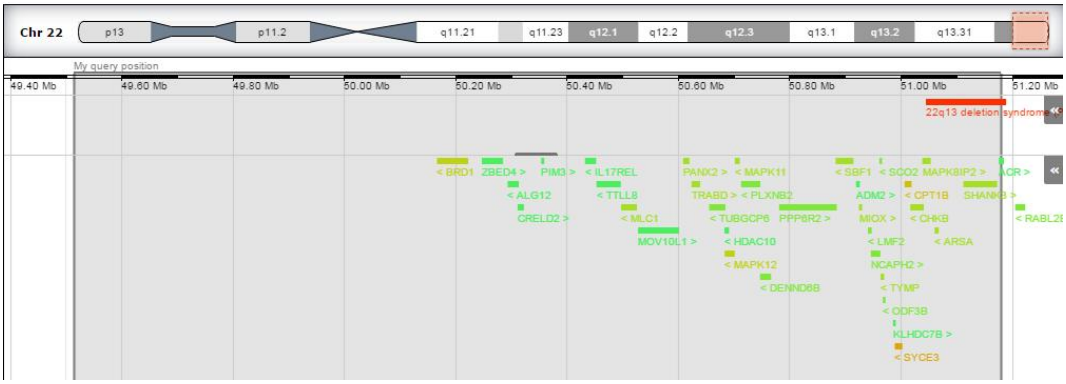


Figure 18 – Deleted 22q13.3 region and the genes involved for patient C9.

PMS patients may have nonspecific clinical presentations, but usually present with neurological deficits, including global DD, ID (moderate to severe), absent or severely delayed speech, normal growth and several minor dysmorphic features (namely asymmetric face, maxillary prognathism, dysmorphic ears, ptosis and bulbous nose) (Phelan, 2008; Phelan and Mcdermid, 2012). Patient C10 presents some clinical overlap (not complete) with the presentation, namely the hypotonia, speech impairment and motor development, but no significant facial dysmorphisms. The differences severity of symptoms from the patient with other PMS patients can be linked to variability in the extent of mitochondrial dysfunction, caused by disturbance of several mitochondrial genes within the 22q13.3 critical region, like *SCO2*, *TYMP* and *CPT1B,* all affected in this patient, as well as *NDUFA6*, *TRMU* and *ACO2* (Frye *et al.*, 2016). This variability in the

extent of the mitochondrial dysfunction can contribute for the incomplete overlap of the patient´s phenotype with that previously described.

9q34 duplications

We detected two non-related patients with 9q34 duplication. One is patient R16, a 16 years old boy with mild ID (IQ= 57), facial dysmorphisms, hypochromic macules and *café-au-lait* spots (figure 13A). Behaviorally, he presents stereotypies, obsessive and aggressive behavior and ADHD. aCGH revealed three *de novo* duplications: a 5.5 Mb duplication at 9q34.13-q34.3, a 1.6 Mb duplication at 14q32.31-q32.33 and a 1.4 Mb duplication at 14q32.33 (figure 13B, C, D, E). All of these findings might possibly contribute to his phenotype, making it difficult to ascertain the specific role of each imbalance. In order to determine if a structural rearrangement originating these three CNVs was present, a conventional karyotype and FISH analysis was performed. We could observe that patient R16 carries a translocation rearrangement between chromosome 9 and chromosome 14 in which the derivative chromosome results from a *de novo* translocation in which the 9q duplicated region is located in 14p. Moreover, the duplicated 14q32.31-q32.33 region is located in tandem and may lead to the disruption of the genes involved in the breakpoints. It was not possible to determine in which chromosome 14 the 14q duplication is located (in the derivative or in the normal one) (figure 13E).

The 9q34.3 duplication encompasses the entire *euchromatic histone-lysine N-methyltransferase 1* (*EHMT1*) gene. Haploinsufficiency for *EHMT1* gene causes Kleefstra syndrome (KS, MIM 610253), which is commonly due to deletions in 9q34.3, affecting the *EHMT1* gene or, more rarely, by *EHMT1* point mutations (Kleefstra *et al.*, 2006, 2009). Duplications encompassing ~600 kb of the distal 9q34.3 region have been reported previously in unaffected Individuals and therefore were proposed to represent likely benign CNVs (Redon *et al.*, 2006; Yatsenko *et al.*, 2012). However, this does not include duplications affecting the 5' region of the *EHMT1* gene, nor duplications or triplications encompassing the entire *EHMT1* gene, which have not been observed in healthy control populations; on the contrary, these imbalances were observed in patients with neurodevelopmental impairment, speech delay, and autism spectrum disorders (ASD), suggesting that increased *EHMT1* dosage is associated with a neurobehavioral phenotype (Yatsenko *et al.*, 2012).

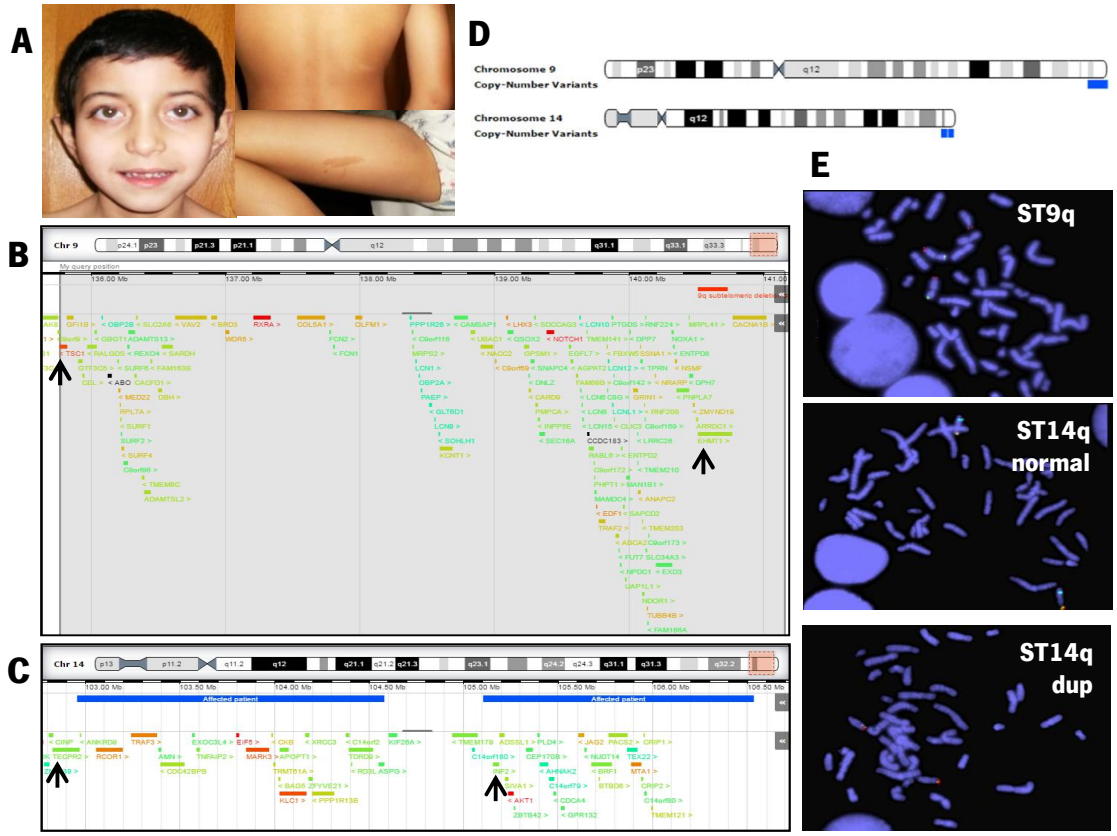


Figure 19 - Clinical features of the patient R16 (A) and images of the 9q31.13-q34 (B), 14q32.31-q32.33 and 14q32.33 (C, D) duplicated regions. Facial appearance of patient R16 and hyperpigmentation spots on the skin (A). Schematic representation of the 9q31.13-q34 duplicated region (B) and 14q32.31-q32.33 and 14q32.33 duplicated regions (C) (the black arrows indicate the genes and location where the primers for expression studies were designed. (D) Schematic representation of the three duplications present in the patient. (E) FISH analysis for the location of the duplicated regions.

In order to determine if duplications found (at the 14q and 9q) in patient R16 would affect the expression of the genes involved, the expression of *EHMT1* and *TSC1* genes (for the 9q dup) and the *TECPR2* and *INF2* genes (for the 14q dup) were studied in peripheral blood cells (figure 14). The expression of *EHMT1* and *TCS1* was found to be increased in the peripheral blood of the patient R16 when compared to controls. Because we know that the 9q34 duplicated region is not located in tandem (due to the FISH studies), this finding is in accordance with the hypothesis that if duplication is located in *trans* is likely that there will be no structural effect in the region and the genes won´t suffer an influence in expression. As for the 14q32 duplicated region, the *INF2* mRNA expression was also found to be increased in the patient, which is an observation in accordance with the fact that the entire gene is located inside the duplicated region.

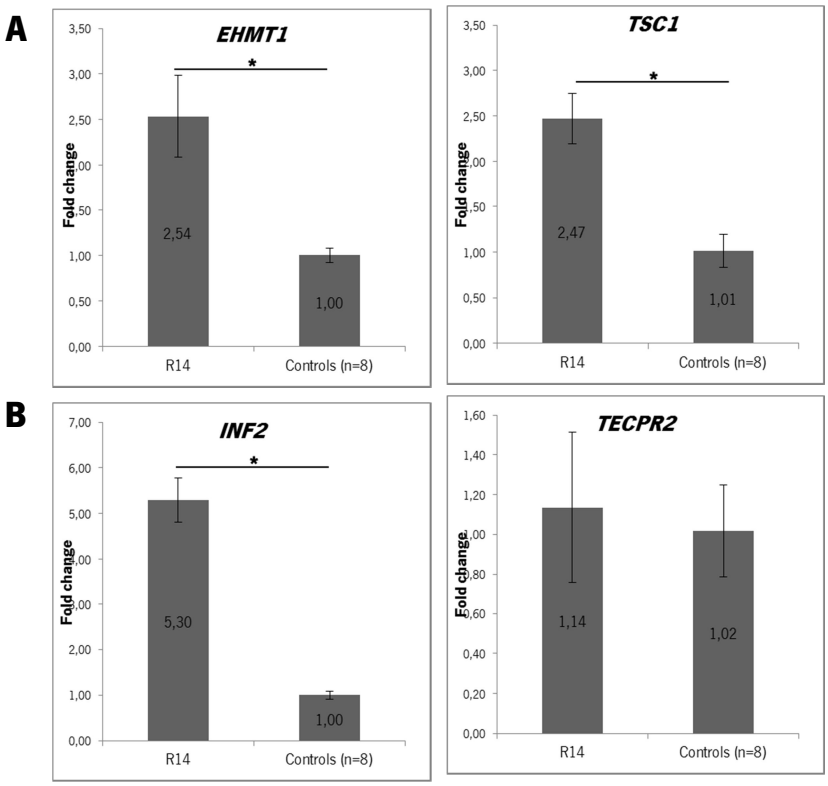


Figure 20 – mRNA expression in patient R16 and controls. (A) *EHMT1* and *TSC1* mRNA increased expression ($p<0.05$*) located in the 9q31.13-q34 region. (B) *INF2* increased expression (*B2M* was used as housekeeping gene; student t-test; $p<0.05$*) in the 14q32.33 region and *TECPR2* unaltered expression in the 14q31.31-q32.33 region.

On the other hand, *TECPR2* expression is not altered in the patient. The portion of the transcript where the primers for *TECPR2* were designed is located outside the duplicated region. Nevertheless, we believe that if the duplicated region affected the expression of the gene this would be still possible to observe since one of the alleles (the one located in the duplicated chromosome) would result in the degradation of the entire mRNA molecule.

The other patient with a 9q34.3 duplication is patient C11, She is a 10 years old girl with moderate ID, speech and motor delay, behavioral problems (ADHD), who has a maternally inherited 118Kb duplication at 9q34.3 that affects the 5' region of the *EHMT1* gene (figure 15A and B). Her clinical presentation overlaps the KS core phenotype, i.e. ID, hypotonia and distinctive facial features (Kleefstra *et al.*, 2006, 2009, 2012), besides hypertelorism, obesity and epilepsy. Until now, only one case of KS caused by an intragenic *EHMT1* duplication was described (Schwaibold *et al.*, 2014); intriguingly, the duplication is very similar to the one presented by C11. Although most cases of KS are caused by *de novo* mutations, two unrelated families were described in which affected children inherited a 9q34.3 deletion from a mildly affected mother who was somatically mosaic for the deletion (M. Willemsen *et al.*, 2011). Patient

C11's mother also carries the duplication and has confirmed psychiatric/cognitive problems, which are frequently observed in elderly KS patients (M. H. Willemsen *et al.*, 2011).

Figure 21 - Facial appearance of patient C11 (A) and 9q34.3 duplicated region affecting *EHMT1* gene (B).

Xp11.22 duplication

R13 is a 13 years old boy with mild non-syndromic ID, no familial history of ID and with significantly low weight. He carries a maternally inherited Xp11.22 duplication that affects the *HUWE1* (HECT, UBA and WWE Domain Containing 1, E3 Ubiquitin Protein Ligase) gene (figure 16), his mother being reported as asymptomatic.

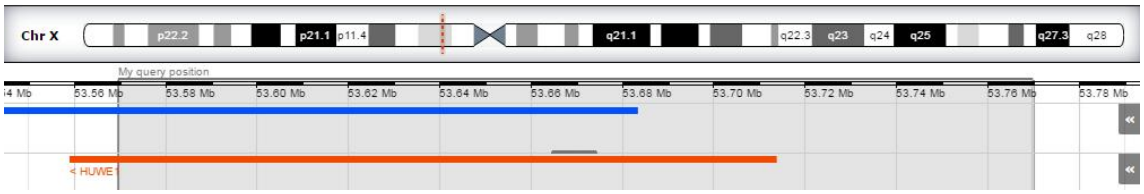


Figure 22 – Duplicated Xp11.22 region and the gene involved.

If located in tandem this duplication could disrupt the gene, but we could not confirm if this was the case. Consistent with our case, microduplications and point mutations in *HUWE1* have also been described in 13 independent families with patients with non-syndromic ID (Froyen *et al.*, 2012; Isrie *et al.*, 2013).

Xp22.31 duplications

Patient C16 is a 17 years old girl with ID (IQ= 57), motor and speech delay and no significant dysmorphisms. She carries a Xp22.31 duplication of 1.695Mb that affects 7 genes, including the *Variable charge, X-linked 3A* (*VCX3A* or *VCXA)* gene (figure 17).

Figure 23 - Duplicated Xp22.31 region and the gene involved.

This gene belongs to the VCX/Y gene family and encodes an RNA-binding protein that specifically binds the 5' end of capped mRNAs to prevent their decapping and decay (Jiao *et al.*, 2009). Regulation of mRNA metabolism is a key point in neurons, with relevance for synaptic structure and function, and its disruption, through absence of RNA-binding proteins, especially FMRP, FMR2P, PQBP1, UFP3B and VCX-A, causes different forms of ID (Bardoni *et al.*, 2012). Although variants in this region can also be present in individuals without phenotypic manifestations (Li *et al.*, 2010), namely in carrier mothers of female probands (Esplin *et al.*, 2014), ID was a common finding in patients with such duplications, in addition to minor facial dysmorphisms, but no major malformations (Faletra *et al.*, 2012).

Likely pathogenic variants

The likely pathogenic variants detected in 5.5% of patients in this study, were mainly composed of large and gene-rich CNVs, both *de novo* and inherited from an affected parent, that encompass OMIM morbid genes and were not found in control databases (table IV). They comprise novel candidate ID-causative *loci* located in 2q11.2-q12.2, 7q33, 10q26.3, 17p11.2, 20q13.12-q13.13, 1p22.1-p21.3, 2p15, 9q33.2-q33.3, 12p13.33, Xq24 and Xq26.3.

Table VII – List of very likely pathogenic variants present in the patients.

Patient(s)	ISCN description (Hg19)	Type	Size (Kb)	Genes	Relevant genes involved	Confirmation	Inheritance	DGV controls	Decipher	Array platform
R17 (♀)	2q11.2-q12.2(101,756,265-106,265,018)x1dn	del	4500	24	*MAP4K4,FHL2, POU3F3, CNOT11*	qPCR	*de novo*	No	251756	1
R18, R19¥,§ (♂),(♀)	7q33(133,176,651-135,252,871)X1mat	del	2076	23	*AGBL3, CNOT4, CALD1, EXOC4*	qPCR	Maternal‡	No	256036	1
R20 (♀)	10q26.3(131,374,701-132,030,468)X1dn	del	600	3	*EBF3*	qPCR	*de novo*	3/6564**	No	1
C17 (♂)	17p11.2(16,757,564-17,178,161)x1mat	del	420	5	*COPS3*	NP	Maternal‡	No	No??	2
	17p11.2(18,478,816-21,255,056)x1mat	del	2770	36	*EPN2, RNF112, ULK2, ALDH3A2, AKAP10*	NP	Maternal‡	No		
R21 (♀)	20q13.12-q13.13(43,283,820-48,850,844)x1dn	del	5500	88	*KCNB1, PIGT, CTSA, SLC2A10, AEFGEF2*	NP	*de novo*	No	309	1
C18 (♀)	1p22.1p21.3(92,227,986-98,689,243)x3	dup	6461	44	*FAM69A, TGFBR3, GLMN, EVI5, RPL5, MTF2, DR1, ABCA4, ABCD3, CNN3, PTBP2, DPYD*	NP	NP	No	3183581	3
C19,C20,C21¥,§ (♂),(♂),(♂)	2p15(61,377,041-61,522,171)x3mat	dup	14	3	*C2orf74; AHSA2; USP34*	qPCR	Maternal	1/6533	256542; 279248(del)	2
R22 (♂)	2q11.2(96,735,183-98,228,265)X3mat	dup	1496	24	*ARID5A, NEURL3, SEMA4C*	NP	Maternal‡	1/6533	254924, 274288	1
C22,C23¥,§ (♂),(♂)	7q33(134598205-134815177)x3mat	dup	216	2	*CALD1, AGBL3*	qPCR	Maternal‡	No	No	3
R23 (♀)	9q33.2-q33.3(123,525,064-127,187,619)X4dn	tri	3600	52	*CRB2, LHX2, LHX6, DENND1A, STRBP, RAB14, GSN, PSMB7, ZBTB26*	qPCR	*de novo*	No	No	1
R24 (♂)	12p13.33p13.32(2,248,863-3,497,525)X3pat	dup*	1248	9	*CACNA1C, TULP3, FOXM1, TSPAN9*	qPCR	Paternal	No	No	1
R25, R26¥ (♂),(♂)	Xq24(119,592,606-119,904,981)X2mat	dup*	300	4	*CUL4B, LAMP2, C1GALT1C1, MCTS1*	qPCR	Maternal	No	No	1
C24 (♂)	Xq26.3(135,293,144-135,863,290)x2mat	dup	570	9	*ARHGEF6, CD40LG, BRS3*	qPCR	Maternal	No	No	2

Legend: Patients R17 to R26: from research cohort; Patients C17 to C24: from clinical cohort; NP: not performed; (‡): inherited from an affected parent; (*): duplication may disrupt gene if located in tandem; (**) doubt regarding the quality of the call in these controls; (¥): siblings; (§): family described in detail elsewhere (R18 and R19 in sub-chapter 2.4). Array platform 1: Agilent 180K; 2: KaryoArray®v3.0 (Agilent 8x60k); 3: Affymetrix Cytoscan 750K.

2q11.2-q12.2 deletion

Patient R17 is a 17 years old girl with syndromic ID, ventricular enlargement, dysmorphic features and hirsutism, whose mother is suspected to have mild ID, although no psychometric assessment was performed.

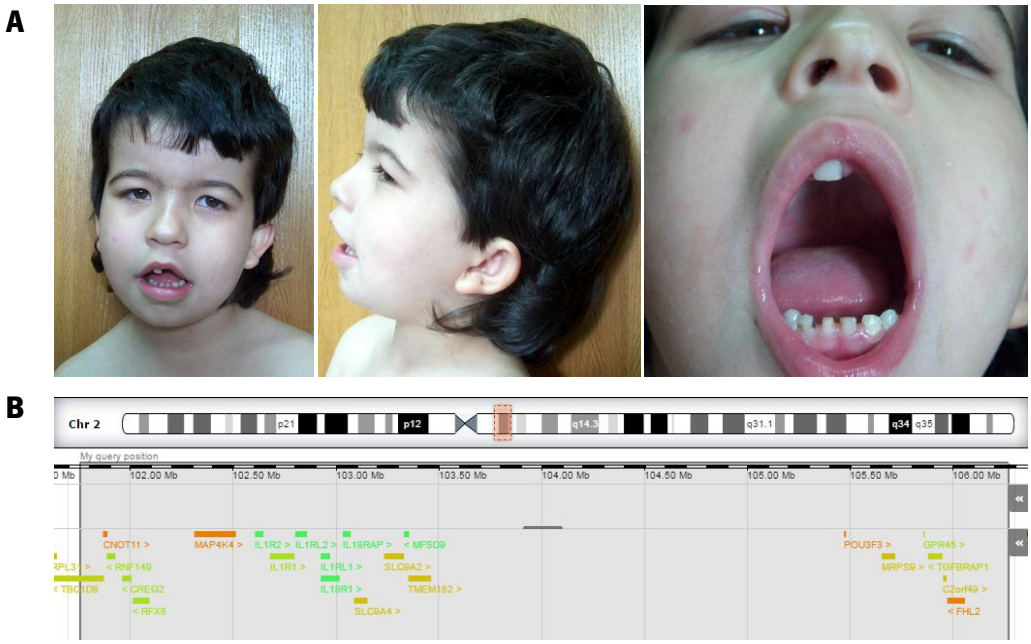


Figure 24 - Facial appearance of the patient (A) and 2q11.2-q12.2 deleted region (B).

She carries a *de novo* 4.5 Mb deletion at 2q11.2-q12.2 affecting 26 genes (figure 18), of which *MAP4K4* (Mitogen-Activated Protein Kinase Kinase Kinase Kinase 4), *FHL2* (Four and a Half LIM Domains 2), *POU3F3* (POU Class 3 Homeobox 3) and *CNOT11* (CCR4-NOT Transcription Complex Subunit 11) have the highest haploinsufficiency score in Decipher. *POU3F3* was previously reported as deleted in a boy with ID and dysmorphic features (such as flat nose, prominent ears, large eyebrows and low hairline) (Dheedene *et al.*, 2014), similar to those of our patient. This gene encodes a transcription factor present in post-mitotic cells and plays a role in neurogenesis and the correct destination of migratory neurons in the cerebral cortex in the mouse (Dominguez *et al.*, 2013). For these reasons POU3F3 stands out as a good candidate for the DD/ID in the patient.

10q26.3 deletions

Patient R20 is a 11 years girl with severe ID (QGD=27 at 7 years of age) born from non-consanguineous parents without history of NDD and with a healthy twin sister. Global developmental delay was noted in the first months (head control at 12 months, sitting at 18

months, independent walking after 30 months, no words spoken at the age of 3 years). She had pyelonephritis at 19 months (renal ultrasound showed no abnormalities), gastroesophageal reflux and recurrent otitis media with conductive hearing loss that required surgical intervention at the age of 2 years 11 months, and a hearing aid. Epilepsy was suspected at 5 months (episodes of suspended activity) but eletroencefalogram (EEG) was normal; she had eyeglasses for strabismus and hypermetropia.

She was first observed at the age of 3 years 5 months (figure 19A-B), at which time she displayed axial hypertonia, mild muscle hypotonia with a hypotonic face, reduced sensitivity to pain and conductive hearing loss. She also presented dysmorphic features (triangular face, small low-set ears with prominent anti-helix, arched eyebrows, anteverted nares, bulbous nasal tip, small mouth with downturned corners, pointed chin; short neck, prominent finger fetal pads) and mild short stature. Brain magnetic resonance imaging (MRI) was performed at the age of 6 years, no abnormalities were noted. At the age of 10 years she was reevaluated (figure 19C). She still had recurrent otitis media, but otherwise was in good global health. Language was very poor, with two word sentences spoken at 8 years. She had significant behavior problems, with stereotypic movements (rotating movements, chewing on clothes, head retropulsion), scoring for a severe autism spectrum disorder (Autism Diagnostic Interview-Revised (ADI-R) and Autism Diagnostic Observation Schedule (ADOS)) at the age of 7 years. She also displayed agitation and aggressive behavior (self-injury and aggression towards others) and was medicated with antipsychotic drugs. An orthopedic surgery was performed for *pes planus*. The facial features were similar to those previously described, with spaced upper central incisors (Figure 19B).

She has a *de novo* 600 kb deletion at 10p26.3 affecting three genes, *MGMT*, *EBF3* and *GLRX,* of which *EBF3 is* the only disease-associated gene. It encodes a member of the highly conserved early B-cell factor transcription factor family and is regulated by *ARX* (Friocourt and Parnavelas, 2011). In mice, knocking out *Ebf3* leads to neonatal lethality and neuronal migration defects, with failure of olfactory neurons project to the dorsal olfactory bulb (Wang *et al.*, 2004). Unfortunately, no description is made of a phenotype in the heterozygous animals, which are actually presented as controls in many of the experiments. We made efforts to obtain and study the neurodevelopmental phenotype of these animals, but were not successful as the Ebf3(O/E2) knock-out line may have been discontinued (Joseph W. Lewcock, personal communication). At the time of aCGH analysis, the existence of the three variants present in DGV affecting the first five exons of *EBF3* gene (figure 19D) (Park *et al.*, 2010; Cooper *et al.*, 2011) as well as the

absence of other known disease causing mutations in this gene, lead us to classify it as a variant of unknown significance (VOUS).

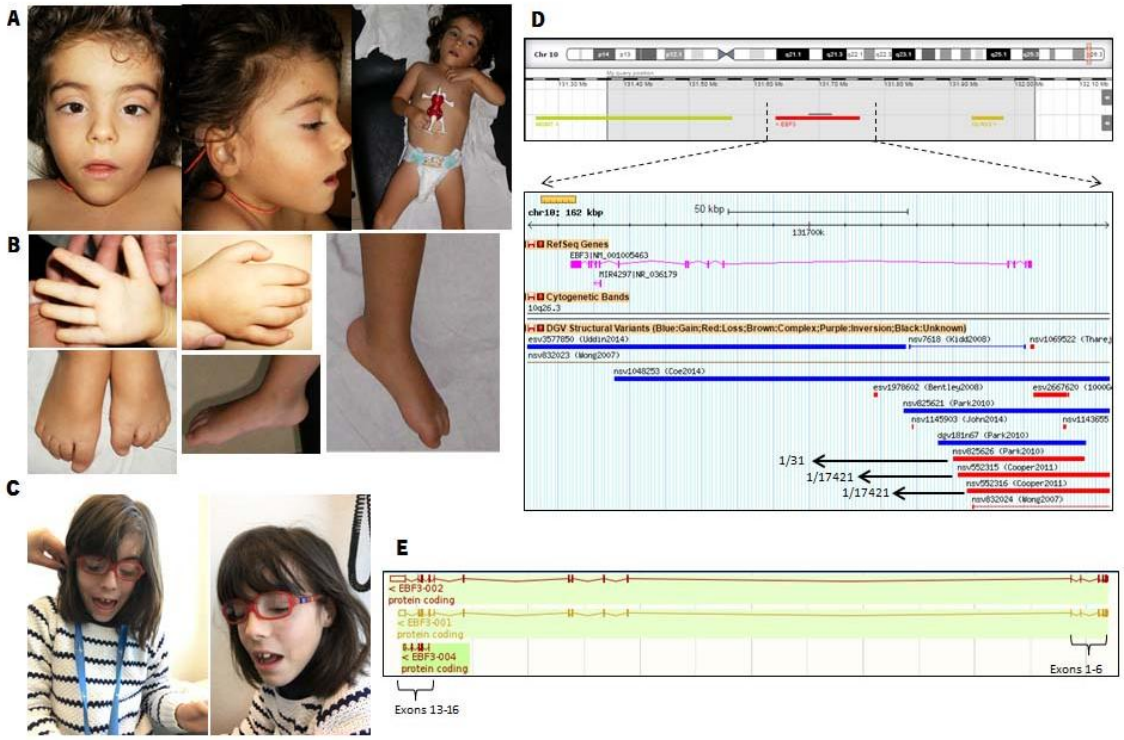


Figure 25 - Appearance of the patient at the age of first consultation (A and B) and currently (C). 10q26.3 deleted region and affected genes (D). *EBF3* gene and the variations present at DGV (E).

Only recently, dominant point mutations and indels have been described in *EBF3* as causative of ID suggesting that haploinsufficiency of this gene causes neurodevelopmental problems (Chao *et al.*, 2017; Harms *et al.*, 2017; Sleven *et al.*, 2017). These patients carrying *de novo* mutations have ID, ataxia, facial dysmorphisms, variable degrees of speech impairment and, in several cases, hypotonia. One of the things that raised doubts about the pathogenicity of this variant in the first place was the existence of population controls bearing deletions of the first six exons of this gene, in heterozygosity (data retrieved from DGV database as of February 2017) (figure 19D) ('Database of Genomic Variants', n.d.) Even though a transcript of *EBF3* starting in Exon13b is listed in the Ensembl database (ENST00000440978.1) (figure 19E), which could explain how deletion of the first exons could eventually result in a normal phenotype, this transcript excludes the DNA binding domain of *EBF3*, and its expression pattern and functional relevance have not been characterized. Upon reassessment however, the CNVs described by Cooper *et al*. (nsv552315 and nsv552316) are considered to be at the threshold of detection by SNP microarray and cannot be the basis for exclusion of a candidate gene, particularly in light of the

strong genetic and functional evidence for the relevance of *EBF3* mutations to disease (Evan Eichler, Greg Cooper and Bradley Coe, personal communication). Hence, we conclude that the data supporting the contribution of *EBF3* deletion for the disease is stronger than the presence of these 3 controls in DGV.

17p11.2 deletions

Patient C17 is a 7 year old boy with clinically described DD, namely language and motor impairment, ataxia and some dysmorphic features, including hypertelorism, strabismus and low ear implantation. He has performed a cerebral magnetic resonance imaging (MRI), which showed no alterations. Although, based on GMDS score he has no cognitive deficit, he presents an impairment in adaptive behavior. He has what appear to be two consecutive deletions at 17p11.2: a 420.6Kb deletion, that encompasses 5 genes, and a 2.77Mb deletion that encompasses 36 genes (figure 20).

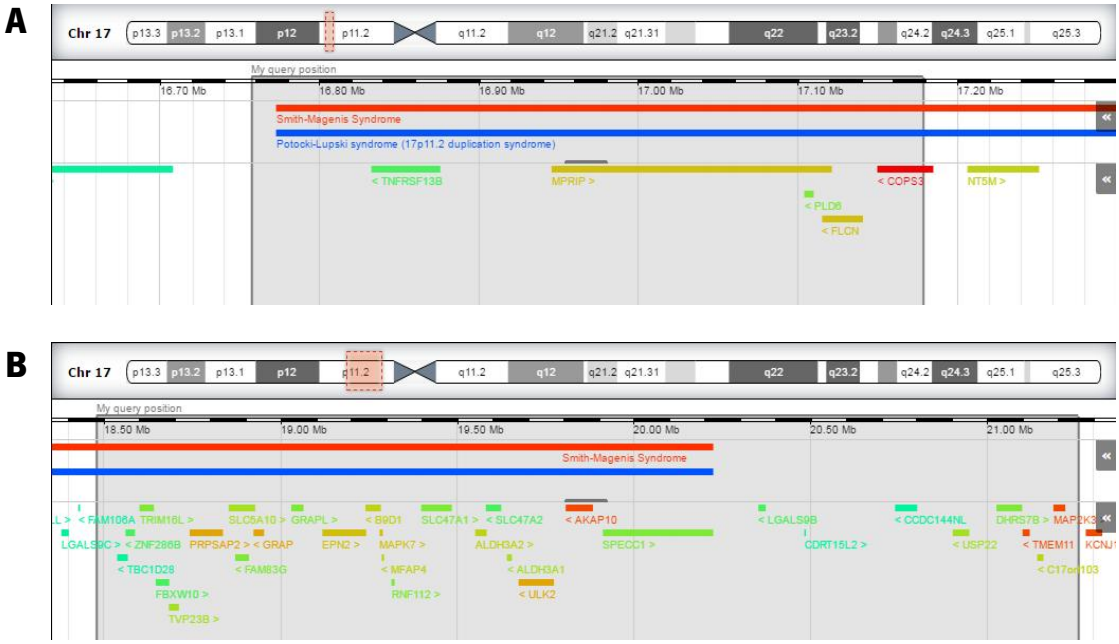


Figure 26 – Both 17p11.2 deleted regions: (A) the 420Kb region and (B) the 2.77Mb region.

He has inherited them from his mother, who has confirmed learning difficulties, although she has completed the 6th grade. These deletions partially overlap the region involved in Smith-Magenis syndrome (SMS), although the phenotype of the patient and mother is not similar to that of SMS, perhaps because the deletion does not affect the *retinoic acid induced 1* (*RAI1*) gene, located within the SMS critical region and thought to cause most of the SMS core phenotype (Potocki *et*

al., 2007; Elsea and Williams, 2011). Among the genes affected by patient C17's deletions, there are several others whose function could potentially contribute for his phenotype. For example *EPN2*, that encodes *epsin2*, a protein thought to be involved in clathrin-mediated endocytosis and found in a brain-derived clathrin-coated vesicle fraction (Rosenthal *et al.*, 1999); *RNF112* (ring finger protein 112, also known as ZNF179), which encodes a member of the RING finger protein family of transcription factors that is abundantly expressed in brain and is involved in neuronal differentiation (Tsou *et al.*, 2016), and *ULK2* (unc-51 like autophagy activating kinase 2), which encodes a protein similar to a serine/threonine kinase, the ortholog of which is known to be involved in axonal elongation in *Caenorhabditis elegans* (Tomoda *et al.*, 2004). Moreover, there are 3 OMIM genes affected by these deletions: *B9D1*, which encodes a B9 domain-containing protein, one of several that are involved in ciliogenesis (Williams *et al.*, 2008), associated with Meckel syndrome (AR) (Hopp *et al.*, 2011) and Joubert syndrome (AR) (Romani *et al.*, 2014); and *ALDH3A2*, which encodes for a fatty aldehyde dehydrogenase that causes Sjögren-Larsson syndrome (AR) when mutated (De Laurenzi *et al.*, 1996). These being recessive disorders the presence of these deletions in heterozygosity are unlikely to cause the phenotype. Nevertheless, a modifying effect cannot be excluded.

20q13.12-q13.13 deletion

Patient R21 is a 16 years old girl with mild ID (IQ= 56), speech delay, microcephaly and facial dysmorphisms as well as astigmatism and hyperactivity. Brain imaging studies revealed no structural alterations. She carries a *de novo* 5.5 Mb deletion at 20q13.12-q13.13 encompassing 123 genes (figure 21). Among these, the genes *KCNB1, PIGT, CTSA, SLC2A10* and *ARFGEF2* were associated with human disease, whereas *MMP9, CSE1L* and *YWHAB* have very high haploinsuffciency scores in Decipher.

KCNB1 (Potassium Channel, Voltage Gated Shab Related Subfamily B, Member 1) encodes a potassium channel in which *de novo* missense mutations were found (in both heterozygosity and homozygosity) in patients with epilepsy and neurodevelopmental disorders (Torkamani *et al.*, 2014).

The *ADP ribosylation factor guanine nucleotide exchange factor 2* (*ARFGEF2*) gene encodes a protein involved in the activation of ADP-ribosylation factors (ARFs) and is required for vesicle and membrane trafficking in the Golgi (Sheen *et al.*, 2004). Homozygous mutations in this gene have

been described in patients with epilepsy, periventricular heterotopia with microcephaly, movement disorders and ID (Banne *et al.*, 2013; Yilmaz *et al.*, 2016).

Even though the clinical is not completely in accordance, the previous description of *KCNB1* and *ARFGEF2* genes with NDD make them good candidates for explaining the phenotype in the patient. However, the contribution of other genes in the deletion cannot be excluded.

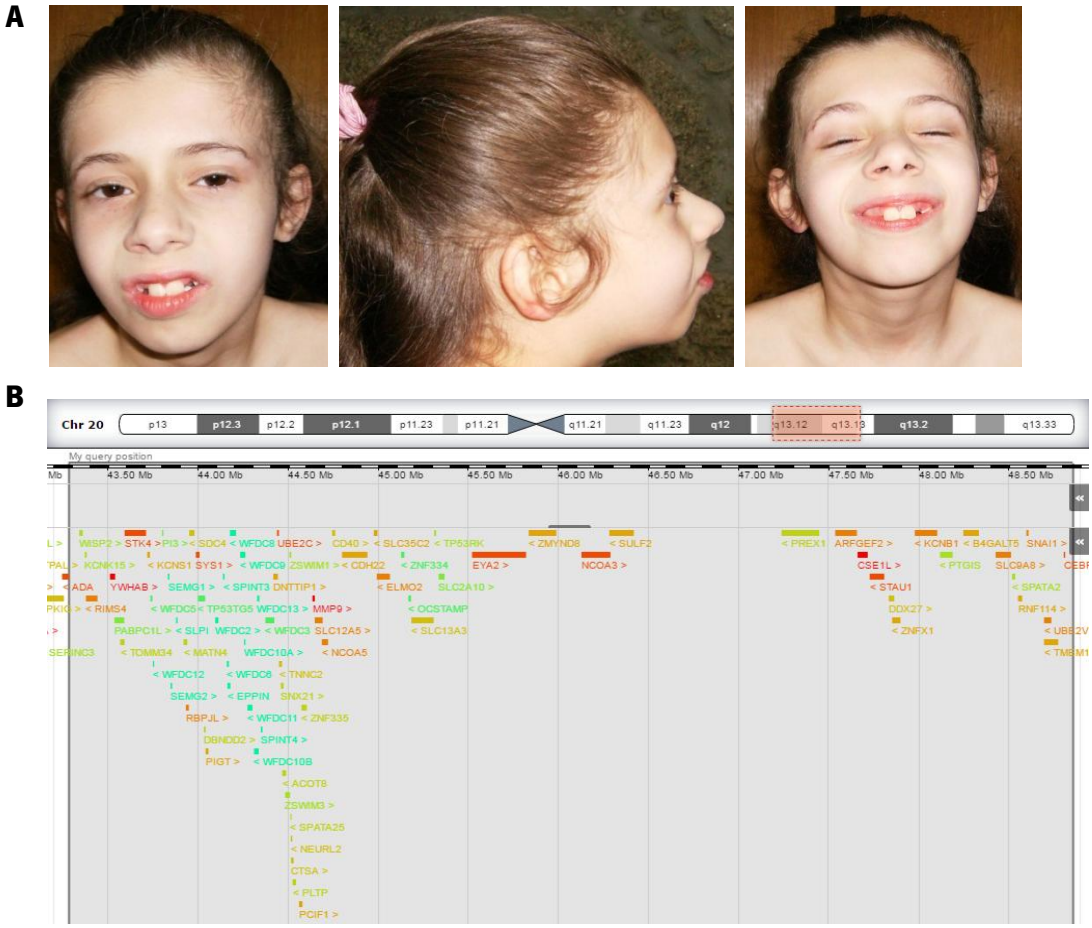


Figure 27 - Facial appearance of the patient (A) and the 20q13.12-q13.13 deleted region (B).

1p22.1p21.3 duplications

Patient C18 is a 5 years old girl with motor and speech delay, who, according to GMDS (evaluated in 2017), has a global DQ of 56. She carries a maternal 1p22.1p21.3 duplication of 6.461Mb that affects 44 genes (figure 22). Her mother has completed the 6th grade although with 2 in-grade retentions and always showing learning difficulties, especially in language skills. The girl has an 8 years old brother suspected of having cognitive deficit: he was not evaluated yet, but he is attending the 2nd grade and does not know how to read. Additionally, there is also a positive history of learning difficulties on the maternal grandfather's family side.

Figure 28 - Facial appearance of the patient (A) and the 1p22.1p21.3 duplicated region (B).

This duplication affects several genes, including the *family with sequence similarity 69 member A* (*FAM69A*) gene, which encodes a member of the FAM69 family of cysteine-rich type II transmembrane proteins. FAM69 proteins are likely to play a fundamental role in the endoplasmic reticulum of most metazoan cells, in addition to specialized roles in the vertebrate nervous system (Tennant-Eyles *et al.*, 2011), according to a brain-specific or brain-including expression pattern (Dudkiewicz *et al.*, 2013). Consistently, several *FAM69* genes have been linked to several neurological disorders in genetic studies: *C3ORF58* (*DIA1*) with autism (Morrow *et al.*, 2008); *CXORF36* (*DIA1R*) with X-linked ID (Thiselton *et al.*, 2002) and *FAM69A* linked to schizophrenia and bipolar disease (Wang *et al.*, 2010). Whether duplication of this gene/region is indeed the cause of this non-syndromic ID remains to be proven.

2q11.2 duplications

Patient R22 is a 14 years old boy with mild non-syndromic ID (IQ=61), no facial dysmorphisms and with one *café-au-lait* spot on the skin. He carries a 1.4 Mb duplication at 2q11.2 inherited from his affected mother that affects 39 genes (figure 23). Of those, the following genes are known to be associated with disease: *ADRA2B* (Adrenoceptor Alpha 2B) causes autosomal dominant adult myoclonic epilepsy 2, MIM 607876; *TMEM127* (Transmembrane Protein 127) confers susceptibility for Pheochromocytoma, MIM 171300; *SNRNP200* (Small Nuclear

Ribonucleoprotein U5 Subunit 200) causes Retinitis pigmentosa 33, MIM 610359; *LMAN2L* (Lectin, Mannose Binding 2 Like) causes AR ID 52, MIM 616887; and *CNNM4* (Cyclin And CBS Domain Divalent Metal Cation Transport Mediator 4) causes AR cone-rod dystrophy and amelogenesis imperfecta, also called Jalili syndrome, MIM 217080. *TMEM127*, *SNRNP200* and *LMAN2L* are the ones with the higher haploinsuficiency scores in Decipher. *LMAN2L* encodes a transmembrane protein that works in the glycoprotein transport in the endoplasmic reticulum (ER) and in which recessive mutations were recently found to cause ID (Rafiullah *et al.*, 2016). *LMAN2L* has also been associated, through GWAS studies, to schizophrenia and bipolar disorder (Lim *et al.*, 2014). Even though so far the contribution of the excess of dosage of *LMAN2L* for NDDS is still unknown, this gene can be considered as a good candidate for the disease in the patient.

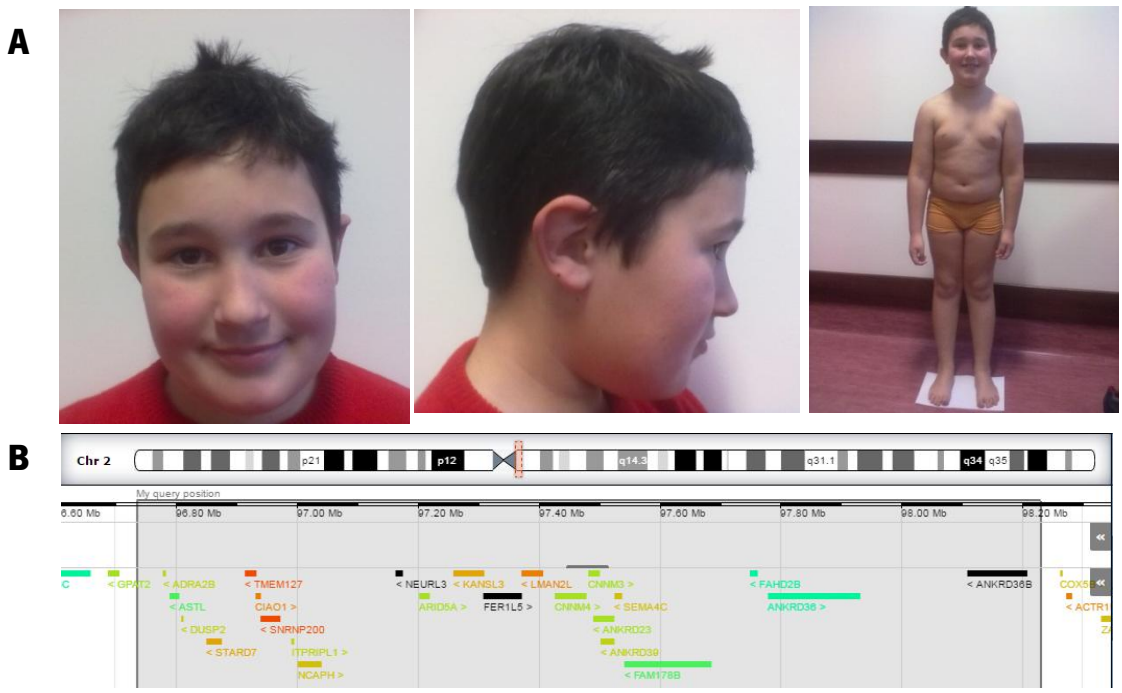


Figure 29 - Facial and body appearance of the patient (A) and the 2q11.2 duplicated region (B).

9q33.2-q33.3 triplication

Patient R23 is a 17 years old girl with mild ID (IQ=53) and familial history of ID. During the perinatal period she presented epilepsy, hypothyroidism and cardiac insufficiency. She has a delay in bone maturation (~3 years), short stature, relative macrocephaly, short neck and dysmorphic face (figure 24). She carries a 3.6 Mb *de novo* triplication at 9q33.2-q33.3 that affects 60 genes. Of those, only the *CRB2* gene is associated with a human disease, but *LHX2,*

DENND1A, STRBP, RAB14, GSN, PSMB7, LHX6 and *ZBTB26* have a very high haploinsuficiency score in Decipher. Moreover, this triplication apparently disrupts the *FBXW2* gene (*F-box and WD repeat domain containing 2*), that encodes for an F-box protein. F-box proteins are one of the four subunits of ubiquitin protein ligases, called "Skp, Cullin, F-box containing complex" (SCFs). SCF ligases bring ubiquitin conjugating enzymes to substrates that are specifically recruited by the different F-box proteins. Components of this complex, such as CUL4B, have been involved in ID pathogenesis (Vulto-van Silfhout *et al.*, 2015).

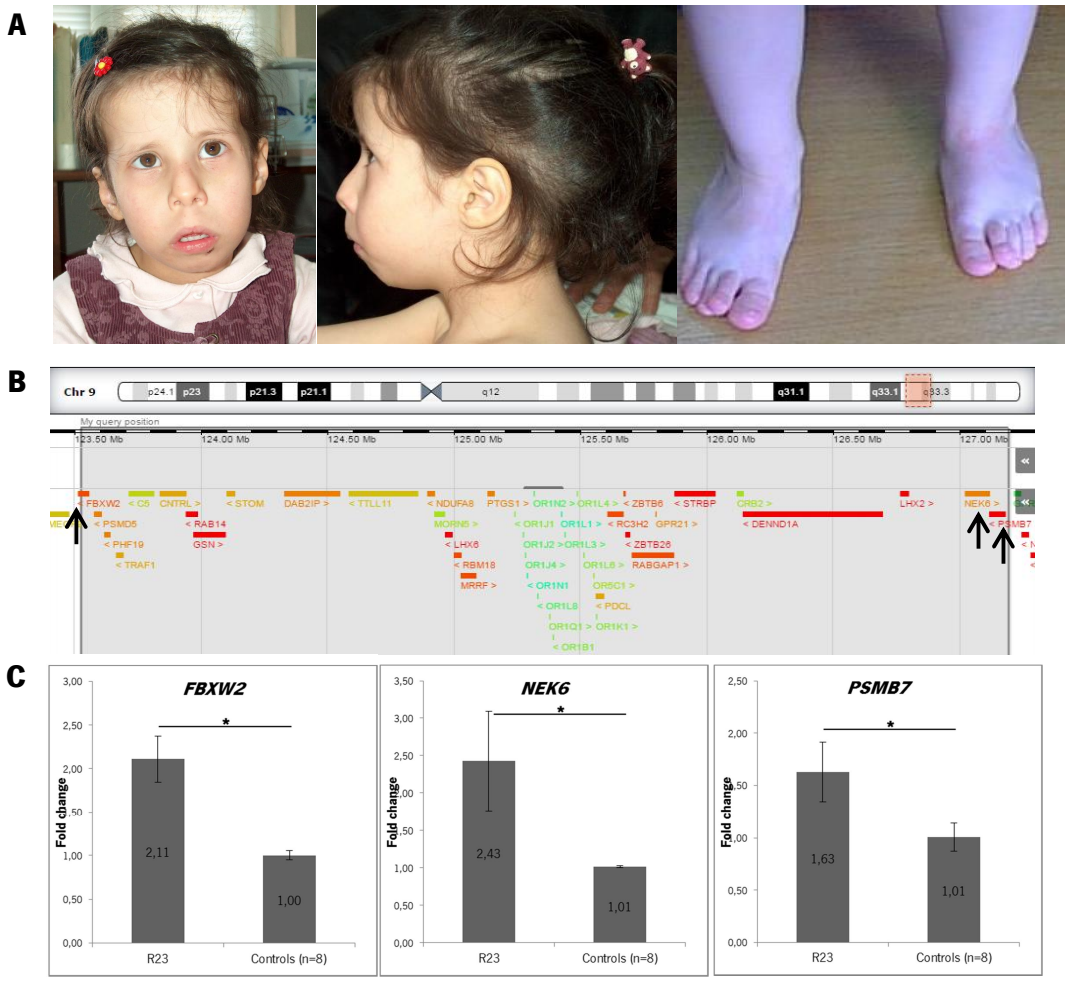


Figure 30 – Facial appearance of the patient (A) and the 9q33.2-q33.3 triplicated region (B). mRNA expression in patient R23 and controls (C). The black arrows indicate the place where the primers for mRNA expression were designed. FBXW2, NEK6 and PSMB7 genes are overexpressed at the mRNA level for patient R23 (B2M and PPIB were used as housekeeping genes; student t-test; p<0.05*).

LHX2 (LIM Homeobox 2) and *LHX6* (LIM Homeobox 6) both encode transcription factors described to play roles in brain development (Flandin *et al.*, 2011; Roy *et al.*, 2014). Additionally, *LHX2* was also described to be involved in osteoclast differentiation and its overexpression inhibits skeletal muscle differentiation (Kim *et al.*, 2014; Kodaka *et al.*, 2015). LHX6 is also

known to play a role in cranial and tooth development (Zhang *et al.*, 2013). Hence these results could be of relevance to the cranioskeletal phenotype of the patient, in addition to the ID/epilepsy aspects.

DENND1A (DENN/MADD Domain Containing 1A) is a guanine nucleotide exchange factor (GEFs) for the early endosomal small GTPase RAB35, that works on the endocytic branch of the synaptic vesicle cycle, and its ablation in hippocampal neurons leads to defects impaired synaptic vesicle endocytosis (Allaire *et al.*, 2006).

RAB14 (RAS associated protein RAB14) encodes a protein that works in the trafficking between the Golgi and the endosomal compartments and is important for early embryonic development (Junutula *et al.*, 2004). Therefore, this genomic imbalance may deregulate vesicle cycling and endocytosis in both neurons and in bone.

Based on the location within the triplication region and the expression levels in the periphery, we selected the *FBXW2* (F-Box and WD Repeat Domain Containing 2)*, NEK6* (NIMA Related Kinase 6) *and PSMB7* (Proteasome Subunit Beta 7) genes to study at the mRNA level in peripheral blood in the patient. We could observe that, in the peripheral blood cells from this patient, these three genes had an increased expression when compared to controls (figure 23C). For *NEK6* these findings are in accordance with the fact it is included inside the triplicated region and, therefore being overexpressed. As for *FBXW2* and *PSMB7* we hypothesize that their expression could be affected since they are located at the breakpoints, which we could observe not to be the case.

NEK6 encodes a protein kinase required for efficient mitotic spindle assembly (Prosser *et al.*, 2016) and was recently described as playing a role in telomere length regulation, making it key player in again and cancer (Hirai *et al.*, 2016). *FBXW2* encodes a F-box protein with multiple WD-40 repeats that in lung cancer is reported to have a tumour suppressor activity (Xu *et al.*, 2017, p. 2) but its function in neuronal cells is unknown. *PSMB7* encodes a proteasome subunit and is thought to play a role in autophagy inhibition in cardiomyocytes (Kyrychenko *et al.*, 2014, p. 7). To the best of our knowledge no mutations in any of these three genes were reported in human NDD or ID-related diseases, making the understanding of their overexpression for our patient difficult.

12p13.33-p13.32 duplication

R24 is a 14 years old boy with non-syndromic mild ID (IQ=63), no facial dysmorphisms and increased activity. He carries a 1.2 Mb duplication at 12p13.33-p13.32 inherited from the father (figure 25).

Figure 31 – The 12p13.33-p13.32 duplicated region and effected genes (A). mRNA expression in patient R24 (B). The black arrows indicate the place where the primers for mRNA expression were designed. TSPAN9 gene does not present altered expression at the mRNA level (B2M and PPIB were used as housekeeping genes; student t-test; $p<0.05$*).

The alteration affects 19 genes, of which only *CACNA1C* has previously been associated with a human disease. Additionally, both *CACNA1*C and *FOXM1* have a very high haploinsuficiency score in Decipher.

CACNA1C (Calcium Channel, Voltage-Dependent, L Type, Alpha 1C Subunit) encodes a calcium channel, and heterozygous point mutations in this gene cause Timothy syndrome (MIM 601005) and Brugada syndrome-3 (MIM 611875)(Splawski *et al.*, 2004; Antzelevitch *et al.*, 2005). *CACNA1C* variants have also been proposed to confer increased risk for autism and schizophrenia (Lancaster *et al.*, 2015; Li *et al.*, 2015). More recently, *CACNA1C* was described to play a role in the survival of young hippocampal neurons (Lee *et al.*, 2016). A *de novo* 2.3 Mb deletion disrupting the last exons of *CACNA1C* was described in a boy with ID, microcephaly, hypotonia and joint laxity. Although the deletion in that patient is smaller than the duplication here reported, they overlap in the *CACNA1C*, *TULP3* and *TSPAN9* genes.

TULP3 (Tubby Like Protein 3) encodes a protein essential for embryonic development in mice, absence of which leads to craniofacial, limb and neural tube defects (Ikeda *et al.*, 2001; Norman

et al., 2009), whereas *TSPAN9* (Tetraspanin 9) is a member of the tetraspanin gene family, which encodes several transmembranar regulators of cell signal transduction; *TSPAN9* is known to be expressed in megakaryocytes and may be involved in platelet function (Protty *et al.*, 2009), however its relevance to nervous system function is unknown. In the periphery, *TSPAN9* gene expression didn´t present any alteration when compared with controls (figure 24B). *CACNA1C* gene has very low expression in the periphery and was not possible to study.

Xq24 duplication

Patient R25 is a 14 years old boy without psychometric criteria for classification as ID (IQ= 80), but who had DD, familial history of ID (two brothers and cousins with ID), an apparently benign cardiac arrhythmia, generalized obesity, stereotypies and hyperactivity. He carries a 0.3Mb maternally inherited duplication at Xq24 affecting four genes (figure 26): *CUL4B, LAMP2, C1GALT1C1* and *MCTS1*. Both point mutations and large deletions in the *CUL4B* gene are described as causative of X-linked ID and cerebral malformations (Tarpey *et al.*, 2007; Zou *et al.*, 2007; Isidor *et al.*, 2009; Ravn *et al.*, 2012) *CUL4B* is a scaffold protein member of the cullin family that works in the formation of protein complex that acts as an E3 ubiquitin ligase catalyzing the polyubiquitination of protein substrates. *CUL4B* was found to be responsible for *TSC2* degradation in neocortical neurons, positively regulating mTOR activity in those cells (Wang *et al.*, 2013). Additionally, *CUL4B* also targets *WDR5* for ubiquitylation leading to its degradation in the neuronal nucleus, which causes impaired neurite outgrowth (Nakagawa and Xiong, 2011). However, to our knowledge, so far there is only one 47.2Mb duplication encompassing *CUL4B* (and other genes) described in a patient with ID (Jin *et al.*, 2015), the present case being the first small, non-disruptive *CUL4B* duplication described in a patient with ID.

CUL4B is entirely duplicated in the patient and its expression in peripheral blood cells is increased, leading to us to believe that the disease in the patient is in fact driven by a dosage increase in *CUL4B*. Also, *LAMP2* (the gene located in the duplication breakpoint) encoding a protein with roles in autophagy/lysosomal function does not present altered expression in the patient, which leads us to suspect that it might not be a contributing factor for this phenotype (figure 26C).

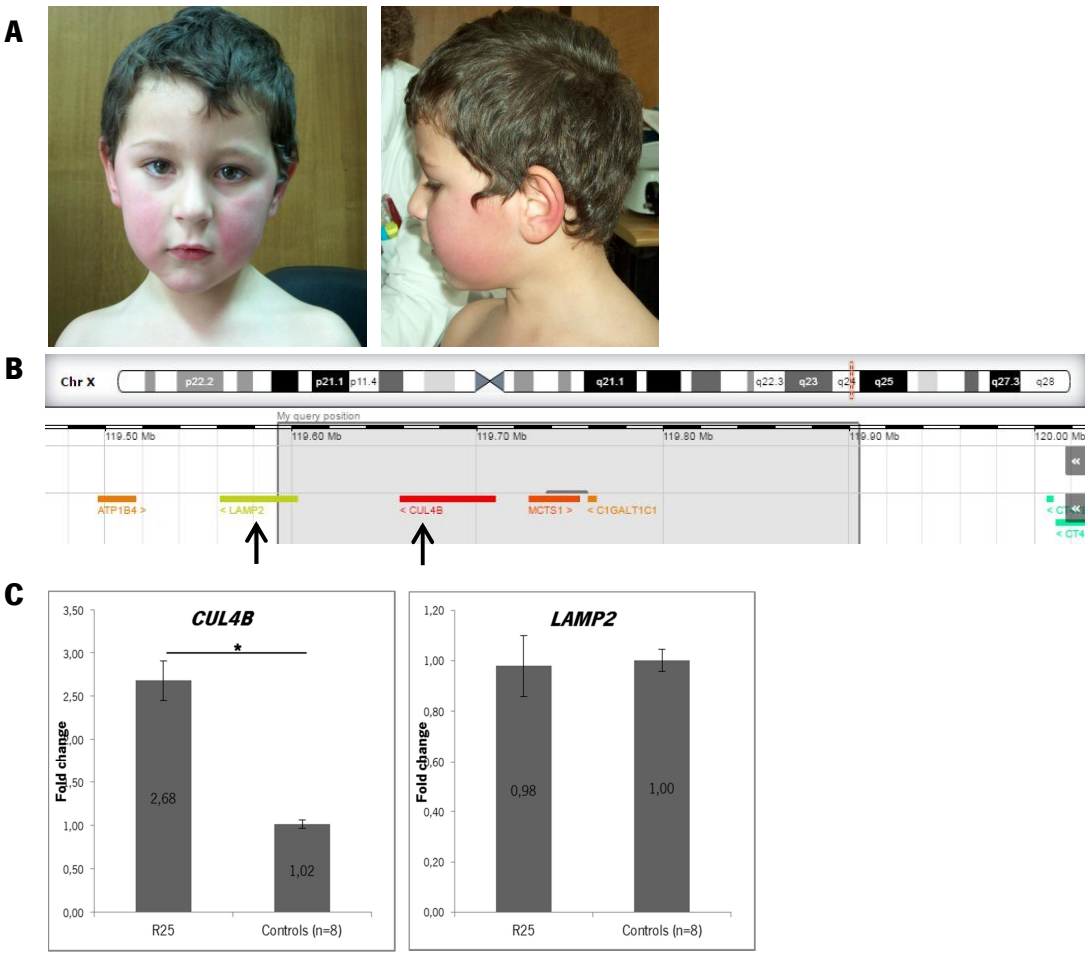


Figure 32 – Facial appearance of patient R25 (A), the Xq24 duplicated region (B) and the mRNA expression level of *CUL4B* and *LAMP2* genes. The black arrows indicate the place where the primers for mRNA expression were designed. *CUL4B* gene is overexpressed in the patient while *LAMP2* gene does not present altered expression at the mRNA level (B2M and PPIB were used as housekeeping genes; student t-test; $p<0.05$*).

Xq26.3 duplication

Patient C24 is a 14 years old boy with a 570.1Kb duplication at Xq26.3 inherited from his mother (not determined if affected or not) (figure 27). He has mild ID, speech delay, dolichocephaly and several dysmorphisms, including micrognatia, syndactyly and clinodactyly. His younger sister also carries the duplication and, and despite presenting some facial dysmorphisms, has no ID and has normal development for her age.

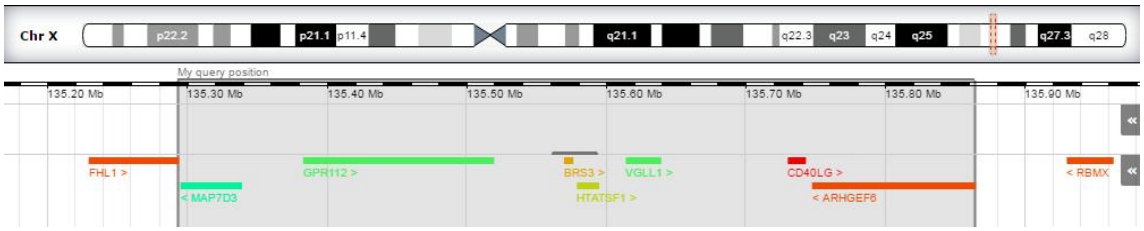


Figure 33 – Xq26.3 duplicated region in patient C24 and the genes affected.

The duplication encompasses the following genes: *FHL1* (*four and a half LIM domains 1*), *MAP7D3* [*MAP7 (microtubule-associated protein 7) domain containing 3*], *GPR112* (*or ADGRG4 - adhesion G protein-coupled receptor G4*), *BRS3* (*bombesin like receptor 3*), *HTATSF1* (*HIV-1 Tat specific factor 1*), *VGLL1* (*vestigial like family member 1*), *MIR934*, *LINC00892, CD40LG* (*CD40 ligand*) and *ARHGEF6* [*Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6*] genes.

ARHGEF6 encodes for a protein that belongs to a family of cytoplasmic proteins which activate the Rho proteins by exchanging bound GDP for GTP. These Rho GTPases play a fundamental role in numerous cellular processes linked to the organization of the cytoskeleton, cell shape, and motility (Murali and Rajalingam, 2014). *ARHGEF6* specifically has been implicated in the regulation of spine morphogenesis and LoF mutations have been found in patients with X-linked ID (Kutsche *et al.*, 2000; Nodé-Langlois *et al.*, 2006). Additionally, a 2.8 Mb duplication in Xq26.2-Xq26.3 was described in two brothers with ID and the *ARHGEF6*, *PHF6*, *HPRT1* and *SLC9A6* genes have pointed out as potential contributors to their patients' phenotype (Madrigal *et al.*, 2010). When compared to this publication, we can see that our patient's duplication is smaller and affects only the *ARHGEF6* gene; nevertheless, the phenotypic similarities between our patient and those described by Madrigal and colleagues (namely the ID, dolichocephaly and facial dysmorphisms) suggest a determinant role for *ARHGEF6* gene in phenotypes associated with Xq26 microduplications (Madrigal *et al.*, 2010). Expression data in the periphery for some of the genes involved in the duplication didn't retrieve results that we could interpret.

Variants of unknown significance (VOUS)

In the VOUS group, we included variants which did not encompass a known CNV region, and/or were described in control databases, and/or were inherited from a parent for whom the clinical presentation was not known, and/or for which pathogenicity was not strongly supported by biological data. In this group, only in six cases the CNV arise *de novo*. This number does not reflect the real number of inherited *de novo* CNVs but instead is due to the unavailability of parent's samples to test inheritance. This was particularly noticed in samples from the CC, since parental DNA was available in all cases for the RC. Regarding the VOUS, several patients presented more than one CNV (4 patients with 2 CNVs, 3 patients with 3 CNVs and 1 patient with 4 CNVs). In the cases in which the patient had CNVs inherited from both the mother and the father there is the question of whether these alterations together could lead to the disease in the child. The list of all the VOUS variants is possible to see in table V.

Table V – List of variants of unknown clinical significance (VOUS)

Patient Cohort	ISCN description (Hg19)	Type	Size (Kb)	Genes (nº)	Genes (name)	Confirmation	Inheritance	DGV controls	Similar case (Decipher)	Array platform
R27 (♂)	1p13.2(112,243,130-112,331,235)X3	dup	88	8	*AK023457, AK092511, BC041890, C1orf183, DDX20, DP103, KCND3, RAP1A*	qPCR	*de novo*	No	No	1
R28 (♂)	1p34.3(36,775,225-369,17,965)X3	dup	143	8	*LSM10, OSCP1, SH3D21, STK40*	NP	ND	1/270 (del)	No	1
C25 (♀)	1q44(245,132,097-245,259,567)x3	dup	127.5	1	*EFCAB2*	NP	ND	6/46505	265207	2
C26 (♂)	2q12.3q13(109,269,051-110,504,320)x3	dup	1240	9	*LIMS1 (PINCH); RANBP2; EDAR; CCDC138; SH3RF3; SEPT10; SOWAHC; MIR4265; SH3RF3-AS1*	NP	ND	No	263424	3
C27 (♂)	2q21.1(131,592,472-131,886,566)x1	del	294.1	3	*ARHGEF4; FAM168B (MANI)*	qPCR	Maternal¥	40/52795	2311, 253247, 263742, 281771, 284907(all slightly bigger)	2
C28 (♀)	2q31.2q31.3(179,933,642-180,709,664)x3	dup	776	2	*SESTD1; ZNF385B*	NP	Paternal¥	No	No	3
R29 (♀)	2q33.1(203,338,376-203,513,494)X3	dup	175	2	*BMPR2, FAM117B*	NP	ND	No	No	1
C29 (♂)	3q23(141,021,128-141,154,103)x3	dup	130	1	*ZBTB38 (CIBZ)*	NP	ND	2/18978 (smaller)	No	3
R30 (♂)	3q26.33(181,357,672-181,466,211)X1	del	108	2	*SOX2; SOX2OT*	qPCR	*de novo*	No	301549 (smaller dup)	1
C30 (♀)	4p16.2(5,468,286-5,812,963)x1	del	344.7	4	*EVC2; EVC; STK32B; C4orf6*	NP	ND	1/17421 (slightly bigger)	No	2
C31 (♂)	5q11.2(54,433,299-57,129,848)x3	dup	2607	20	*CDC20B; GPX8; MIR449A; MIR449B; MIR449C; CCNO; DHX29; SKIV2L2; PPAP2A; RNF138P1; SLC38A9; DDX4; IL31RA; IL6ST; ANKRD55; MAP3K1; C5orf35; MIER3; GPBP1; ACTBL2*	NP	Maternal¥	No	No	2
C32 (♂)	5q23.1q23.2(119,470,295-121,553,124)x1	del	2100	6	*PPR16; FTMT; ZNF474; SRFBP1; LOX; LOC100505841*	NP	ND	Smaller CNVs; none comprising all the genes	No	3
C33 (♂)	5q31.1(132,540,588-132,890,257)x3	dup	349.7	2	*FSTL4; MIR1289*	NP	ND	1/17421	No	3
C34 (♂)	6p22.3(18,596,430-22,732,054)x1	del	4136	7+2	*CDKAL1 ; ID4; SOX4; E2F3; PRL; MBOAT1; HDGFL1; LOC729177; LINC00340*	NP	Maternal¥	No	249613	2

(Cont.)

C35 (♂)	7q36.1(151,768,386-152,077,451)x3	dup	309.1	2	*KMT2C (MLL3); GALNT11*	NP	Paternal¥	4/29957	248263	3
C36 (♂)	8p23.1(9,687,615-10,112,430)x3	dup	424.8	3	*LINC00599; MIR124; MSRA*	NP	Paternal¥	No	No	2
R31 (♂)	9p24.1(6,668,082-7,112,536)X1	del	444	1	*KDM4C*	NP	ND	2/ 29084	No	1
R32 (♂)	10p11.23(30,659,736-30,761,192)X3	dup	101	7	*MAP3K8*	qPCR	Paternal¥	1/369	No	1
R33 (♂)	10q24.2(100,014,985-100,100,985)X3	dup	86	1	*LOXL4*	NP	ND	No	No	1
R34 (♂)	11p11.2(44,601,486-44,779,120)X3	dup	178	1	*CD82*	NP	ND	No	No	1
R35 (♂)	11q23.3(119,415,826-119,560,414)X3	dup	145	1	*PVRL1*	NP	ND	No	No	1
R36 (♂)	13q32.3(99,451,824-99,530,240)X3	dup	78	3	*DOCK9*	NP	ND	No (only smaller dup)	No	1
R37 (♂)	14q24.2(71,814,635-71,927,259)X3	dup	113	1	*SNORD56B*	NP	ND	No	No	1
C37 (♂)	14q32.32-q32.33(103,997,076-105,608,966)x3	dup	1610	27	*TRMT61A; BAG5; C14orf153; KLC1, XRCC3, ZFYVE21, PPP1R13B, C14orf2, TDRD9, ASPG, MIR203, KIF26A, C14orf180, TMEM179, INF2, ADSSL1, SIVA1, AKT1, ZBTB42, MGC23270, KIAA0284, PLD4, AHNAK2, C14orf79, CDCA4, GPR132, JAG2*	NP	Paternal¥	1/29084 (del bigger)	UOM272256 but has additional alterations	2
C38 (♂)	15q26.3(100,269,795-100,956,135)x3	dup	686	5	*LYSMD4, DNM1P46, ADAMTS17, FLJ42289, CERS3*	NP	Paternal¥	1/29084 (bigger; several smaller dups)	331128	4
C39 (♀)	16q11.2q12.1(46,906,585-47,199,337)x3	dup	292.8	4	*GPT2, DNAJA2, ITFG1, NETO2*	NP	*de novo*	1/29084 (bigger)	No	3
C40 (♂)	17p13.3(2,287,362-2,287,555)x3	dup	194	1	*MNT*	NP	ND	5/46874 (bigger)	No	2
R38 (♂)	17p13.1(6,955,115-7,409,331)X1 dn	del	454	30	*DLG4, GABARAP, DULLARD, NEURL4, NLGN2, CHRNB1*	NP	*de novo*	1/181 1/1557 1/17421	260507, 2346, 3474 all smaller del)	1
R39 (♀)	17q11.2(27,429,294-27,516,778)X3	dup	87	1	*MYO18A*	qPCR	Paternal¥	No	No	1
R40 (♂)	17q21.31(43,696,388-43,979,132)X3	dup	283	7	*MAPT*	NP	ND	1/ 29084 (smaller)	No	1
R41 (♂)	19p13.2(7,077,066-7,727,437)X3 mat	dup	605	14	*KHSRP, PSPN, TUBB4, STXBP2*	NP	Maternal	6/771 1/29084	253443	1
	19p13.3-p13.2(6,332,716-6,993,284)X3 mat	dup	660	22	*PNPLA6, ARHGEF18, KIAA1543*	NP	Maternal	3/181 (half the size)	253443	
R42 (♀)	19q13.12(37,775,477-37,942,465)X3	dup	167	3	*HKR1, ZNF527, ZNF569*	NP	ND	1/270 (del)	1989	1

(Cont.)

C41 (♂)	19q13.43(56,549,717-57,146,408)x3	dup	596.7	14	*NLRP5, ZNF787,ZNF444, GALP, ZSCAN5B, ZSCAN5A, ZNF542, ZNF582, ZNF583, ZNF667, ZNF471, ZFP28, ZNF470, ZNF71*	NP	Maternal¥	No	259335	2
R43 (♂)	19q13.43(58,443,388-58,669,835)X3	dup	226	8	*C19orf18, ZNF135, ZNF256, ZNF329, ZNF418, ZNF606, ZSCAN1, ZSCAN18*	NP	*de novo*	1/ 29084	289634	1
R44 (♀)	Xp11.21-p11.1(56,304,820-56,964,968)X3	dup	660	4	*KLF8, UBQLN2*	NP	ND	No	274061; 307669	1
C42 (♂)	Xp21.1(32,826,352-33,936,518)x3	dup	1110	1	*DMD*	NP	Maternal¥	No	No	3
C43(♀)	Xp22.33(480,164-785,059)x3	dup	304.9	1	*SHOX*	NP	ND	No	279033 (but has more variants)	3
C44 (♀)	4p16.1(8,080,960-8,416,608)x3	dup	335.6	4	*ABLIM2; SH3TC1; HTRA3; ACOX3*	NP	ND	1/29084	No	2
	12p13.33(2,802,013-3,123,690)x3	dup	321.7	8	*CACNA1C; FKBP4; ITFG2; NRIP2; FOXM1; C12orf32; TULP3;TEAD4*	NP	ND	Smaller CNVs; none comprising all the genes	UOM272277	
R45 (♂)	Xq28(153,230,586-153,282,378)X2	dup	52	3	*HCFC1, IRAK1*	NP	ND	1/265 (del)	No	1
R46 (♂)	5q23.3(128,758,178-129,350,165)X3	dup	592	2	*ADAMTS19, CHSY3*	NP	ND	2/6533 (smaller)	No	1
	9q33.3(128,474,074-128,515,941)X1	del	419	1	*PBX3*	NP	ND	No	No	1
R47 (♂)	6q21(108,431,203-108,722,841)X3	dup	291	5	*AF520419, LACE1, NR2E1, SNX 3A, SNX3*	NP	Maternal¥	No	No	1
	9p24.1(6,802,781-6,943,275)X1	del	140	3	*JMJD2C, KDM4C, KIAA0780*	NP	Paternal	2/29084	No	1
	16q24.3(89,867,584-89,916,614)X3	dup	49	2	*FANCA, SPIRE2*	NP	Maternal¥	1/1557; 1/17421	307678 (del)	1
	Xp22.33(1,373,567-1,430,376)X2	dup	56	1	*CSF2RA*	NP	Maternal¥	1/2 (del)	No	1
R48 (♂)	3p21.31(48,464,967-48,574,235)X3	dup	109	7	*PLXNB1*	NP	Maternal¥	1/ 17421 (del)	No	1
	7p22.3(1,565,982-1,701,871)X3	dup	136	5	*MAFK*	NP	Paternal	1/270 (del)	284772	1
	10p12.31(20,641,191-21,122,699)X3	dup	481	1	*NEBL, PLXDC2*	NP	Maternal¥	No	289047 (smaller)	1
	14q31.3(88,794,387-88,853,440)X3	dup	59	1	*SPATA7, KCNK10*	NP	*de novo*	No	No	1

(Cont.)

R49 (♀)	1p22.1(92753417-92916646)X3	dup	163	2	*GLMN, RPAP2*	NP	Paternal ¥	1/ 17421 (smaller)	284999, 276287 (partially overlap)	1
	8q21.11(76,470,859-77,036,939)X3	dup	566	1	*HNF4G*	NP	Maternal ¥	1/ 17421 1/3017	No	1
	9q21.13(78,311,144-78,695,190)X1	del	384	2	*PCSK5*	NP	Paternal ¥	No	No	1
	15q21.3(57,639,792-58,142,922)X3	dup	503	4	*CGNL1, GRINL1A*	NP	Maternal ¥	1/ 29084	No	1
R50 (♂)	18q12.1(29,316,291-29,569,853)X1	del	254	2	*TRAPPC8, SLC25A52*	NP	Paternal ¥	No	No	1
	20q11.23(36,531,120-36,618,758)X3	dup	88	2	*VSTM2L*	NP	*de novo*	1/ 29084 (del)	No	1
R51 (♀)	5p15.1(16,112,927-16,260,219)X3	dup	147	1	*MARCH11*	NP	ND	No	No	1
	13q31.1(84,644,861-84,723,563)X3	dup	79	1	*MIR548F1*	NP	ND	1/ 29084	No	1
R52 (♂)	9q34.3(137,932,744-138,316,317)X3	dup	383	1	*OLFM1*	NP	ND	1/ 29084; 1/ 17421	299027 (smaller)	1
	Xq28(153,130,545-153,282,378)X2	dup	151	8	*IRAK1, L1CAM*	NP	ND	1/265 (del)	323738	1
R53 (♀)	10q25.3(118,404,726-119,052,432)X3	dup	647	7	*KCNK18, SLC18A2, VAX1*	NP	ND	No	No	1
	10q26.11(119,297,989-119,351,151)X3	dup	53	2	*EMX2, EMX2OS*	NP	ND	1/ 29084; 1/31	No	1
	Xp22.33(1,549,311-1,641,335)X3	dup	92	2	*ASMTL, P2RY8*	NP	ND	2/1 (del)	306493, 288916, 288492, 288579	1
R54 (♀)	11q23.3(119,415,826-119,546,072)X3	dup	130	1	*PVRL1*	NP	ND	No	No	1
	13q31.3(92,352,559-92,668,273)X1	del	315	1	*GPC5*	NP	ND	1/ 17421 (smaller) 1/3017 (smaller)	301689	1

Legend: Patients R26to R60: from research cohort; Patients C26 to C53: from clinical cohort; dup: duplication; del: deletion; NP: not performed; ND: not determined; .Array platform 1: Agilent 180K; 2: KaryoArray®v3.0 (Agilent 8x60k); 3: Affymetrix CytoScan HD array; 4: Affymetrix CytoScan 750K; ¥: presumably healthy

DISCUSSION

This study of a cohort of ID patients in whom most common causes of disease had been excluded allowed us to find a reliable cause of disease in 9.3% of patients and to propose novel candidate ID *loci* in 5.1%. The pathogenic variants associated with known syndromes were essentially large CNVs, most of them arising *de novo*, and affecting in general dozens of genes. Patients carrying these CNVs usually presented the major clinical features established for a particular syndrome. However, even these well-established pathogenic CNVs can be associated with a broad phenotypic presentation, as observed in patients R6 and C4, both with WBS associated variants but not presenting the full-blown phenotype for this syndrome, or C11 and R16, both with 9q34 gains affecting the *EHMT1* gene and presenting overlapping features with Kleefstra syndrome. In this note, we believe that the main contributions of this work are: **(I)** the reporting of new patients with CNVs in regions known to be associated with known syndromes but with different clinical presentations; **(II)** the reporting of novel candidate ID-causative *loci* at 2q11.2-q12.2 (del), 7q33 (del7dup), 10q26.3 (del), 17p11.2 (del), 20q13.12-q13.13 (del), 1p22.1-p21.3 (dup), 2p15 (dup), 9q33.2-q33.3 (tri), 12p13.33 (dup), Xq24 (dup) and Xq26.3 (dup); **(III)** the study in patients with copy number gains of the mRNA expression at the periphery for genes located either inside the duplicated/triplicated regions and/or at the breakpoints, making it possible to determine if there is an actual effect of gene dosage at the transcription level.

Comparison of yield of diagnosis using aCGH in a research vs. a clinical scenario

The yield for known pathogenic CNVs was slightly higher in the CC (11.7%) than in the RC (8.5%). This can be due to the fact that all the patients included in the RC were at the end of the diagnostic road and the most common genetic causes of NDDs had been already excluded before their incorporation in the study. Even though many of the patients from the CC have performed other molecular analyses before aCGH, this was not the case for every single one of the patients, as it was for in the RC. Considering the likely pathogenic variants and VOUS, the yields were similar in both cohorts. Many of the variants here detected by aCGH were rare and restricted to one patient/family. Therefore they have been, accordingly, classified as VOUS and their clinical significance needs to be carefully addressed in future studies. Additionally, rare intermediate-size CNVs, relatively frequent in the general population, and not necessarily assigned *a priori* as pathogenic, have been recently associated with ID and negatively with

educational attainment (Mannik *et al.*, 2015), and so even these should not be excluded as cause of disease but rather re-assessed in the face of accumulating information to establish useful genotype-phenotype correlations.

mRNA expression studies of genes affected by copy number gains

The expression analysis performed for some of the novel duplications here proposed to be ID-causative aimed to determine the effect of the CNV in gene expression focusing either in key genes within the CNV or in genes located in or near the breakpoint region. While a deletion most probably causes haploinsufficiency, the effect of a duplication is not so obvious, as in these cases disease can be caused by triplosensitivity, gene disruption or gene fusion breakpoints (Newman *et al.*, 2015). What lead us to study the expression of genes affected by the duplication was the fact that intragenic duplications were reported to often disrupt the gene reading frame leading to a loss of function and consequently to haploinsufficiency (Newman *et al.*, 2015). Based on this, we hypothesized that if a duplication was located in tandem this could lead to expression alterations in the genes located in the breakpoints, nearby and inside the duplicated region. We concluded that for all the genes clearly located inside the duplicated region their expression was always increased when compared to healthy adult controls. This finding should be interpreted with care and be subjected to interpretation only in the studied genes and in the present duplications. These findings cannot be extrapolated to other overlapping duplications (even if they affect the same genes studied) as the breakpoint location might be different and possibly lead to different impact in gene expression. Another important point to mention is that these results concern the gene's expression in the periphery and a different situation can occur in brain.

As for patient R16, in whom the whole *EHMT1* gene (9q34 duplications) is duplicated, the expression was increased when compared to healthy adult controls. This finding thus reinforce the idea advanced by Yatsenko and colleagues, that increased dosage of *EHMT1* may be responsible for neurodevelopmental impairments (Yatsenko *et al.*, 2012). *TSC1* gene expression was also increased in patient R16 being also a good candidate for the phenotype. In a fibroblast cellular model, overexpression of hamartin (*TSC1* gene product) lead to growth inhibition and changes in cell morphology (Benvenuto *et al.*, 2000). In neurons, overexpression of Tsc1/Tsc2 results in suppression of axon formation (Choi *et al.*, 2008). Since cell morphology adaptation and axon growth are essential for brain development and function, we can hypothesize that *TSC1* overexpression might also be a contributor for the patient's phenotype. Because we know that

the 9q34 duplicated region is not located in tandem (based on the FISH studies), these finding are in accordance with the hypothesis that if duplication is located in *trans* is likely that there will be no structural effect in the region and the genes will not suffer an influence in expression.

As for the 14q32 duplicated region, the *INF2* mRNA expression was also found to be increased in the patient, which is an observation in accordance with the fact that the entire gene is located inside the duplicated region. On the other hand, *TECPR2* expression is not altered in the patient. The portion of the transcript where the primers for *TECPR2* were designed is located outside the duplicated region. Nevertheless, we believe that if the duplicated region affected the expression of the gene this would be still possible to observe, since one of the alleles (the one located in the duplicated chromosome) would possibly result in the degradation of the entire mRNA molecule.

One of the limitations that we found using this approach was the fact that not all the genes located in the breakpoints were possible to analyse at the expression level because they presented very low expression in the periphery. When we were faced with this situation we need to move to a second choice which often was a gene located as close as possible to the breakpoint with considerable expression in peripheral blood and/or interesting function for ID (this last criteria being only applied when possible).

The increased expression observed for *FBXW2*, *NEK6* and *PSMB7* genes in patient R23 is consistent with the presence of a triplication in *9q33.2-q33.3*. All of them, as mentioned before, may contribute to the phenotype of the patient. For *NEK6* these findings are in accordance with the fact it is included inside the triplicated region and, therefore being overexpressed. As for *FBXW2* and *PSMB7* we had hypothesized that their expression could be negatively affected by the position of the duplicated region (since they are located at the breakpoints), which we could observe not to be the case.

As for patient R24 (12p13.33-p13.32 duplication), *TSPAN9* gene expression in periphery was not altered, leading us to believe that a compensatory mechanism can occur in order to compensate for the extra copy of *TSPAN9* (since the gene in completely located inside the duplication. Still in this region we tried to analyze the *CACNA1C* gene (located in the breakpoint) expression. However, *CACNA1C* expression in the periphery is very low, making it impossible to analyze.

CUL4B was entirely duplicated in patients R25 and R26 and its expression in peripheral blood cells was increased (study only performed in patient R25), leading to us to believe that the disease in this patient is in fact driven by an increased *CUL4B* dosage. To our knowledge, this was the first time that *CUL4B* overexpression was determined in an ID patient. Also, *LAMP2* (the

gene located in the duplication breakpoint) does not present altered expression, which might indicate that the gene is not disrupted by the duplication, and possibly that the duplicated region is inserted somewhere else in the genome.

The likely pathogenic CNVs here proposed as novel candidate *loci* for ID encompass several genes that either were already associated with NDDs or were proposed to have a role in ID and which can be grouped according to their function in several cellular aspects.

Transcriptional factors/cell cycle regulators/DNA repair proteins (POU3F3, EBF3, LHX2, LHX6, TULP3, FOXM1, ARID5A)

Transcriptional regulation is an essential component of the neuronal differentiation programs and of the response to stimulation patterns underlying neuronal plasticity. Therefore, it is not surprising that genes involved in transcription regulation are known to associate to several cancers, due to their roles in cell proliferation (Wang and Baker, 2015), but have also been implicated (usually in the reverse manner) in well-known NDDs, as is the case of *FOXL2* in BPES (D'haene *et al.*, 2009; Verdin and Baere, 2012; Zahanova *et al.*, 2012) and *BAZ1B* in WBS (Lalli *et al.*, 2016). More recently, *de novo* mutations in *EBF3,* were described by several groups as a cause of a neurodevelopmental syndrome, that includes ID, abnormal genitalia, and structural central nervous system (CNS) malformations (Chao *et al.*, 2017; Harms *et al.*, 2017; Sleven *et al.*, 2017). Therefore, genes affecting transcription could play a role in NDD phenotypes.

Chromatin modifiers/chromatin remodeling proteins (known genes: EHMT1, ARID1B)

An excess of mutation genes encoding proteins involved in chromatin regulation have been described in NDDs (Lasalle, 2013; Sajan *et al.*, 2017). *EHMT1* and *ARID1B* belong to this category and are known to be associated with ID for many years. Here we describe two more patients with duplications affecting the *EHMT1* gene, in one of which it was possible to show *EHMT1* overexpression in the periphery. The patient carrying the *ARID1B* deletion helped for the association of *ARID1B* haploinsufficiency with Coffin-Siris syndrome (Wieczorek *et al.*, 2013).

Ubiquitin signaling (CUL4B, CNOT4, UBE2C, NEURL3, FBXW2)

Ubiquitin-mediated degradation of proteins is a crucial mechanism for cell maintenance and viability, but ubiquitin signaling is also relevant for subcellular localization activation or exocytosis of key neuronal proteins (Clague *et al.*, 2012). Several genes belonging to this pathway are described to be associated with NDDs, as is the case of *CUL4B* (Tarpey *et al.*, 2007), shown here to be duplicated in patients R25 and R26. ID can be caused by deletions affecting genes

encoding for several types of ubiquitin-conjugating enzymes, such as *UBE2A, UBE2C* and *UBE3A* (De Leeuw *et al.*, 2010; LaSalle *et al.*, 2015; Bruinsma *et al.*, 2016). *NEURL3* and *CNOT4* encode for proteins with E3 ubiquitin-protein ligase activity; as for *FBXW2*, it encodes for an F-box protein and these proteins are one of the four types of subunits of SCF ubiquitin-protein ligases. Our findings thus reinforce the idea that genes encoding for proteins belonging to the ubiquitin-proteasome system are possible new candidate genes for NDD phenotypes.

Cytoskeleton regulation and organization, cell shape and motility (USP34, TSC1, ARHGEF6)

Several NDDs, including many severe epilepsy syndromes, are caused by mutations in genes regulating neuronal migration, which often encode for proteins involved in the function of the cytoskeleton. Cell division and axon/dendrite formation, which similarly depend on cytoskeletal functions, may also be affected (Stouffer *et al.*, 2015). We described one patient with a duplication encompassing *ARHGEF6* gene. *ARHGEF6* was for a long time associated with X-linked ID (XLID) (Kutsche *et al.*, 2000), but lately its role as a monogenic cause of XLID was considered questionable (Piton *et al.*, 2013). The duplication present in patient C24 encompasses other genes and for this reason *ARHGEF6* cannot be considered the cause of disease in the patient (other genes involved might be playing a role alone or together with *ARHGEF6*). Nevertheless, the description of one more patient with a genomic imbalance affecting the gene might be useful for elucidating its role in ID.

Intracellular vesicular trafficking and exocytosis (ARFGEF2, EXOC4, EXOC6B)

Disruption of intracellular vesicular trafficking can be related with several conditions. In this work we report a patient with a deletion encompassing *ARFGEF2,* previously described associated with epilepsy and ID (in the case of homozygous mutations) (Banne *et al.*, 2013; Yilmaz *et al.*, 2016). The collection of patients here presented also allowed the description for the first time of *EXOC6B* gene haploinsufficiency in association with DD/ID (reported in detail in a dedicated publication) (Wen *et al.*, 2013).

Signaling mediators/transducers/receptor activity/transmembrane proteins (B9D1, GPR45, TNFRSF13B, CSE1L, SPATA2, FAM69A, SEMA4C, CACNA1C, AHSA2, AKT3, STX1A)

Disruption of synaptogenesis has been associated with ID and NDD (Yoshida *et al.*, 2011). Several genes associated with synaptogenesis were here described in different patients/CNVs, highlighting once again the crucial contribution of proper synapse development and function for ID/DD. Proteins as syntaxins and semaphorins play a role in this process. Syntaxins act in the vesicle fusion process while semaphorins regulate cell motility and play crucial roles in axon

guidance during synaptogenesis as well as in immune cell regulation and tumor progression (Kruger *et al.*, 2005).

aCGH technology has for long been used in the research and clinical contexts allowing the delineation of many new microdeletion and microduplication syndromes. In the last decade it was possible to observe the decrease in the rate at which new syndromes were described, being that less distinct microdeletion and microduplication syndromes were published, making it more difficult to publish isolated or fewer cases (van Ravenswaaij-Arts and Kleefstra, 2009). We believe that publications of this nature are important for the ID field as they allow the sharing of patients with very rare (often private) CNVs that many times may reveal candidate genes for disease.

References

Allaire PD, Ritter B, Thomas S, Burman JL, Denisov AY, Legendre-Guillemin V, et al. Connecden , A Novel DENN Domain-Containing Protein of Neuronal Clathrin-Coated Vesicles Functioning in Synaptic Vesicle Endocytosis. J. Neurosci. 2006; 26: 13202–13212.

Allen AS, Berkovic SF, Cossette P, Delanty N, Dlugos D, Eichler EE, et al. De novo mutations in the classic epileptic encephalopathies. Nature 2013; 501: 217–221.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-5 (5th ed.). Arlington, VA: American Psychiatric Publishing.; 2013.

Antzelevitch C, Brugada P, Borggrefe M, Brugada J, Brugada R, Corrado D, et al. Brugada Syndrome Report of the Second Consensus Conference. Circulation 2005; 111: 659–670.

Banne E, Atawneh O, Henneke M, Brockmann K, Gärtner J, Elpeleg O, et al. West syndrome, microcephaly, grey matter heterotopia and hypoplasia of corpus callosum due to a novel ARFGEF2 mutation. J. Med. Genet. 2013; 50: 772–775.

Bardoni B, Abekhoukh S, Zongaro S, Melko M. Intellectual disabilities, neuronal posttranscriptional RNA metabolism, and RNA-binding proteins: Three actors for a complex scenario. Prog. Brain Res. 2012; 197: 29–51.

Baynam G, Overkov A, Davis M, Mina K, Schofield L, Allcock R, et al. A germline MTOR mutation in Aboriginal Australian siblings with intellectual disability, dysmorphism, macrocephaly, and small thoraces. Am J Med Genet Part A 2015; 167: 1659–1667.

Benvenuto G, Li S, Brown SJ, Braverman R, Vass WC, Cheadle JP, et al. The tuberous sclerosis-1 (TSC1) gene product hamartin suppresses cell growth and augments the expression of the TSC2 product tuberin by inhibiting its ubiquitination. Oncogene 2000; 19: 6306–6316.

Bruinsma CF, Savelberg SMC, Kool MJ, Jolfaei MA, Van Woerden GM, Baarends WM, et al. An essential role for UBE2A/HR6A in learning and memory and mGLUR-dependent long-term depression. Hum. Mol. Genet. 2016; 25: 1–8.

Chao H-T, Davids M, Burke E, Pappas JG, Rosenfeld JA, McCarty AJ, et al. A Syndromic Neurodevelopmental Disorder Caused by De Novo Variants in EBF3. Am. J. Hum. Genet. 2017; 100: 1–10.

Chen Z, Schwahn BC, Wu Q, He X, Rozen R. Postnatal cerebellar defects in mice deficient in methylenetetrahydrofolate reductase. Int J Dev Neurosci 2005; 23: 465–474.

Choi S-Y, Pang K, Kim JY, Ryu JR, Kang H, Liu Z, et al. Post-transcriptional regulation of SHANK3 expression by microRNAs related to multiple neuropsychiatric disorders. Mol. Brain 2015; 8: 1–12.

Choi Y-J, Di Nardo A, Kramvis I, Meikle L, Kwiatkowski DJ, Sahin M, et al. Tuberous sclerosis complex proteins control axon formation. Genes Dev. 2008; 22: 2485–2495.

Clague MJ, Coulson JM, Urbé S. Cellular functions of the DUBs. J. Cell Sci. 2012; 125: 277–286.

Coe BP, Girirajan S, Eichler EE. The genetic variability and commonality of neurodevelopmental disease. Am. J. Med. Genet. C Semin. Med. Genet. 2012; 160 C: 118–129.

Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. Nat. Genet. 2011; 43: 838–846.

Costa-Mattioli M, Monteggia L. mTOR complexes in neurodevelopmental and neuropsychiatric disorders. Nat Neurosci 2013; 16: 1537–43.

Delio M, Pope K, Wang T, Samanich J, Haldeman-Englert CR, Kaplan P, et al. Spectrum of Elastin Sequence Variants and Cardiovascular Phenotypes in 49 Patients With Williams – Beuren Syndrome. Am. J. Hum. Genet. Part A 2013; 161A: 527–533.

D'haene B, Attanasio C, Beysen D, Dostie J, Lemire E, Bouchard P, et al. Disease-causing 7.4 kb Cis-Regulatory Deletion Disrupting Conserved Non-Coding Sequences and Their Interaction with the FOXL2 Promotor: Implications for Mutation Screening. PLOS Genet. 2009; 5: e1000522.

D'haene B, Vandesompele J, Hellemans J. Accurate and objective copy number profiling using real-time quantitative PCR. Methods 2010; 50: 262–270.

Dheedene A, Maes M, Vergult S, Menten B. A de novo POU3F3 Deletion in a Boy with Intellectual Disability and Dysmorphic. Mol. Syndromol. 2014; 5: 32–35.

Dominguez MH, Ayoub AE, Rakic P. POU-III Transcription Factors (Brn1, Brn2, and Oct6) Influence Neurogenesis, Molecular Identity, and Migratory Destination of Upper-Layer Cells of the Cerebral Cortex. Cereb. Cortex 2013; 23: 2632–2643.

Dudkiewicz M, Lenart A, Pawłowski K. A Novel Predicted Calcium-Regulated Kinase Family Implicated in Neurological Disorders. PLoS ONE 2013; 8: 1–10.

Elsea SH, Williams SR. Smith – Magenis syndrome: haploinsufficiency of RAI1 results in altered gene regulation in neurological and metabolic pathways. Expert Rev. Mol. Med. 2011; 13: 1–14.

Esplin ED, Li B, Slavotinek A, Novelli A, Battaglia A, Clark R, et al. Nine patients with Xp22.31 microduplication, cognitive deficits, seizures, and talipes anomalies. Am. J. Med. Genet. A. 2014; 164: 2097–2103.

Faletra F, D'Adamo AP, Rocca MS, Carrozzi M, Perrone MD, Pecile V, et al. Does the 1.5 Mb microduplication in chromosome band Xp22.31 have a pathogenetic role? New contribution and a review of the literature. Am. J. Med. Genet. A. 2012; 158A: 461–464.

Ferrero GB, Howald C, Micale L, Biamino E, Augello B, Fusco C, et al. An atypical 7q11.23 deletion in a normal IQ Williams-Beuren syndrome patient. Eur. J. Hum. Genet. EJHG 2010; 18: 33–8.

Flandin P, Zhao Y, Vogt D, Jeong J, Long J, Potter G, et al. Article Lhx6 and Lhx8 Coordinately Induce Neuronal Expression of Shh that Controls the Generation of Interneuron Progenitors. Neuron 2011; 70: 939–950.

Fregeau B, Kim BJ, Hernández-García A, Jordan VK, Cho MT, Schnur RE, et al. De Novo Mutations of RERE Cause a Genetic Syndrome with Features that Overlap Those Associated with Proximal 1p36 Deletions. Am. J. Hum. Genet. 2016; 98: 963–970.

Friocourt G, Parnavelas J. Identification of Arx targets unveils new candidates for controlling cortical interneuron migration and differentiation. Front. Cell. Neurosci. 2011; 5: 1–12.

Froyen G, Belet S, Martinez F, Santos-Rebouças CB, Declercq M, Verbeeck J, et al. Copy-Number Gains of HUWE1 Due to Replication- and Recombination-Based Rearrangements. Am. J. Hum. Genet. 2012; 91: 252–264.

Frye RE, Cox D, Slattery J, Tippett M, Kahler S, Granpeesheh D, et al. Mitochondrial Dysfunction may explain symptom variation in Phelan-McDermid Syndrome. Sci. Rep. 2016; 6: 1–12.

Halbedl S, Schoen M, Feiler MS, Boeckers TM, Schmeisser MJ. Shank3 is localized in axons and presynaptic specializations of developing hippocampal neurons and involved in the modulation of NMDA receptor levels at axon terminals. J. Neurochem. 2016; 137: 26–32.

Harms FL, Girisha KM, Hardigan AA, Kortum F, Shukla A, Alawi M, et al. Mutations in EBF3 disturb transcriptional profiles and underlie a novel syndrome of intellectual disability, ataxia and facial dysmorphism. Am J Hum Genet 2017; 100: 117–127.

Hirai Y, Tamura M, Otani J, Ishikawa F. NEK6-mediated phosphorylation of human TPP1 regulates telomere length through telomerase recruitment. Genes Cells Devoted Mol. Cell. Mech. 2016; 21: 874–889.

Hoebeeck J, van der Luijt R, Poppe B, De Smet E, Yigit N, Claes K, et al. Rapid detection of VHL exon deletions using real-time quantitative PCR. Lab. Invest. 2005; 85: 24–33.

Hoeffer CA, Klann E. mTOR signaling : At the crossroads of plasticity , memory and disease. Trends Neurosci. 2010; 33: 67–75.

Hopp K, Heyer CM, Hommerding CJ, Henke SA, Sundsbak JL, Patel S, et al. B9D1 is revealed as a novel Meckel syndrome (MKS) gene by targeted exon-enriched next-generation sequencing and deletion analysis. Hum. Mol. Genet. 2011; 20: 2524–2534.

Huang N, Lee I, Marcotte EM, Hurles ME. Characterising and predicting haploinsufficiency in the human genome. PLoS Genet. 2010; 6: e1001154.

Ikeda A, Ikeda S, Gridley T, Nishina PM, Naggert JK. Neural tube defects and neuroepithelial cell death in Tulp3 knockout mice. Hum. Mol. Genet. 2001; 10: 1325–1334.

Isidor B, Pichon O, Baron S, David A, Le Caignec C. Deletion of the CUL4B Gene in a Boy With Mental Retardation , Minor Facial Anomalies , Short Stature , How to Cite this Article : Am. J. Med. Genet. Part A 2009; 152A: 175–180.

Isrie M, Kalscheuer VM, Holvoet M, Fieremans N, Esch H Van, Devriendt K. HUWE1 mutation explains phenotypic severity in a case of familial idiopathic intellectual disability. Eur. J. Med. Genet. 2013; 56: 379–382.

Jiao X, Chen H, Chen J, Herrup K, Firestein BL, Kiledjian M. Modulation of neuritogenesis by a protein implicated in X-linked mental retardation. J. Neurosci. 2009; 29: 12419–27.

Jin Z, Yu L, Geng J, Wang J, Jin X, Huang H. A novel 47.2 Mb duplication on chromosomal bands Xq21.1–25 associated with mental retardation. Gene 2015; 567: 98–102.

Jordan VK, Zaveri HP, Scott DA. 1p36 deletion syndrome: an update. Appl. Clin. Genet. 2015; 8: 189–200.

Junutula JR, Mazie AM De, Peden AA, Ervin KE, Advani RJ, van Dijk SM, et al. Rab14 Is Involved in Membrane Trafficking between the Golgi Complex and Endosomes. Mol. Biol. Cell 2004; 15: 2218–2229.

Kim BJ, Zaveri HP, Shchelochkov OA, Yu Z, Seymour ML, Oghalai JS, et al. An Allelic Series of Mice Reveals a Role for RERE in the Development of Multiple Organs Affected in Chromosome 1p36 Deletions. PLoS Med. 2013; 8: e57460.

Kim JH, Youn BU, Kim K, Moon JB, Lee J, Nam K-I, et al. Lhx2 regulates bone remodeling in mice by modulating RANKL signaling in osteoclasts. Cell Death Differ. 2014; 21: 1613–1621.

Kirov G, Rees E, Walters JTR, Escott-Price V, Georgieva L, Richards AL, et al. The penetrance of copy number variations for schizophrenia and developmental delay. Biol Psychiatry 2014; 75: 378–85.

Kleefstra T, Brunner HG, Amiel J, Oudakker AR, Nillesen WM, Magee A, et al. Loss-of-Function Mutations in Euchromatin Histone Methyl Transferase 1 (EHMT1) Cause the 9q34 Subtelomeric Deletion Syndrome. Am. J. Hum. Genet. 2006; 79: 370–377.

Kleefstra T, Kramer JM, Neveling K, Willemsen MH, Koemans TS, Vissers LELM, et al. Disruption of an EHMT1-Associated Chromatin- Modification Module Causes Intellectual Disability. Am. J. Hum. Genet. 2012; 91: 73–82.

Kleefstra T, Nillesen WM, Houge G, Foulds N, Dooren M Van, Willemsen MH, et al. Further clinical and molecular delineation of the 9q subtelomeric deletion syndrome supports a major contribution of EHMT1 haploinsufficiency to the core phenotype. J. Med. Genet. 2009; 46: 598–606.

Kodaka Y, Tanaka K, Kitajima K, Tanegashima K, Matsuda R, Hara T. LIM homeobox transcription factor Lhx2 inhibits skeletal muscle differentiation in part via transcriptional activation of Msx1 and Msx2. Exp. Cell Res. 2015; 331: 309–319.

Kruger RP, Aurandt J, Guan K. Semaphorins Command Cells To Move. Nat. Rev. Mol. Cell Biol. 2005; 6: 789–800.

Kutsche K, Yntema H, Brandt A, Jantke I, Nothwang HG, Orth U, et al. Mutations in ARHGEF6, encoding a guanine nucleotide exchange factor for Rho GTPases, in patients with X-linked mental retardation. Nat. Genet. 2000; 26: 247–250.

Kuzman MR, Muller DJ. Association of the MTHFR gene with antipsychotic-induced metabolic abnormalities in patients with schizophrenia. Pharmacogenomics 2012; 13: 843–846.

Kyrychenko VO, Nagibin VS, Tumanovska LV, Pashevin DO, Gurianova VL, Moibenko AA, et al. Knockdown of PSMB7 induces autophagy in cardiomyocyte cultures: possible role in endoplasmic reticulum stress. Pathobiol. J. Immunopathol. Mol. Cell. Biol. 2014; 81: 8–14.

Lalli MA, Jang J, Park JHC, Wang Y, Guzman E, Zhou H, et al. Haploinsufficiency of BAZ1B contributes to Williams syndrome through transcriptional dysregulation of neurodevelopmental pathways. Hum. Mol. Genet. 2016; 25: 1294–1306.

Lancaster TM, Foley S, Tansey KE, Linden DEJ, Caseras X. CACNA1C risk variant is associated with increased amygdala volume. Eur. Arch. Psychiatry Clin. Neurosci. 2015; DOI: 10.10

Lasalle JM. Autism genes keep turning up chromatin. OA Autism 2013; 1: 14.

LaSalle JM, Reiter LT, Chamberlain SJ. Epigenetic regulation of UBE3A and roles in human neurodevelopmental disorders. Epigenomics 2015; 7: 1213–1228.

De Laurenzi V, Rogers GR, Hamrock DJ, Marekov LN, Steinert PM, Compton JG, et al. Sjögren–Larsson syndrome is caused by mutations in the fatty aldehyde dehydrogenase gene. Nat. Genet. 1996; 12: 52–57.

Lee AS, Jesús-Cortés H De, Kabir ZD, Knobbe W, Orr M, Burgdorf C, et al. The Neuropsychiatric Disease-Associated Gene cacna1c Mediates Survival of Young. eNeuro 2016; 3: 1–11.

Lee C, Iafrate AJ, Brothman AR. Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. Nat. Genet. 2007; 39: 48–54.

De Leeuw N, Bulk S, Green A, Jaeckle-Santos L, Baker LA, Zinn AR, et al. UBE2A deficiency syndrome: Mild to severe intellectual disability accompanied by seizures, absent speech, urogenital, and skin anomalies in male patients. Am. J. Med. Genet. A. 2010; 152 A: 3084–3090.

Leung TY, Pooh RK, Wang CC. Classification of pathogenic or benign status of CNVs detected by microarray analysis. Expert Rev Mol Diagn 2010; 10: 717–721.

Li F, Shen Y, Köhler U, Sharkey FH, Menon D, Coulleaux L, et al. Interstitial microduplication of Xp22.31: Causative of intellectual disability or benign copy number variant? Eur. J. Med. Genet. 2010; 53: 93–99.

Li J, Zhao L, You Y, Lu T, Jia M, Yu H. Schizophrenia Related Variants in CACNA1C also Confer Risk of Autism. PLoS ONE 2015; DOI:10.137: 1–12.

Lim CH, Zain SM, Reynolds GP, Zain MA, Roffeei SN, Zainal NZ, et al. Genetic association of LMAN2L gene in schizophrenia and bipolar disorder and its interaction with ANK3 gene polymorphism. Prog. Neuropsychopharmacol. Biol. Psychiatry 2014; 54: 157–162.

Madrigal I, Fernández-Burriel M, Rodriguez-Revenga L, Cabrera JC, Martí M, Mur A, et al. Xq26.2-q26.3 microduplication in two brothers with intellectual disabilities: clinical and molecular characterization. J. Hum. Genet. 2010; 55: 822–826.

Malhotra D, Sebat J. CNVs: Harbingers of a rare variant revolution in psychiatric genetics. Cell 2012; 148: 1223–1241.

Mannik K, Magi R, Macé A, Cole B, Guyatt AL, Shihab HA, et al. Copy Number Variations and Cognitive Phenotypes in Unselected Populations. JAMA 2015; 313: 2044–2054.

Merla G, Brunetti-Pierri N, Micale L, Fusco C. Copy number variants at Williams – Beuren syndrome 7q11 . 23 region. Hum. Genet. 2010; 128: 3–26.

Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. Am. J. Hum. Genet. 2010; 86: 749–764.

Morrow EM. Genomic Copy Number Variation in disorders of cognitive development. J Am Acad Child Adolesc Psychiatry 2010; 49: 1091–1104.

Morrow EM, Yoo S-Y, Flavell SW, Kim T-K, Lin Y, Hill RS, et al. Identifying autism loci and genes by tracing recent shared ancestry. Sci. N. Y. NY 2008; 321: 218–223.

Murali A, Rajalingam K. Small Rho GTPases in the control of cell shape and mobility. Cell Mol Life Sci 2014; 71: 1703–1721.

Nakagawa T, Xiong Y. X-Linked Mental Retardation Gene CUL4B Targets Ubiquitylation of H3K4 Methyltransferase Component WDR5 and Regulates Neuronal Gene Expression. Mol. Cell 2011; 43: 381–391.

Newman S, Hermetz KE, Weckselblatt B, Rudd MK. Next-generation sequencing of duplication CNVs reveals that most are tandem and some create fusion genes at breakpoints. Am. J. Hum. Genet. 2015; 96: 208–220.

Nodé-Langlois R, Muller D, Boda B. Sequential implication of the mental retardation proteins ARHGEF6 and PAK3 in spine morphogenesis. J. Cell Sci. 2006; 119: 4986–4993.

Norman RX, Ko HW, Huang V, Eun CM, Abler LL, Zhang Z, et al. Tubby-like protein 3 (TULP3) regulates patterning in the mouse embryo through inhibition of Hedgehog signaling. Hum. Mol. Genet. 2009; 18: 1740–1754.

Park H, Kim J-I, Ju YS, Gokcumen O, Mills RE, Kim S, et al. Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing. Nat. Genet. 2010; 42: 400–405.

Pfaffl M, Lange I, Daxenberger A, Meyer H. Tissue-specific expression pattern of estrogen receptors (ER): quantification of ER alpha and ER beta mRNA with real-time RT-PCR. APMIS 2001; 109: 345–355.

Phelan K, Mcdermid HE. The 22q13.3 Deletion Syndrome (Phelan-McDermid Syndrome). Mol. Syndromol. 2012; 2: 186–201.

Phelan MC. Deletion 22q13.3 syndrome. Orphanet J. Rare Dis. 2008; 3: 1–6.

Piton A, Redin C, Mandel J-L. XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. Am. J. Hum. Genet. 2013; 93: 368–83.

Potocki L, Bi W, Treadwell-deering D, Carvalho CMB, Eifert A, Friedman EM, et al. Characterization of Potocki-Lupski Syndrome Critical Interval That Can Convey an Autism Phenotype. Am. J. Hum. Genet. 2007; 80: 633–649.

Prasad AN, Rupar CA, Prasad C. Methylenetetrahydrofolate reductase (MTHFR) deficiency and infantile epilepsy. Brain Dev. 2011; 33: 758–769.

Prosser SL, O'Regan L, Fry AM. Novel insights into the mechanisms of mitotic spindle assembly by NEK kinases. Mol. Cell. Oncol. 2016; 3: e1062952.

Protty MB, Watkins NA, Colombo D, Thomas SG, Heath VL, Herbert JMJ, et al. Identification of Tspan9 as a novel platelet tetraspanin and the collagen receptor GPVI as a component of tetraspanin microdomains. Biochem. J. 2009; 417: 391–400.

Rafiullah R, Aslamkhan M, Paramasivam N, Thiel C, Mustafa G, Wiemann S, et al. Homozygous missense mutation in the LMAN2L gene segregates with intellectual disability in a large consanguineous Pakistani family. J. Med. Genet. 2016; 53: 138–144.

van Ravenswaaij-Arts CMA, Kleefstra T. Emerging microdeletion and microduplication syndromes; the counseling paradigm. Eur. J. Med. Genet. 2009; 52: 75–76.

Ravn K, Lindquist S, Nielsen K, Dahm T, Tumer Z. Deletion of CUL4B leads to concordant phenotype in a monozygotic twin pair. Clin. Genet. 2012; 82: 292–294.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature 2006; 444: 444–54.

Romani M, Micalizzi A, Kraoua I, Dotti MT, Cavallin M, Sztriha L, et al. Mutations in B9D1 and MKS1 cause mild Joubert syndrome : expanding the genetic overlap with the lethal ciliopathy Meckel syndrome. Orphanet J. Rare Dis. 2014; 9: 1–4.

Rosenblatt D, Lue-Shing H, Arzoumanian A, Low-Nang L, Matiaszuk N. Methylenetetrahydrofolate Reductase (MR) Deficiency: Thermolability of Residual MR Activity, Methionine Synthase Activity, and Methylcobalamin Levels in Cultured Fibroblasts. Biochem. Med. Metab. Biol. 1992; 47: 221–225.

Rosenthal JA, Chen H, Slepnev VI, Pellegrini L, Salcini EA, Di Fiore PP, et al. The Epsins Define a Family of Proteins That Interact with Components of the Clathrin Coat and Contain a New Protein Module. J. Biol. Chem. 1999; 274: 33959–33965.

Roy A, Gonzalez-Gomez M, Pierani A, Meyer G, Tole S. Lhx2 Regulates the Development of the Forebrain Hem System. Cereb. Cortex 2014; 24: 1361–1372.

Sajan SA, Jhangiani SN, Muzny DM, Gibbs RA, Lupski JR, Glaze DG, et al. Enrichment of mutations in chromatin regulators in people with Rett syndrome lacking mutations in MECP2. Genet. Med. Off. J. Am. Coll. Med. Genet. 2017; 19: 13–19.

Schwaibold EMC, Smogavec M, Hobbiebrunken E, Winter L, Zoll B, Burfeind P, et al. Intragenic duplication of EHMT1 gene results in Kleefstra syndrome. Mol. Cytogenet. 2014; 7: 3–7.

Sheen VL, Ganesh VS, Topcu M, Sebire G, Bodell A, Hill RS, et al. Mutations in ARFGEF2 implicate vesicle trafficking in neural progenitor proliferation and migration in the human cerebral cortex. Nat. Genet. 2004; 36: 69–76.

Sheng M, Kim E. The Shank family of scaffold proteins. J. Cell Sci. 2000; 113: 1851–1856.

Simões MR, Albuquerque CP, Pinho MS, Pereira M, Seabra-Santos M, Alberto I. Bateria de Avaliação Neuropsicológica de Coimbra (BANC).

Sleven H, Welsh SJ, Yu J, Churchill MEA, Wright CF, Henderson A, et al. De Novo Mutations in EBF3 Cause a Neurodevelopmental Syndrome. Am. J. Hum. Genet. 2017; 100: 130–150.

Splawski I, Timothy KW, Sharpe LM, Decher N, Kumar P, Bloise R, et al. CaV1.2 Calcium Channel Dysfunction Causes a Multisystem Disorder Including Arrhythmia and Autism. Cell 2004; 119: 19–31.

Stouffer MA, Golden JA, Francis F. Neurobiology of Disease Neuronal migration disorders: Focus on the cytoskeleton and epilepsy. Neurobiol. Dis. 2015; 1: doi: 10.1016/j.nbd.2015.08.003.

Svec D, Tichopad A, Novosadova V, Pfaffl MW, Kubista M. How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. Biomol. Detect. Quantif. 2015; 3: 9–16.

Tarpey PS, Raymond FL, Meara SO, Edkins S, Teague J, Butler A, et al. Mutations in CUL4B, Which Encodes a Ubiquitin E3 Ligase Subunit, Cause an X-linked Mental Retardation Syndrome Associated with Aggressive Outbursts, Seizures, Relative Macrocephaly, Central Obesity, Hypogonadism, Pes Cavus, and Tremor. Am. J. Hum. Genet. 2007; 80: 345–352.

Taylor S, Wakem M, Dijkman G, Alsarraj M, Nguyen M. A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines. Methods 2010; 50: S1–S5.

Tennant-Eyles AJ, Moffitt H, Whitehouse CA, Roberts RG. Characterisation of the FAM69 family of cysteine-rich endoplasmic reticulum proteins. Biochem. Biophys. Res. Commun. 2011; 406: 471–477.

Thiselton DL, McDowall J, Brandau O, Ramser J, d'Esposito F, Bhattacharya SS, et al. An integrated, functionally annotated gene map of the DXS8026-ELK1 interval on human Xp11.3-Xp11.23: potential hotspot for neurogenetic disorders. Genomics 2002; 79: 560–572.

Tomoda T, Kim JH, Zhan C, Hatten ME. Role of Unc51.1 and its binding partners in CNS axon outgrowth. Genes Dev. 2004; 18: 541–558.

Torkamani A, Bersell K, Jorge B, Bjork Jr RL, Friedman JR, Bloss CS, et al. De Novo KCNB1 Mutations in Epileptic Encephalopathy. Ann. Neurol. 2014; 76: 529–540.

Torres F, Barbosa M, Maciel P. Recurrent copy number variations as risk factors for neurodevelopmental disorders: critical overview and analysis of clinical implications. J Med Genet 2016; 53: 73–90.

Tsang BL, Devine OJ, Cordero AM, Marchetta CM, Mulinare J, Mersereau P, et al. Assessing the association between the methylenetetrahydrofolate reductase (MTHFR) 677C> T polymorphism and blood folate concentrations: a systematic review and meta-analysis of trials and observational studies. Am J Clin Nutr 2015: 1–9.

Tsou J-H, Yang Y-C, Pao P-C, Lin H-C, Huang N-K, Lin S-T, et al. Important Roles of Ring Finger Protein 112 in Embryonic Vascular Development and Brain Functions. Mol Neurobiol 2016; Mar 7.: doi:10.1007/s12035-016-9812-7.

Verdin H, Baere E De. FOXL2 Impairment in Human Disease. Horm. Res. Paediatr. 2012; 77: 2–11.

Victorino D, Godoy M, Goloni-Bertollo E, Pavarino E. Meta-analysis of Methylenetetrahydrofolate reductase maternal gene in Down syndrome: increased susceptibility in women carriers of the MTHFR 677T allele. Mol Biol Rep 2014; 41: 5491–5504.

Vilhais-Neto GC, Maruhashi M, Smith KT, Vasseur-Cognet M, Peterson AS, Workman JL, et al. Rere controls retinoic acid signalling and somite bilateral symmetry. Nature 2010; 463: 953–957.

Vulto-van Silfhout AT, Nakagawa T, Bahi-Buisson N, Haas SA, Hu H, Bienek M, et al. Variants in CUL4B are associated with cerebral malformations. Hum. Mutat. 2015; 36: 106–117.

Wang HL, Chang NC, Weng YH, Yeh TH. XLID CUL4B mutants are defective in promoting TSC2 degradation and positively regulating mTOR signaling in neocortical neurons. Biochim. Biophys. Acta - Mol. Basis Dis. 2013; 1832: 585–593.

Wang KS, Liu XF, Aragam N. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. Schizophr. Res. 2010; 124: 192–199.

Wang LH, Baker NE. E Proteins and ID Proteins: Helix-Loop-Helix Partners in Development and Disease. Dev. Cell 2015; 35: 269–280.

Wang SS, Lewcock JW, Feinstein P, Mombaerts P, Reed RR. Genetic disruptions of O/E2 and O/E3 genes reveal involvement in olfactory receptor neuron projection. Dev. Camb. Engl. 2004; 131: 1377–1388.

Wechsler D. Wechsler Intelligence Scale for Children—Third Edition. 1991

Wen J, Lopes F, Soares G, Farrell S a, Nelson C, Qiao Y, et al. Phenotypic and functional consequences of haploinsufficiency of genes from exocyst and retinoic acid pathway due to a recurrent microdeletion of 2p13.2. Orphanet J. Rare Dis. 2013; 8: 100.

Wieczorek D, Bögershausen N, Beleggia F, Steiner-Haldenstätt S, Pohl E, Li Y, et al. A comprehensive molecular study on coffin-siris and nicolaides-baraitser syndromes identifies a broad molecular and clinical spectrum converging on altered chromatin remodeling. Hum. Mol. Genet. 2013; 22: 5121–5135.

Willemsen M, Beunders G, Callaghan M, de Leeuw N, Nillesen W, Yntema H, et al. Familial Kleefstra syndrome due to maternal somatic mosaicism for interstitial 9q34 . 3 microdeletions. Clin. Genet. 2011; 80: 31–38.

Willemsen MH, Silfhout ATV, Nillesen WM, Bokhoven H Van, Philip N, Kini U, et al. Update on Kleefstra Syndrome. Mol. Syndromol. 2011; 2: 202–212.

Williams CL, Winkelbauer ME, Schafer JC, Michaud EJ, Yoder BK. Functional Redundancy of the B9 Proteins and Nephrocystins in Caenorhabditis elegans Ciliogenesis. Mol. Biol. Cell 2008; 19: 2154–2168.

Williams NM, Franke B, Mick E, Anney RJL, Freitag CM, Gill M, et al. Genome-Wide Analysis of Copy Number Variants In Attention Deficit Disorder: The Role of Rare Variants and Duplications at 15q13.3. Am J Psychiatry 2012; 169: 195–204.

Wu X, Wang X, Chan Y, Jia S, Luo Y, Tang W. Folate metabolism gene polymorphisms MTHFR C677T and A1298C and risk for Down syndrome offspring: a meta-analysis. Eur. J. Obstet. Gynecol. 2013; 167: 154–159.

Xu J, Zhou W, Yang F, Chen G, Li Haomin, Zhao Y, et al. The β-TrCP-FBXW2-SKP2 axis regulates lung cancer cell growth with FBXW2 acting as a tumour suppressor. Nat. Commun. 2017; 8: 14002.

Yatsenko SA, Hixson P, Roney EK, Scott DA, Schaaf CP, Ng Y-T, et al. Human subtelomeric copy number gains suggest a DNA replication mechanism for formation: beyond breakage - fusion-bridge for telomere stabilization. Hum. Genet. 2012; 131: 1895–1910.

Yilmaz S, Gokben S, Serdaroglu G, Eraslan C, Mancini GM, Tekin H, et al. The expanding phenotypic spectrum of ARFGEF2 gene mutation: Cardiomyopathy and movement disorder. Brain Dev. 2016; 38: 124–127.

Yoshida T, Yasumura M, Uemura T, Lee S-JSJ, Ra M, Taguchi R, et al. IL-1 receptor accessory protein-like 1 associated with mental retardation and autism mediates synapse formation by trans-synaptic interaction with protein tyrosine phosphatase δ. J. Neurosci. 2011; 31: 13485–13499.

Zahanova S, Meaney B, Łabieniec B, Verdin H, Baere E De, Nowaczyk MJM. Blepharophimosis-ptosis-epicanthus inversus syndrome plus: deletion 3q22.3q23 in a patient with characteristic

facial features and with genital anomalies, spastic diplegia, and speech delay. Clin. Dysmorphol. 2012; 21: 48–52.

Zhang Z, Gutierrez D, Li X, Bidlack F, Cao H, Wang J, et al. The LIM Homeodomain Transcription Factor LHX6 A TRANSCRIPTIONAL REPRESSOR THAT INTERACTS WITH PITUITARY HOMEOBOX 2 (PITX2) TO REGULATE ODONTOGENESIS. J. Biol. Chem. 2013; 288: 2485–2500.

Zou Y, Liu Q, Chen B, Zhang X, Guo C, Zhou H, et al. Mutation in CUL4B , Which Encodes a Member of Cullin-RING Ubiquitin Ligase Complex , Causes X-Linked Mental Retardation. Am. J. Hum. Genet. 2007; 80: 561–566.

The Deciphering Developmental Disorders (DDD) Study - DECIPHER v9.12 [Internet]. [cited 2017 Feb 7] Available from: https://decipher.sanger.ac.uk/ddd#ddgenes

Database of Genomic Variants [Internet]. [cited 2014 Aug 5] Available from: http://dgv.tcag.ca/dgv/app/home

Supplementary Tables

Supplementary Table S1 – Primers used for quantitative real-time PCR confirmation of gene dosage.

Chromosome	Gene	Reference sequence	Primer location	Primer Forward 5'→3'	Primer Reverse 5'→3'	Amplicon size (bp)
Chr 20	*SDC4*	ENSG00000124145	Exon4	ACCGAACCCAAGAAACTAGA	GTGCTGGACATTGACACCT	101bp
Chr 3	*ZNF80*	ENSG00000174255	Exon1	GCTACCGCCAGATTCACACT	AATCTTCATGTGCCGGGTTA	182bp
Chr X	*CUL4B*	NM_003588.3	Exon3	CTTCAACCTCGTCCTTCTGC	GTTGCAGCAGTTGGTGAAGA	166bp
Chr X	*CUL4B*	NM_003588.3	Exon21	ATTGATGCTGCAATTGTTCG	TGTTTGCAAGATTTGTGTCTGA	182bp
Chr X	*HUWE1*	NM_031407.5	Exon69	TGTTGACATCCCACTCTTGTTC	TTGTTTACAAAGGGTATAACCCAGA	152bp
Chr X	*HUWE1*	NM_031407.5	Exon75	GGCACACATCAAGGACGAG	GCAAAGCGAAGGAACTTCTG	153bp
Chr 10	*EBF3*	NM_001005463.2	Intron15	CTCTCTGCTGGGTGCTGAG	GCGTCCCTTCATACGCTAAC	169bp
Chr 1	*AKT3*	NM_005465.4	Exon7	TCTGGGCTTAACCTCTTCCA	TGTTAAAAAGGGATGTCTAGTGTTC	162bp
Chr 1	*AKT3*	NM_005465.4	Exon8	CCTTGAAATATTCCTTCCAGACA	CCATGCAAATACTGGATTTACTTCT	101bp
Chr 1	*AKT3*	NM_005465.4	Exon9	AGAGAGCGGGTGTTCTCTGA	CCTTGAGATCACGGTACACAA	106bp
Chr 1	*AKT3*	NM_005465.4	Exon10	CAGTTGGAGAATCTAATGCTGGA	AATGGAACCGAAGCCTACCT	150bp
Chr 20	*EBF4*	NM_001110514.1	Exon16	GCTGCCTCCTCCATGTCC	AAGGCGCTCCTCTGTTTGAC	101bp
Chr 12	*MED13L*	NM_015335.4	Exon3/Intron3	GGAAGAAGGACTCTGGGAAAA	CAGGAAACTCTCGGTATCTAGCA	151bp
Chr 7	*BAZ1B*	NM_032408.3	Exon3	TCCTGCCTGGTATGAGAAGC	TCCCACAGCATATTTGGTCA	112bp
Chr 2	*GPR45*	NM_007227.3	Exon1	ACGTCCCTTGAGGCTTACAC	ACGATGATGCAGACCACAGT	161bp
Chr 7	*CNOT4*	NM_001190850.1	Exon10	CACCGAGCGGTTTATAATTCA	AGACCTGTGTTGTGCTGTGG	164bp
Chr 9	*LHX2*	NM_004789.3	Exon3	GCTCGGGACTTGGTTTATCA	GTTGAAGTGTGCGGGGTACT	156bp
Chr 12	*TULP3*	NM_003324.4	Exon3	GGCTACTACTTGAGAAGAGGCAAA	TGACATTGCTGTGGGGAGTA	150bp
Chr 3	*SOX2*	NM_003106.3	Exon1	CCCACCTACAGCATGTCCTA	CTGATCATGTCCCGGAGGT	164bp

(Cont.)

Chr 7	*OCM*	NM_001097622.1	Intron1/exon2	CTCTGTTCTTCAGACCCAGACA	GCTTACTTAAGCTCTTCTTCATCCA	152bp
Chr 10	*MAP3K8*	NM_005204.3	Exon3	TGGAGTACATGAGCACTGGAA	TTGACACATGGTCATTAGACTGG	152bp
Chr 17	*MYO18A*	NM_00134765.1	Exon2	TGTCAAGCGCTTTTCCTTCT	AGAGTCCTCACCTCCACCTG	111bp
Chr 2	*ARHGEF4*	NM_015320.3	Exon14	TTCTGGCACAGCATCAGC	CACTTGCAGGCAGAGGAAG	144bp
Chr X	*ARHGEF6*	NM_004840.2	3'UTR	CTTGAAATGTCCCGCTGAAT	AACAACAGCAAATGCCCAAG	162bp
Chr 7	*CALD1*	NM_033139.3	Exon4	GAATGACGATGATGAAGAGGAG	ACAGTACCTGTTCTGGGCATTC	139bp
Chr 11	*KIRREL3*	NM_032531.3	Exon17/3'UTR	GATGCAGACTCACGTCTAAGGA	CTTGATCAGAGCTTCGAAGGAA	179bp
Chr 1	*MAST2*	NM_015112.2	Exon3	AGCTGCTCCCTTTGTCCAG	GCCACCTTTATGAACACTTACCAG	158bp
Chr 2	*USP34*	NM_014709.3	Exon80	ATGAAGGAGCAACTCCCATT	GCTCAGTTCCTGGATCAATAAT	168bp
Chr 15	*SNRPN*	NM_003097.3	5'UTR	CTTTCCTGTCTGTCATTTTGC	GTCCCTTCTCTGTGCAGC	160bp
Chr 1	*PRKAB2*	NM_005399.4	Exon4	AGCCATAATGACTTTGTTGCCA	GCCCATCAGTCTTGACAGAAA	174bp
Chr 9	*ZNF658*	NM_001317916.1	5'UTR	ACCTCTTTGGTATAAACGTTCCAT	AGGACAGGGAGTCACATCTCTC	119bp
Chr X	*PNPLA4*	NM_001142389.1	Exon6	CACCAACGCTCTTCCCAT	CACCATGATATCCTGCTTGG	136bp
Chr 1	*FAM69A*	NM_001006605.4	Exon3	AGACTGGAGTTATTGATGGGC	CTGGAATGTTTATTCATAATGGC	130bp

Supplementary table S2 – Primers used for mRNA expression studies by qRT-PCR.

Chromosome	Gene	Reference sequence	Primer Forward 5'→3'	Primer location	Primer Reverse 5'→3'	Primer location	Amplicon size (bp)
Chr 1	*FAM69A*	NM_001006605.4	AGACTGGAGTTATTGATGGGC	Exon3	CACAACACCTGGTAGATTATCCC	Exon4	134bp
Chr 1	*DPYD*	NM_000110.3	GCAGCAATTTGCTACTGAGG	Exon5	CCCAGCACCAAAAAGAGC	Exon6	123bp
Chr 1	*TGFBR3*	NM_003243.4	GCCTTGATGGAGAGCTTCAC	Exon3	GGGATTCAGGTGAAGTGTGAC	Exon4	146bp
Chr 5	*PPIB*	NM_000942.4	TGACCTACGAATTGGAGATGAAG	Exon2	TGCTGTTTTTGTAGCCAAATCCT	Exon3	130bp
Chr 7	*CALD1*	NM_033139.3	GCAGAAAAGCAGTGGTGTCA	Exons8/9	CCTTCAGCAGGAACAGGAAG	Exon10	152bp
Chr 7	*AGBL3*	NM_178563.3	TCCATTGACTCTCTGACTTACCTTC	Exon12	ATCTGGTTCATTTGGCCTTG	Exon15	194bp
Chr 7	*CNOT4*	NM_001190850.1	CCTGCATGTAGAAAGCCATATCC	Exons2/3	GTACACTAGCCAAATGTTTGCG	Exon3	150bp
Chr 7	*EXOC4*	NM_021807.3	CACTACACAGAATTGACGACAGC	Exon2	TTTCCGAAGCTCATCCCGTTT	Exon3	146bp
Chr 9	*EHMT1*	NM_024757.4	CTGCATGCAGCCAGTAAAGATC	Exons3/4	CTGCTGTCGTCCAAAGTCAG	Exon4	104bp
Chr 9	*CACNA1B*	NM_000718.3	TGGTGTCTGGGATTCCAAG	Exon4	CCATGTAGAACTCCAGGCCA	Exon5	134bp
Chr 9	*TSC1*	NM_000368.4	GATAGAACTGAAGAAGGCCAAC	Exon19	GTGCTTGTTCTGCAGTTGTTCC	Exon20	177bp
Chr 14	*TECPR2*	NM_014844.3	GGGGAAGACGGAATCTATCA	Exons2/3	GTGACATCAAATCTCCGAAGCT	Exons3/4	149bp
Chr 14	*INF2*	NM_022489.3	GACCACTTCTACCTCCTCCTG	Exon12	TGAGGAAGTTCCCAATTCTC	Exon14	201bp
Chr 9	*FBXW2*	NM_012164.3	CTTGTGACAGGCTCCTTTGAC	Exon4	ATTGTAGTCCACGCTAAATACCG	Exons4/5	111bp
Chr 9	*NEK6*	NM_001145001.2	AAGATAGGCCGAGGACAGTTC	Exon4	CCATCATCTCAAAGATCTG	Exons4/5	99bp
Chr 9	*PSMB7*	NM_002799.2	TTTCTCCGCCCATACACAGTG	Exon7	AGCACCTCAATCTCCAGAGGA	Exon8	119bp
Chr 12	*TSPAN9*	NM_006675.4	AACATCATCCAGGCTGAG	Exon6	GAGTTCTCCATGCAGCAG	Exon7	106bp
Chr X	*LAMP2*	NM_002294.2	ACCACTGTGCCATCTCCTAC	Exon5	GAGTCTAAGTAGAGCAGTGTGAG	Exon6	215bp
Chr X	*CUL4B*	NM_003588.3	GCATTCTTCTCTTGATTGAGAGG	Exon8	GAGCCGGTTAGTTTCTTCC	Exon9	142bp
Chr X	*FHL1*	NM_001159702.2	AAGAACCGCTTCTGGCATGAC	Exon4	CCCCTTGTACTCCACGTTTTG	Exon5	188bp
Chr X	*ARHGEF6*	NM_004840.2	TCCTCGCTGAAAAATGGGGTA	Exon1	CTTGGAGGGTTGCACATCCT	Exon2	147bp
Chr X	*MAP7D3*	NM_024597.3	TTGTCATCTGCAGGCCTTC	Exon6	GCATTACATAATTGGTGACGC	Exon8	159bp

Supplementary table S3 – OMIM entrance, haploinsufficiency score and constraint metrics for the selected genes in patient R17.

2q11.2-q12.2 deletion	**List of all the genes affected**	***C2orf29, C2orf49, CREG2, FHL2, GPR45, IL18R1, IL18RAP, IL1R1, IL1R2, IL1RL1, IL1RL2, LOC150568, MAP4K4, MFSD9, MRPS9, POU3F3, RFX8, RNF149, SLC9A2, SLC9A4, SNORD89, TBC1D8, TGFBRAP1, TMEM182***							
Gene	**Morbid gene**	**OMIM**	**% HI score**	**DDG2P**	**ClinVar**	**Constraint Metrics**			
						Synonymous (z)	**Missense (z)**	**LoF (pLI)**	**CNV (z)**
MAP4K4	No	-	20-30%	-	1del/7dups	-0.83	4.01	1	0.19
FHL2	No	-	20-30%	-	4dels/7dups/32SNVs	-0.15	0.35	0	0.53
POU3F3	No	-	20-30%	-	4dels/6dups	ND	ND	ND	ND
CNOT11	No	-	20-30%	-	2dels9dups	1.66	3.76	0.99	0.18

OMIM: Online Mendelian Inheritance in Man; HI score: Haploinsufficiency Score index - high ranks (e.g. 0-10%) indicate a gene is more likely to exhibit haploinsufficiency, low ranks (e.g. 90-100%) indicate a gene is more likely to NOT exhibit haploinsufficiency (retrieved from Decipher); LoF: Loss of function; CNVs: copy number variations; z: Z score is the deviation of observed counts from the expected number for one gene (positive Z scores = gene intolerance to variation, negative Z scores = gene tolerant to variation) (retrieved from ExAC); pLI: probability that a given gene is intolerant of loss-of-function variation (pLI closer to one = more intolerant the gene is to LoF variants, pLI >= 0.9 is extremely LoF intolerant) (retrieved from ExAC); del – deletion; dup – duplication; SNV – single nucleotide variant; ins – insertion; indel – insertion/deletion

Supplementary table S4 - OMIM entrance, haploinsufficiency score and constraint metrics for the selected genes in patient R20.

10q26.3 deletion	**List of all the genes affected**	***MGMT, EBF3, GLRX3***							
Gene	**Morbid gene**	**OMIM**	**% HI score**	**DDG2P**	**ClinVar**	**Constraint Metrics**			
						Synonymous (z)	**Missense (z)**	**LoF (pLI)**	**CNV (z)**
EBF3	Yes	617330, Hypotonia, ataxia, and delayed development syndrome	0-10%	-	27dels/13dups/15SNVs	1.84	4.89	1	-0.15

OMIM: Online Mendelian Inheritance in Man; HI score: Haploinsufficiency Score index - high ranks (e.g. 0-10%) indicate a gene is more likely to exhibit haploinsufficiency, low ranks (e.g. 90-100%) indicate a gene is more likely to NOT exhibit haploinsufficiency (retrieved from Decipher); LoF: Loss of function; CNVs: copy number variations; z: Z score is the deviation of observed counts from the expected number for one gene (positive Z scores = gene intolerance to variation, negative Z scores = gene tolerant to variation) (retrieved from ExAC); pLI: probability that a given gene is intolerant of loss-of-function variation (pLI closer to one = more intolerant the gene is to LoF variants, pLI >= 0.9 is extremely LoF intolerant) (retrieved from ExAC); del – deletion; dup – duplication; SNV – single nucleotide variant; ins – insertion; indel – insertion/deletion

Supplementary table S5 - OMIM entrance, haploinsufficiency score and constraint metrics for the selected genes in patient C17.

17p11.2 deletions (both)	List of all the genes affected	***TNFRSF13B, MPRIP, PLD6, FLCN, COPS3 + TRIM16L, ZNF286B, TBC1D28, FBXW10, TVP23B, PRPSAP2, SLC5A10, GRPA, FAM83G, GRAPL, EPN2, B9D1, MAPK7, MFAP4, RNF112, SLC47A1, ALDH3A2, SLC47A2, ALDH3A1, ULK2, AKAP10, SPECC1, LGALS9B, CDRT15L2, CCDC144NL, USP22, DHRS7B, TMEM11, C17orf103, MAP2K3***							
Gene	**Morbid gene**	**OMIM**	**% HI score**	**DDG2P**	**ClinVar**	**Constraint Metrics**			
						Synonymous (z)	**Missense (z)**	**LoF (pLI)**	**CNV (z)**
COPS3	No	-	0-10%	-	33dels/29dups	0.06	1.85	0.99	0.91
EPN2	No		40-50%	-	27dels/28dups	0.16	1.09	0.18	0.58
B9D1	Yes	-614209, ?Meckel syndrome 9; 617120, Joubert syndrome 27	50-60%	Probable	29dels/28dups/23SNVs	-0.27	0.20	0.12	0.42
RNF112	No	-	70-80%	-	27dels/28dups	-1.07	1.15	0	0.46
ULK2	No	-	30-40%	-	27dels/30dups/1SNV	0.03	0.61	0	-0.3
ALDH3A2	Yes	270200, Sjogren-Larsson syndrome (AR)	50-60%	Yes	47dels/34dups/62SNVs/2indel/7ins	1.06	0.47	0.01	-0.76
AKAP10	Yes	115080, Cardiac conduction defect, susceptibility to	10-20%	-	27dels/30dups/2SNVs	-0.75	0.38	0.93	0.41
MAP2K3	No	-	10-20%	-	3dels/4dups/1SNV	-0.13	-0.23	0	-4.29
TMEM11	No	-	10-20%	-	2dels/4dups	0.22	2.06	0.78	0.24

OMIM: Online Mendelian Inheritance in Man; HI score: Haploinsufficiency Score index - high ranks (e.g. 0-10%) indicate a gene is more likely to exhibit haploinsufficiency, low ranks (e.g. 90-100%) indicate a gene is more likely to NOT exhibit haploinsufficiency (retrieved from Decipher); LoF: Loss of function; CNVs: copy number variations; z: Z score is the deviation of observed counts from the expected number for one gene (positive Z scores = gene intolerance to variation, negative Z scores = gene tolerant to variation) (retrieved from ExAC); pLI: probability that a given gene is intolerant of loss-of-function variation (pLI closer to one = more intolerant the gene is to LoF variants, pLI >= 0.9 is extremely LoF intolerant) (retrieved from ExAC); del – deletion; dup – duplication; SNV – single nucleotide variant; ins – insertion; indel – insertion/deletion

Supplementary table S6 - OMIM entrance, haploinsufficiency score and constraint metrics for the selected genes in patient R21.

20q13.12-q13.13 deletion	List of all the genes affected	***ACOT8, ARFGEF2, B4GALT5, C20orf123, C20orf165, C20orf199, CD40, CDH22, CEBPB, CSE1L, CTSA, DBNDD2, DDX27, DNTTIP1, ELMO2, EYA2, KCNB1, KCNK15, KCNS1, LOC100131496, LOC100240726, LOC284749, MATN4, MIR1259, MMP9, NCOA3, NCOA5, NEURL2, PABPC1L, PCIF1, PI3, PIGT, PLTP, PREX1, PTGIS, RBPJL, RIMS4, RNF114, SDC4, SEMG1, SEMG2, SLC12A5, SLC13A3, SLC2A10, SLC35C2, SLC9A8, SLPI, SNAI1, SNORD12, SNORD12B, SNORD12C, SNX21, SPATA2, SPINLW1, SPINT3, SPINT4, STAU1, STK4, SULF2, SYS1, SYS1-DBNDD2, TMEM189, TMEM189-UBE2V1, TNNC2, TOMM34, TP53RK, TP53TG5, UBE2C, UBE2V1, WFDC10A, WFDC10B, WFDC11, WFDC12, WFDC13, WFDC2, WFDC3, WFDC5, WFDC6, WFDC8, WFDC9, WISP2, YWHAB, ZMYND8, ZNF334, ZNF335, ZNFX1, ZSWIM1, ZSWIM3***							
Gene	**Morbid gene**	**OMIM**	**% HI score**	**DDG2P**	**ClinVar**	**Constraint Metrics**			
						Synonymous (z)	**Missense (z)**	**LoF (pLI)**	**CNV (z)**
KCNB1	Yes	616056, Epileptic encephalopathy, early infantile, 26	20-30%	Yes	1del/5dups/53SNVs/1ins	1.48	5.15	0.98	-0.56
PIGT	Yes	615398, Multiple congenital anomalies-hypotonia-seizures syndrome 3; 615399, ?Paroxysmal nocturnal hemoglobinuria 2	30-40%	Yes	1del/3dups/9SNV/1ins	-0.66	-0.37	0	0.26
CTSA	Yes	256540, Galactosialidosis	40-50%	Yes	10dels/3dups/41SNVs/2indel/3ins	1.13	1.45	0	-0.98
SLC2A10	Yes	208050, Arterial tortuosity syndrome	70-80%	Yes	9dels/3dups/155SNVs/1ins	-0.51	-1.01	0.01	-0.48
ARFGEF2	Yes	608097, Periventricular heterotopia with microcephaly	20-30%	Probable	6dels/6dups/138SNVs/3ins	1.18	2.9	1	-0.04

OMIM: Online Mendelian Inheritance in Man; HI score: Haploinsufficiency Score index - high ranks (e.g. 0-10%) indicate a gene is more likely to exhibit haploinsufficiency, low ranks (e.g. 90-100%) indicate a gene is more likely to NOT exhibit haploinsufficiency (retrieved from Decipher); LoF: Loss of function; CNVs: copy number variations; z: Z score is the deviation of observed counts from the expected number for one gene (positive Z scores = gene intolerance to variation, negative Z scores = gene tolerant to variation) (retrieved from ExAC); pLI: probability that a given gene is intolerant of loss-of-function variation (pLI closer to one = more intolerant the gene is to LoF variants, pLI >= 0.9 is extremely LoF intolerant) (retrieved from ExAC); del – deletion; dup – duplication; SNV – single nucleotide variant; ins – insertion; indel – insertion/deletion

Supplementary table S7 - OMIM entrance, haploinsufficiency score and constraint metrics for the selected genes in patient C18.

1p22.1-p21.3 duplication	**List of all the genes affected**	***TGFBR3, BRDT, EPHX4, BTBD8, KIAA1107, C1orf146, GLMN, RPAP2, GFI1, EVI5, RPL5, FAM69A, SNORD21, SNORA66, MTF2, TMED5, CCDC18, LOC100131564, DR1, FNBP1L, BCAR3, LOC100129046, MIR760, DNTTIP2, GCLM, ABCA4, ARHGAP29, ABCD3, F3, SLC44A3, CNN3, LOC729970, ALG14, TMEM56, TMEM56-RWDD3, RWDD3, FLJ31662, PTBP2, DPYD, DPYD-AS1, MIR137HG, MIR2682, MIR137, LOC729987***							
Gene	**Morbid gene**	**OMIM**	**% HI score**	**DDG2P**	**ClinVar**	**Constraint Metrics**			
						Synonymous (z)	**Missense (z)**	**LoF (pLI)**	**CNV (z)**
FAM69A	No	-	10-20%	-	1del/3dups/16SNVs	0.38	0.55	0.09	nan
TGFBR3	No	-	10-20%	-	4dups/1SNV	-0.44	0.55	0.01	-1.04
GLMN	Yes	138000, Glomuvenous malformations	10-20%	Yes	6dels/5dups/29SNVs/1indel/ 2ins	-1.61	-0.18	0	0.58
EVI5	No	-	10-20%	-	1del/2dups	0.16	-0.09	0	0.73
RPL5	Yes	612561, Diamond-Blackfan anemia 6	0-10%	-	3dels/3dups/31SNVs/1indel/ 2ins	0.37	1.52	0.99	0.61
MTF2	No	-	0-10%	-	1del/2dups	-0.29	2.29	0.99	1.18
DR1	No	-	0-10%	-	1del/2dups	-0.63	2.04	0.84	0.69
ABCA4	Yes	604116, Cone-rod dystrophy 3; 248200, Retinal dystrophy, Stargardt disease 1, Fundus flavimaculatus; 601718, Retinitis pigmentosa 19; 153800, Macular degeneration, age-related, 2	10-20%	-	82dels/15dups/688SNVs/3indels/27ins	-0.74	-1.50	0	0.12
ABCD3	Yes	-616278, ?Bile acid synthesis defect, congenital, 5	10-20%	-	2dels/3dups/4SNVs/1ins	-0.22	2.42	1	0.17
CNN3	No	-	10-20%	-	2dups	1.92	2.38	0.86	0.69
PTBP2	No	-	0-10%	-	3dels/2dups	-0.26	3.27	0.99	0.59
DPYD	Yes	274270, Dihydropyrimidine dehydrogenase deficiency, 5-fluorouracil toxicity	0-10%	-	23dels/7dups/131SNVs/1indel/5ins	-0.68	-1.51	0	nan

OMIM: Online Mendelian Inheritance in Man; HI score: Haploinsufficiency Score index - high ranks (e.g. 0-10%) indicate a gene is more likely to exhibit haploinsufficiency, low ranks (e.g. 90-100%) indicate a gene is more likely to NOT exhibit haploinsufficiency (retrieved from Decipher); LoF: Loss of function; CNVs: copy number variations; z: Z score is the deviation of observed counts from the expected number for one gene (positive Z scores = gene intolerance to variation, negative Z scores = gene tolerant to variation) (retrieved from ExAC); pLI: probability that a given gene is intolerant of loss-of-function variation (pLI closer to one = more intolerant the gene is to LoF variants, pLI >= 0.9 is extremely LoF intolerant) (retrieved from ExAC); del – deletion; dup – duplication; SNV – single nucleotide variant; ins – insertion; indel – insertion/deletion

Supplementary table S8 - OMIM entrance, haploinsufficiency score and constraint metrics for the selected genes in patient R23.

9q33.2-q33.3 triplication	List of all the genes affected	*C5, FBXW2, LOC253039, LOC402377, PHF19, PSMD5, TRAF1, C9orf45, CEP110, CRB2, DAB2IP, DENND1A, GGTA1, GPR21, GSN, LHX2, LHX6, LOC100129034, MIR548D1, MIR600, MIR601, MORN5, MRRF, NDUFA8, NEK6, OR1B1, OR1J1, OR1J2, OR1J4, OR1K1, OR1L1, OR1L3, OR1L4, OR1L6, OR1L8, OR1N1, OR1N2, OR1Q1, OR5C1, PDCL, PSMB7, PTGS1, RAB14, RABGAP1, RBM18, RC3H2, SNORD90, STOM, STRBP, TTLL11, ZBTB26, ZBTB6*							
Gene	**Morbid gene**	**OMIM**	**% HI score**	**DDG2P**	**ClinVar**	**Constraint Metrics**			
						Synonymous (z)	**Missense (z)**	**LoF (pLI)**	**CNV (z)**
CRB2	Yes	616220, Focal segmental glomerulosclerosis 9; 219730, Ventriculomegaly with cystic kidney disease	60-70%	Yes	2dels/17dups/7SNVs/1ins	1.14	1.64	0	0.84
LHX2	No	-	0-10%	-	1del/16dups	2.63	4.59	0.95	0.5
DENND1A	No	-	0-10%	-	2dels/16dups	-0.25	1.31	0.90	0.80
STRBP	No	-	0-10%	-	2dels/16dups	0.39	3.19	1	1.02
RAB14	No	-	0-10%	-	2dels/15dups	1.01	2.90	0.97	1.02
GSN	Yes	105120, Amyloidosis, Finnish type	0-10%	-	1del/18dups/43SNVs/9ins	0.78	1.75	0	0.51
PSMB7	No	-	0-10%	-	1del/16dups	0.05	0.89	0.94	0.33
LHX6	No	-	0-10%	-	1del/15dups	2.69	4.43	0.95	0.46
ZBTB26	No	-	0-10%	-	1del/16dups	-0.93	1.77	0.02	0.66

OMIM: Online Mendelian Inheritance in Man; HI score: Haploinsufficiency Score index - high ranks (e.g. 0-10%) indicate a gene is more likely to exhibit haploinsufficiency, low ranks (e.g. 90-100%) indicate a gene is more likely to NOT exhibit haploinsufficiency (retrieved from Decipher); LoF: Loss of function; CNVs: copy number variations; z: Z score is the deviation of observed counts from the expected number for one gene (positive Z scores = gene intolerance to variation, negative Z scores = gene tolerant to variation) (retrieved from ExAC); pLI: probability that a given gene is intolerant of loss-of-function variation (pLI closer to one = more intolerant the gene is to LoF variants, pLI >= 0.9 is extremely LoF intolerant) (retrieved from ExAC); del – deletion; dup – duplication; SNV – single nucleotide variant; ins – insertion; indel – insertion/deletion

Supplementary table S9 - OMIM entrance, haploinsufficiency score and constraint metrics for the selected genes in patient R24.

12p13.33-p13.32 duplication	**List of all the genes affected**	***C12orf32, CACNA1C, FKBP4, FOXM1, ITFG2, NRIP2, TEAD4, TSPAN9, TULP3***							
Gene	**Morbid gene**	**OMIM**	**% HI score**	**DDG2P**	**ClinVar**	**Constraint Metrics**			
						Synonymous (z)	**Missense (z)**	**LoF (pLI)**	**CNV (z)**
CACNA1C	Yes	611875, Brugada syndrome 3; 601005, Timothy syndrome	0-10%	Yes	25dels/27dups/483SNVs/3ins	-0.32	6.41	1	-0.77
FOXM1	No	-	0-10%	-	9dels/21dups	-0.69	0.38	0.80	-0.32
TULP3	No	-	60-70%	-	9dels/21dups	0.12	-0.83	0	-0.56
TSPAN9	No	-	30-40%	-	8dels/21dups	-0.18	1.28	0.78	-2.06

OMIM: Online Mendelian Inheritance in Man; HI score: Haploinsufficiency Score index - high ranks (e.g. 0-10%) indicate a gene is more likely to exhibit haploinsufficiency, low ranks (e.g. 90-100%) indicate a gene is more likely to NOT exhibit haploinsufficiency (retrieved from Decipher); LoF: Loss of function; CNVs: copy number variations; z: Z score is the deviation of observed counts from the expected number for one gene (positive Z scores = gene intolerance to variation, negative Z scores = gene tolerant to variation) (retrieved from ExAC); pLI: probability that a given gene is intolerant of loss-of-function variation (pLI closer to one = more intolerant the gene is to LoF variants, pLI >= 0.9 is extremely LoF intolerant) (retrieved from ExAC); del – deletion; dup – duplication; SNV – single nucleotide variant; ins – insertion; indel – insertion/deletion

Supplementary table S10 - OMIM entrance, haploinsufficiency score and constraint metrics for the selected genes in patients R25 and R26.

Xq24 duplication	**List of all the genes affected**	***C1GALT1C1, CUL4B, LAMP2, MCTS1***							
Gene	**Morbid gene**	**OMIM**	**% HI score**	**DDG2P**	**ClinVar**	**Constraint Metrics**			
						Synonymous (z)	**Missense (z)**	**LoF (pLI)**	**CNV (z)**
CUL4B	Yes	300354, Mental retardation, X-linked, syndromic 15 (Cabezas type)	0-10%	Yes	56dels/48dups/20SNVs/1ins	0.65	3.88	1	**nan**
LAMP2	Yes	300257, Danon disease	50-60%	Yes	73dels/75dups/158SNVs/13ins	0.15	0.41	0.95	**nan**
C1GALT1C1	Yes	300622, Tn polyagglutination syndrome, somatic	20-30%	-	51dels/46dups/5SNVs	-0.78	0.46	0.69	**nan**
MCTS1	No	-	10-20%	-	51dels/46dups	0.74	1.86	0.83	**nan**

OMIM: Online Mendelian Inheritance in Man; HI score: Haploinsufficiency Score index - high ranks (e.g. 0-10%) indicate a gene is more likely to exhibit haploinsufficiency, low ranks (e.g. 90-100%) indicate a gene is more likely to NOT exhibit haploinsufficiency (retrieved from Decipher); LoF: Loss of function; CNVs: copy number variations; z: Z score is the deviation of observed counts from the expected number for one gene (positive Z scores = gene intolerance to variation, negative Z scores = gene tolerant to variation) (retrieved from ExAC); pLI: probability that a given gene is intolerant of loss-of-function variation (pLI closer to one = more intolerant the gene is to LoF variants, pLI >= 0.9 is extremely LoF intolerant) (retrieved from ExAC); del – deletion; dup – duplication; SNV – single nucleotide variant; ins – insertion; indel – insertion/deletion

Supplementary table S11 - OMIM entrance, haploinsufficiency score and constraint metrics for the selected genes in patient C24.

Xq26.3 duplication	List of all the genes affected	*FHL1, MAP7D3, GPR112, BRS3, HTATSF1, VGLL1, MIR934, CD40LG, ARHGEF6*							
Gene	Morbid gene	OMIM	% HI score	DDG2P	ClinVar	Constraint Metrics			
						Synonymous (z)	Missense (z)	LoF (pLI)	CNV (z)
ARHGEF6	Yes	300436, Mental retardation, X-linked 46	10-20%	-	56dels/56dups/46SNVs	-0.25	0.77	1	nan
CD40LG	Yes	308230, Immunodeficiency, X-linked, with hyper-IgM	0-10%	-	59dels/55dups/17SNVs/7ins	0.82	0.92	0.86	nan
BRS3	no	-	30-40%	-	52dels/50dups/2ins	-0.49	0.90	0.89	nan
FHL1	Yes	300696, Emery-Dreifuss muscular dystrophy 6, X-linked; Myopathy, X-linked, with postural muscle atrophy; 300717, Reducing body myopathy, X-linked 1a, severe, infantile or early childhood onset ¸300718, Reducing body myopathy, X-linked 1b, with late childhood or adult onset; 300695, Scapuloperoneal myopathy, X-linked dominant	10-20%	Yes	58dels/50dups/52SNV/5ins/1indel	0.59	1.29	0.92	nan

OMIM: Online Mendelian Inheritance in Man; HI score: Haploinsufficiency Score index - high ranks (e.g. 0-10%) indicate a gene is more likely to exhibit haploinsufficiency, low ranks (e.g. 90-100%) indicate a gene is more likely to NOT exhibit haploinsufficiency (retrieved from Decipher); LoF: Loss of function; CNVs: copy number variations; z: Z score is the deviation of observed counts from the expected number for one gene (positive Z scores = gene intolerance to variation, negative Z scores = gene tolerant to variation) (retrieved from ExAC); pLI: probability that a given gene is intolerant of loss-of-function variation (pLI closer to one = more intolerant the gene is to LoF variants, pLI >= 0.9 is extremely LoF intolerant) (retrieved from ExAC); del – deletion; dup – duplication; SNV – single nucleotide variant; ins – insertion; indel – insertion/deletion

The effect of CNVs at 1q43-q44 in neurodevelopment phenotypes and head circumference alterations

The effect of CNVs at 1q43-q44 in neurodevelopment phenotypes and head circumference alterations

Fátima Lopes, Fátima Torres, Gabriela Soares, Clara D. van Karnebeek, Cecília Martins, Diana Antunes, João Silva, Luís Filipe Botelho, Susana Sousa, Paula Rendeiro, Purificação Tavares, Hilde Van Esch, Evica Rajcan-Separovic, Patricia Maciel

In preparation

Abstract

Microdeletions at 1q43q44 have been described as resulting in a clinically recognizable phenotype of intellectual disability, facial dysmorphisms and microcephaly. In the last years, the report of several patients with copy number variations affecting the 1q43q44 region highlighted *AKT3* gene as a likely key player in head size anomalies. However, genotype-phenotype correlations from previous reports have not been consistent.

We report five patients with copy number variations in the 1q43q44 region: one patient with a larger deletion (3.7 Mb), two patients with smaller deletions affecting *AKT3* and *SDCCAG8* (0.16 and 0.18 Mb), one patient with a smaller quadruplication (1 Mb) that affects the entire *AKT3* gene and one patient with a translocation breakpoint at the 1q43-q44 region. All patients with deletions presented microcephaly without structural brain abnormalities, whereas the patient with quadruplication had macrocephaly but his carrier father had normal head circumference and the patient with the translocation had a normal head size. These observations, together with previously published cases and Decipher database entries reveal that *AKT3* genomic disruption doesn't necessarily follow the rule of "deletion equals microcephaly" and "duplication equals macrocephaly". In fact, there are several cases of small genomic imbalances inherited from asymptomatic progenitors, implicating *AKT3* as a contributor for the ID/DD phenotype but raising doubts about the consistency of its role in head size.

In this chapter we aim to contribute to the clarification the role of *AKT3* in 1q43q43 CNV-related neurodevelopmental disorders.

Introduction

The 1q43q44 microdeletion syndrome is characterized by variable degrees of intellectual disability (ID), growth retardation, microcephaly (MIC), anomalies of the corpus callosum (ACC), seizures (SZR) and abnormal facial features such as round face with low-set ears, prominent forehead and flat nasal bridge, epicanthal folds and hypertelorism (De Vries *et al.*, 2001; Ballif *et al.*, 2012). The first report of pathogenic deletions at 1q43-q44 described a large microscopically observed deletion in a female patient with motor and mental impairment, MIC, SZR and several dysmorphisms (Mankinen *et al.*, 1976). In the recent years, many cases with submicroscopic deletions in this region were reported (Boland *et al.*, 2007; Hill *et al.*, 2007; van Bon *et al.*, 2008; Orellana *et al.*, 2009; Nagamani *et al.*, 2012) and the consequent determination of the possible genes associated with the 1q43q44 deletion syndrome became possible. In 2012, Ballif and collaborators described a cohort of patients with 1q43-q44 microdeletions and have proposed three critical regions for MIC, ACC and SZR (Ballif *et al.*, 2012). Overall, the gene most consensually associated with head size anomalies is the V-AKT murine thymoma viral oncogene homolog 3 (*AKT3*) gene (Ballif *et al.*, 2012; Wang *et al.*, 2013; Chung *et al.*, 2014) while *CEP170* (CENTROSOMAL PROTEIN, 170-KD) and *ZBTB18* (ZINC FINGER- AND BTB DOMAIN-CONTAINING PROTEIN 18, also named *ZNF238*) were considered more likely causative of ACC (Nagamani *et al.*, 2012; Perlman *et al.*, 2013). More recently, single nucleotide variations in *ZNF238* were identified in two patients: one with developmental and speech delay, microcephaly and dysmorphic features and the other with severe ID, breathing disturbances and microcephaly without structural anomalies (de Munnik *et al.*, 2014; Lopes *et al.*, 2016). *HNRNPU* (HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN U), *HNRNPU-AS1* (HNRNPU antisense RNA 1) and *FAM36A* (FAMILY WITH SEQUENCE SIMILARITY 36, MEMBER A) were proposed as good candidates for the epilepsy and ID phenotype within the 1q43-q44 microdeletion syndrome (Thierry *et al.*, 2012; Poot and Kas, 2013).

Even though the vast majority of the 1q43-q44 deletion cases described so far with microcephaly (MIC) do carry genomic rearrangements that do disrupt the *AKT3* gene, there are two patients described in the literature with *AKT3* disruption that do not display MIC (Ballif *et al.*, 2012). Conversely, there are also patients with 1q43q44 deletion who display MIC even though they carry deletions that do not comprise the *AKT3* gene (Poot *et al.*, 2008; van Bon *et al.*, 2008; Ballif *et al.*, 2012). Therefore, despite its known biological function and the strong evidence that *AKT3* is a key gene for MIC in patients with 1q43-q44 deletions, other factors must play a role in

the arising of the phenotype, resulting in incomplete penetrance and variable expressivity, perhaps consequent to different genetic or epigenetic backgrounds of the individuals (van Bon *et al.*, 2008; Ballif *et al.*, 2012). Even though the Ballif et al collection of cases allowed the determination of candidate regions with a high amount of certainty, there were still patients described that did not fit the profile (Ballif *et al.*, 2012). In this perspective, the description of more patients with 1q43-q44 deletions is crucial for the delineation of more precise phenotype-genotype correlations.

We discuss five patients with 1q43q44 CNVs and try to establish genotype-phenotype correlations in the hope to bring further insight to the role of *AKT3* in brain abnormalities and ID.

Methods

Molecular karyotyping

Genomic DNA was extracted from peripheral blood using the Citogene® DNA isolation kit (Citomed, Portugal) for patient 1, QIAsymphony SP apparatus (QIAGEN GmbH, Germany) for patient 2 and 3 and ArchivePure DNA blood kit for patient 4. The aCGH hybridization and analysis was performed using: Patient 1 - aCGH Agilent 180K custom array (GEO GL15397, across-array methodology [(Buffart *et al.*, 2008; Krijgsman *et al.*, 2013), Nexus Copy Number 5.0 software with FASST Segmentation algorithm for data analysis; Patient 2 and 3 – Affymetrix CytoScan 750K Platform (750.000 markers distributed throughout the genome, with a medium resolution of 8-20Kb), Chromosome Analysis Suite (ChAS 3.0) software (Affymetrix); Patient 4 – Affymetrix CytoScanHD microarray platform.

Array paiting

For patient 5, array painting was used to characterize the breakpoints as previously described (Backx *et al.*, 2007). The q arm of the aberrant chromosome 1 and the p arm of the aberrant chromosome 9 were microdissected and amplified. The correctness of the location of the dissected chromosome was controlled by reverse painting to normal human metaphase chromosomes and spectrumOrange-dUTP was used to label the amplified dissected chromosomes by a second DOP-PCR amplification. After, the aberrant chromosomes were labeled with Cy5-dCTPs and a genome-wide array was performed. Arrays used in this study (clone preparation, hybridization, and data analysis) were constructed as described before (Vermeesch *et al.*, 2005).

Quantitative PCR confirmations

Primers for qPCR were designed using Primer3Plus software (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and taking into account standard recommendations for qPCR primer development (Jovanovic *et al.*, 2003).

Four sets of primers were designed for the *AKT3* gene (ENSG00000117020) for exon 7, 8, 9 and 10. The reference genes used were *SDC4* (ENSG00000124145) and *ZNF80* (ENSG00000174255) localized in the 20q12-q13 and 3p12 regions, respectively (primers designed for this purpose are listed in supplementary table 1). qPCR reactions were carried out in a 7500-FAST Real Time PCR machine (Applied Biosystems) using Power SYBR Green® (Applied Biosystems). The specificity of each of the reactions was verified by the generation of a melting curve for each of the amplified fragments. The primer efficiency was calculated by the generation of a standard curve fitting the accepted normal efficiency percentage. Quantification was performed as described elsewhere (Hoebeeck *et al.*, 2005). Ct values obtained for each test were analyzed in DataAssist ™ software (Applied Biosystems, Foster City, CA, USA).

Informed consent

Informed consent was obtained from the children's parents for blood sampling, genetic analyses and publication.

Results

Clinical description and molecular findings

Patient 1

A girl was evaluated at 9 years of age for learning disabilities and microcephaly (head circumference -2,5 SD). Height and weight were in the 25th centile. She had a mildly sloping forehead and large upper central incisors (figure 1). Evaluation with the Wechsler Intelligence Scale for Children (third edition) showed a full scale IQ of 63. Cerebral MRI was normal except for a discrete global atrophy. Pregnancy and delivery were uncomplicated at 35 weeks. Head circumference at birth was in the 3rd centile. Family history was unremarkable. Parents were young, healthy and non-consanguineous. The patient had a healthy younger brother. Peripheral blood chromosome analysis demonstrated a normal 46,XX karyotype. Congenital

cytomegalovirus infection was excluded using PCR on the DNA obtained from the newborn metabolic disease screening Guthrie card.

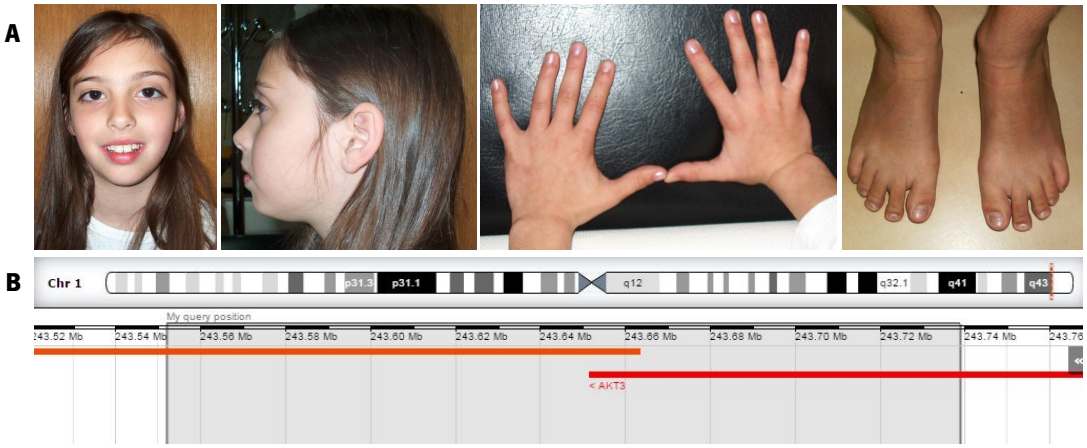


Figure 34 – Patient 1. Facial features of patient 1, hands and feet (A). Schematic representation of the 1q43-q44 deletion affecting the *AKT3* and *SDCCAG8* genes (B).

aCGH revealed a 0.18 Mb *de novo* deletion at chromosome region 1q43-q44 (chr1:243,552,007-243,738,675) containing 2 genes: *AKT3* and *SDCCAG8*. Additionally, she also showed a 0.12 Mb duplication at cytoband Xp11.4 but this alteration was not considered relevant for the phenotype since it did not contain genes and there was a control with a similar deletion described in DGV database. The 1q43-q44 deletion was confirmed by qPCR for the *AKT3* gene (exons 7, 8, 9 and 10). This analysis showed the deletion breakpoint to be located between exons 8 and 9. Analysis of the same fragments in both parents showed that the deletion occurred *de novo*. Mutation analysis by Sanger sequencing of *AKT3* showed no mutations.

Patient 2

Boy with 10y, intellectual disability and microcephaly.

aCGH revealed a 0.16 Mb deletion (chr1:243,592,147-243,749,968) affecting the both *AKT3* and *SDCCAG8* genes.

Patient 3

This boy was referred for evaluation around 3 to 4 months of age for evident microcephaly. Currently he presents weight and height within the normal range, but OFC has been below P5 since he was 8 months old. Concerning family history he has a maternal uncle with epilepsy. The patient presents a hyperkinetic behavior and global development delay, language being the most strikingly affected; he is currently undertaking speech therapy. On examination, the patient revealed a type B bilateral tympanogram which is consistent with a middle ear pathology. At the age of evaluation he didn´t know colors or numbers, was described by parents as “clumsy” and

by the teacher as aggressive. MRI evaluation of the brain demonstrated a cerebral volume that is in accordance with a decreased cephalic perimeter, without enlarged cerebrospinal fluid spaces (figure 2). The cerebral hemispheres appeared otherwise unremarkable without noticeable malformations of cortical development, no signs of hypoxic-ischemic or infectious lesions. The corpus callosum was completed formed and displayed a normal thickness. No abnormalities were seen in the posterior fossa. The white matter revealed a normal pattern of myelination. The ventricles and the subarachcnoid spaces had a normal configuration. There was no evidence of significant skull abnormalities, other than the identified smaller dimensions and a slight left positional plagiocephaly. EEG was also normal.

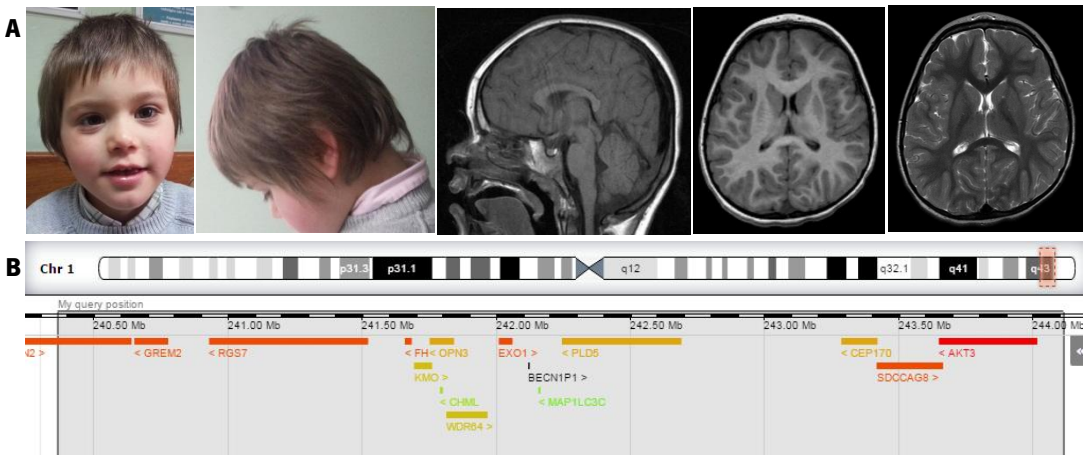


Figure 35 – Patient 3. (A) Facial features and brain imaging of patient 3. Notice the relatively small head (on the left) and MRI for T1-weighted SE sagital image showing a normal corpus callosum (on the right). (B) Schematic representation of the 1q43-q44 deletion.

aCGH revealed a 3.7 Mb deletion (chr1:240,366,425-244,111,022) that proved to be *de novo* after qPCR confirmation for the *AKT3* gene (same assay used for patient 1).

Patient 4

This boy was born at term, after an uneventful pregnancy and delivery, with poor Apgar scores, absence of gag reflex. Initially he presented hypotonia and apnea, having developed seizures and dystonia afterwards. He has developmental delay, macrocephaly (MAC) and white matter lesions of brain including thalamic lesions.

aCGH revealed a 1 Mb paternally inherited quadruplication (breakpoints 242,950,674 - 243,810,942), the patient´s father being considered healthy by the referring clinician (no MAC). Meanwhile, he underwent trio whole exome sequencing analysis which retrieved no diagnosis.

Patient 5

This girl, 7 years old at the time of observation, presented with ID, developmental delay (especially of her speech) and feeding problems. She was born to healthy parents, and family history and pregnancy were unremarkable. Growth parameters are within the normal range but head circumference is at the 25th centile. She follows special school and has concentration problems. She has a happy character and good social skills. She is dysmorphic with a broad and short nose with beaked nose tip and prominent columella, thin upper lip and large, low-set ears. She also has short terminal phalanges and a hyper nasal speech without velopharyngeal insufficiency. MRI scan for the patient revealed no abnormalities.

The patient has a *de novo* t(1;9) (q44;p13.3) and aCGH revealed no additional pathogenic genomic imbalances. Breakpoint cloning was performed by arraypainting. On chromosome 9p, the breakpoint is located in a gene-desert at 9p21.1 (between RP11-774M21 and RP11-492E22). Since no genes are located in the 2 Mb interval surrounding the 9p21.1 breakpoint, this suggests that the 1q44 breakpoint is the one most involved in the patient's phenotype. This is located on chromosome 1q44, between clones RP11-655L13 and RP11-705K21. The *ZNF238* gene is located in this interval and might also be disrupted by the breakpoint (BP). The corrected karyotype based on these data is t(1;9) (q44; p21.1). Quantitative PCR was performed for exons 7, 8, 9 and 10 revealing no changes in dosage for any of these exons. Mutation analysis of *ZNF238* and *AKT3* showed no mutations on the other allele. Real time quantitative polymerase chain reaction (qRT-PCR) on mRNA extracted from EBV immortalized lymphocytes of the patient was performed, to explore a potential difference in expression of the *AKT3* or *ZNF238* genes; this analysis showed a reduced expression for the *AKT3* gene on mRNA of the patient, but in two controls out of five as well. An extremely low expression of *ZNF238* in Epstein-Barr virus (EBV) immortalized lymphocytes hampered an accurate quantitative expression analysis.

Table I presents a summary and comparison of the clinical features of the patients.

Table II summarizes the molecular findings present in the patients described here.

Table VIII - Comparison of the clinical features of the patients in the current series with patients with *AKT3* deletions described in the literature.

	Clinical feature	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Van Bon, 2008, Patient 12	Van Bon, 2008, Patient 13	Nagamani, 2012 Patient 6	Gai, 2015
	Aberration	Deletion	Deletion	Deletion	Quadruplication	Translocation breakpoint	Deletion	Deletion	Deletion	Deletion
Clinical Overview	**Gender**	♀	♂	♂	♂	♀	♀	♀	♀	♂
	Consanguinity	No	No	No	No	No	No	No	NA	NA
	Birth	35w (uneventful)	NA	40 w	NA	NA	At term	At term	NA	NA
	Measurements at birth (heigh, weigth, OFC)		NA	2710g/47 cm/32,5 cm	NA	NA	2250g (-2.5SD)	2900	NA	3940g (50-57th centile)
	Age of observation	9y	NA	3-4 m (Evident microcephaly)	NA	NA	5m, 4y, 7y, 8y	10m, 1y, 7y	7.5y	4y
	ID	Mild (IQ=63)	NA	Mild (IQ=61)	NA	NA	NA	IQ between 40-60	Mild DD	IQ between 70-79
	Weigth (centile)	P25	NA	P3-P98	NA	NA	NA	NA	P57	17Kg 10-25th centile)
	Heigth (centile)	P25	NA	P3-P98	NA	NA	88cm (-5.5SD) at 8y	Normal	P75	116cm (90th centile)
Cranio-facial abnormalities	**Head/ OFC (centile)**	Microcephaly, P<3	NA	Microcephaly (P<3)	NA	NA	Microcephaly (-4.5SD)	Microcephaly (-2.6SD)	Normal, P50	Microcephaly, 2nd centile
	Structural brain abnormalities	No	NA	No	NA	NA	Thin corpus callosum, generalized atrophy, wide ventricles, myelination abnormalities, brain calcifications	No anomalies (brain CT)	No	pineal cyst (13 mm X 11 mm)

(Cont.)

	Facial dysmorphisms	Mildly sploping forehead; large upper incisors	NA	Upward palpebral fissures; retrognatia; poor hearing (conduction); dental caries	NA	NA	Round face, small nose, low set ears	No major dysmorphisms	No	Sloping skull contour, large ears, synophrys, mild micrognatia
Others	**Behavior**	No alterations	NA	Aggressivity; Auto-mutilation hyperactivity	NA	NA	NA	NA	Anxiety	No
	Seizures	No	NA	No	NA	NA	Yes (developed at 4y)	No	No	No
	Hypotonia	No	NA	No	NA	NA	NA	NA	No	Yes
	Genitalia	Normal	NA	Normal	NA	NA	NA	NA	NA	NA
	Heart	NA	NA	Never evaluated	NA	NA	Pericardial effusion with congestive heart failure (cause of death)	NA	Normal	NA

NA: not available;

Table IX - Comparison of the molecular findings of the patients in the current series with patients with *AKT3* deletions described in the literature.

Molecular findings	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Van Bon, 2008, Patient 12	Van Bon, 2008, Patient 13	Nagamani, 2012 Patient 6	Gai, 2015
Aberration	Deletion	Deletion	Deletion	Quadruplication	Translocation breakpoint	Deletion	Deletion	Deletion	Deletion
Heritability	*de novo*	ND	*de novo*	paternal	*de novo*	*Maternal* (unaffected mother)	*Maternal*	Not maternal, Father unavailable	*Paternal* (unaffected father)
Confirmation	qPCR	qPCR	qPCR	ND	arraypainting	SNP array	ND	NA	FISH
Start (hg19)	243,552,007	243,592,147	240,366,425	243,415,063	NA	NA	NA	243,439,056-243,456,375	244,182,939-244,229,143
End (hg19)	243,738,675	243,749,968	244,111,022	244,478,355	NA	NA	NA	243,668,684-243,675,405	243,724,990-243,786,019
Size (Mb)	0.18	0.16	3.7	1	NA	0.4	0.4	0.23	0.4
Genes affected	*AKT3, SDCCAG8*	*AKT3, SDCCAG8*	*FMN2, GREM2, RGS7, FH, KMO, OPN3, CHML, WDR64, EXO1, MAP1LC3C, PLD5, LOC731275, CEP170, SDCCAG8, AKT3*	*CEP170, AKT3, SDCCAG8, ZNF238*	*AKT3*	*AKT3, SDCCAG8*	*AKT3, SDCCAG8*	*AKT3* (only the last exon), *SDCCAG8*	*AKT3*

NA: not available; ND: not determined

Figure 3 displays the relative size and position of the genomic imbalances detected in the five patients.

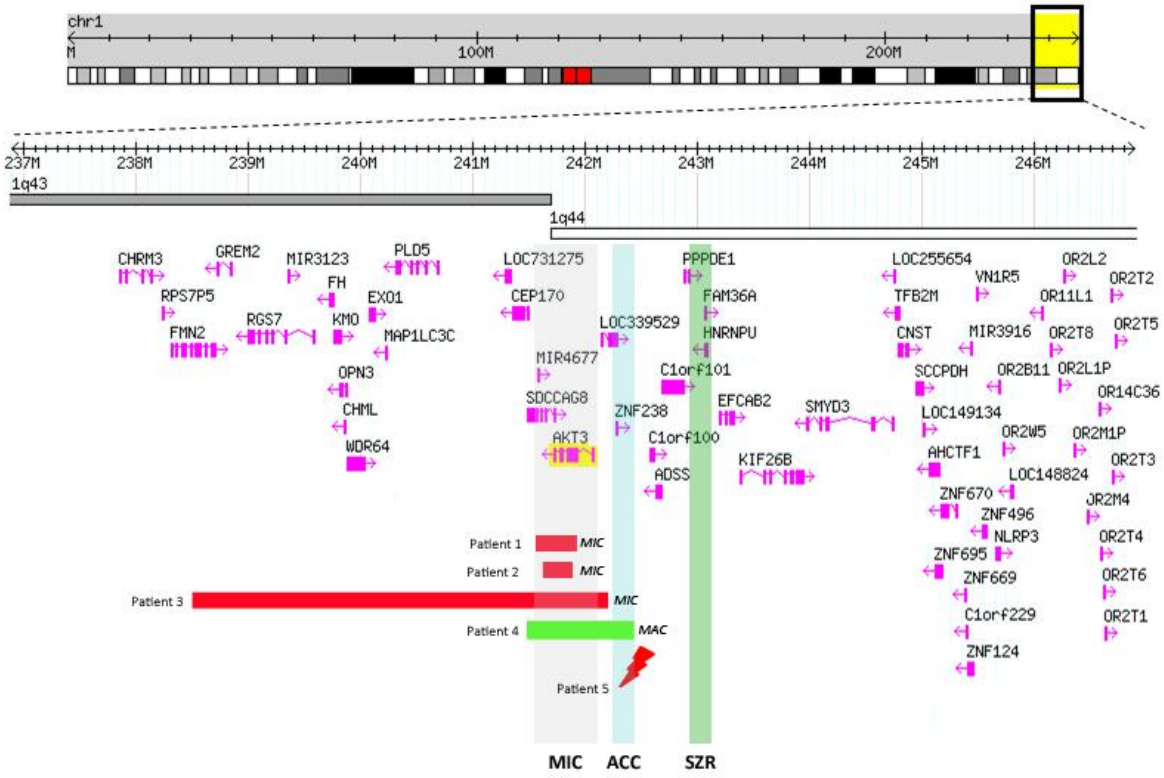


Figure 36 - Schematic representations of the CNVs found in the patients and overlap with the critical regions proposed by Ballif B. *et al*, 2011. A 10 Mb genomic portion encompassing cytobands 1q43-q44 is shown. RefSeq genes (in pink and the transcriptional direction is shown by the arrows) present within the genomic region are shown. Shaded in grey is the proposed critical region for microcephaly (MIC) (affecting *AKT3* gene), in blue the critical region for anomalies of the corpus callosum (ACC) (affecting the *ZNF238* gene) and in green the critical region for seizures (SZR) (affecting the *C1ORF199* gene). Individual red horizontal bars represent deletions and the green bar a quadruplication. In each CNV the corresponding patient is indicated.

Discussion

Patients with CNVs in the 1q43-q44 citoband have been described in literature, however a significant variability is observed (Boland *et al.*, 2007; Orellana *et al.*, 2009; Caliebe *et al.*, 2010; Lall *et al.*, 2011; Ballif *et al.*, 2012; Nagamani *et al.*, 2012; Wang *et al.*, 2013). Considering the data published in the last years, it was inferred that deletions affecting either *CEP170* or *ZNF238* genes often result in CCA (corpus callosum anomalies), while CNVs encompassing *AKT3* some of the genes located in 1q43-q44 region were likely implicated with OFC (occipitofrontal circumference) alterations (Ballif *et al.*, 2012; Nagamani *et al.*, 2012; Wang *et al.*, 2013). Additionally, it was reported that 1q43q44 deletions expanding towards the centromere might be implicated in delayed myelination (Shimojima *et al.*, 2012). Both deletions and translocation breakpoints have been described in 1q44 affecting the *AKT3* gene in patients with MIC and agenesis of the corpus callosum.

Nagamani and colleagues defined three potentially critical regions for MIC, ACCs and SZR (represented in figure 2); the authors described the association of MIC with deletion of the *AKT3* gene in 93% of the cases; ACCs in 86% of the cases with deletion affecting *ZNF238* gene and of SZR with deletions affecting the *FAM36A*, *C1ORF199* and *HNRNPU* in genes 87% of the patients (Ballif *et al.*, 2012). In this work, however, not all the patients who had a deletion mapping in this region displayed agenesis of the corpus callosum. Some of them presented MIC without anomalies in corpus callosum, particularly one patient (referred to as patient 5 in the publication), who had an intragenic deletion of *AKT3* without evidence of CCAs (Nagamani *et al.*, 2012). More recently, it has been proposed that deletions affecting solely *AKT3* may be incompletely penetrant, and often inherited, while larger 1q43q44 deletions occured mostly *de novo* and appeared to be fully penetrant (Gai *et al.*, 2015).

The two cases described that did not fit the proposed model of *AKT3*'s causative role in MIC were (i) a family in which the two differently affected siblings and their healthy mother carried an *AKT3* deletion described by van Bon and colleagues (van Bon *et al.*, 2008), and (ii) a patient described by Nagamani and colleagues (referred to as patient 6 in the original publication) with a deletion affecting the 2 last exons of the *AKT3* gene (considering the largest transcript), without MIC or other structural brain abnormalities (Nagamani *et al.*, 2012). More recently, a deletion affecting exclusively the *AKT3* gene was described in a patient with MIC and ID and in his asymptomatic father, being the first report of a paternally inherited pure *AKT3* deletion of incomplete penetrance (Gai *et al.*, 2015). A comparison of these three cases with the ones described here is made in tables I and II. Furthermore, there are cases in Decipher with deletions inherited from parents (Decipher#317423, Decipher#252432 and Decipher#277172, the two latter ones being reportedly healthy progenitors) and two cases

(Decipher#277653 and Decipher#257255) carrying deletions inherited from parents with a similar phenotype (detailed description in table III). Possibly, in both the families reported in Decipher and in our patients 4 and 5, additional factors must be playing a role, supporting the concept that *AKT3* partial or pure deletions may be subject to incomplete penetrance and/or differential expressivity driven by the different genetic and epigenetic backgrounds of the individuals.

Table X – Summary of the Decipher patients with overlapping CNVs.

Decipher number	Del/Dup	Size (Mb)	Genes affected	Inheritance	Pathogenicity	Index phenotype	Parent phenotype
317423	Deletion	0.09	*AKT3*	Paternal	Likely pathogenic	Global developmental delay, microcephaly, autism	Not described
307197	Deletion	0.1	*SDCCAG8*	Unknown	Likely pathogenic	Cognitive impairment	-
252432	Deletion	0.12	*SDCCAG8*, *AKT3*	Inherited from normal parent	Not described	Not described	Normal
277653	Deletion	0.16	*AKT3*	Inherited from parent with similar phenotype to child	Not described	Cognitive impairment, microcephaly	Cognitive impairment
257255	Deletion	0.2	*CEP170*, *SDCCAG8*, *AKT3*	Inherited from parent with similar phenotype to child	Not described	Not described	Not described
277172	Deletion	0.27	*SDCCAG8*, *AKT3*	Inherited from normal parent	Not described	Not described	Normal
252434	Deletion	0.35	*CEP170*, *SDCCAG8*, *AKT3*	De novo	Not described	Not described	-
301634	Deletion	1.6	*SDCCAG8*, *AKT3* + 114 genes	De novo	Pathogenic	Intellectual disability, absent speech, movement abnormality	-
331004	Duplication	0.3	*SDCCAG8*, *AKT3*	Unknown	Likely pathogenic	Global developmental delay	-

In this study, the findings in patients 1, 2 and 3 would support the conclusion that haploinsufficiency of *AKT3* gene is indeed associated with microcephaly. Patient 3 fits quite well in the established 1q43-q44 microdeletion syndrome regarding the phenotype and size of the deletion. Comparing the patient described by Nagamani et al in 2011 (patient 6 in the publication) and patients 1 and 2 (the ones with the smaller deletions) of this report, MIC may be explained by the presence of a deletion that affects a bigger portion of *AKT3* (it affects the last 4 exons of *AKT3,* while the one present in the Naganami case only affects its last 2 exons), as well as by the additional variables hypothesized before. Moreover,

patient 4 in our series might indicate that *AKT3* quadruplication might not always be associated with the mirror phenotype (macrocephaly). As for patient 3, despite having a larger deletion that includes not only the *AKT3* gene but also the *CEP170* gene, he has evident MIC but an apparently normal corpus callosum. Therefore, although there seems to be a connection between *AKT3* haploinsufficiency and MIC, the role of yet unknown additional factors that may lead to differences regarding the presence and severity of symptoms should be also emphasized. In patient 5, the breakpoint is located in an interval including the *AKT3* and *ZNF238* genes. Quantitative PCR located in internal exons of *AKT3* did not revealed any CNVs, however the 5'UTR and 3'UTR regions were not tested yet. Expression analysis of the *AKT3* and *ZNF238* genes could not support nor exclude their contribution to the patient's phenotype. It was only possible to assess the expression in one lymphoblastoid cell line, which might not recapitulate the expression pattern of the brain, the primary tissue of interest concerning the neurodevelopmental phenotype. Because *ZNF238* expression was not possible to detect in the periphery and that of *AKT3* was reduced in the patient but also in two of the five healthy controls tested, our results were not conclusive regarding the contribution of this gene for the neurodevelopment perturbation seen in patient 5. However, if we assume that the observed expression in the periphery reflects that of the nervous systems, we could say that reduced expression of *AKT3* alone is not sufficient to cause MIC, since patient 5 has a normal head size.

The minimal overlapping region of all the patients with del/dups described in this study encompasses only one additional gene to *AKT3*: *SDCCAG8* (serological defined colon cancer antigen 8). This gene encodes for a protein thought to be a stable centrosomal component with a structural role in the centrosomal architecture or the microtubule-organizing activities of the centrosome matrix (Kenedy *et al.*, 2003). Mutations in *SDCCAG8* were described in patients with nephronophthisis-related ciliopathies; even though the clinical features of these patients include ID, the alterations causing disease in those cases are recessive (Otto *et al.*, 2010). Recessive variants in *SDCCAG8* gene were also associated with Bardet-Biedl syndrome and with an increased risk for schizophrenia (Schaefer *et al.*, 2011; Hamshere *et al.*, 2013). More recently, *SDCCAG8* has been described to play a role in neuronal polarization and migration of the developing mouse cortex (Insolera *et al.*, 2014), which would be consistent with the described genetic effects in humans. The AKT3 protein belongs to the protein kinase B (PKB/Akt) family, involved in cell survival, proliferation and growth (Nakatani *et al.*, 1999). In mice, both Akt1 and Akt3 play a role in determination of organ size. However, while Akt1 null mice have a decrease of all the organs, Akt3 null mice have a selective 20% decrease in brain size, Akt3 being the predominant Akt protein expressed in cortex and hippocampus. Unfortunately, the authors showed no data concerning

the brain size in heterozygous animals which would be relevant for the interpretation the findings in humans (Easton *et al.*, 2005). Akt3-null and heterozygous mice also have angiogenesis impairment, showing a dose-dependent reduction in vessel number (5 fold decrease in homozygous and 2.5 fold decrease in heterozygous), something that was never evaluated in any of the reported patients (Corum *et al.*, 2014).

Conclusion

In summary, we describe five patients with 1q43q44 deletions, three of which with quite consistent outcomes with those of the perceived core 1q43q44 deletions affecting *AKT3* phenotype, and two others (patients 4 and 5) that, in combination with five Decipher cases, raise doubts about the straightforward implication of *AKT3* copy number changes in human OFC. This phenotypic variability may result from the nature of the genomic imbalance (deletions, a quadruplication and a translocation breakpoint), different deletion sizes (even though very similar, the size is not exactly the same between the cases), sequence variants present in the other allele, possible modifier loci located somewhere else in the genome or even sex influenced penetrance. In fact, the small differences in the deletion sizes also raise the possibility of the existence of regulatory elements nearby and within the *AKT3* gene that, when compromised, may result in different outcomes in brain development. The implication of *AKT3* haploinsufficiency in MIC without ACCs appears to be clear for the vast majority of the cases, even though not absolute.

References

Backx L, Van Esch H, Melotte C, Kosyakova N, Starke H, Frijns J-P, et al. Array painting using microdissected chromosomes to map chromosomal breakpoints. Cytogenet. Genome Res. 2007; 116: 158–166.

Ballif BC, Rosenfeld JA, Traylor R, Theisen A, Bader PI, Ladda RL, et al. High-resolution array CGH defines critical regions and candidate genes for microcephaly, abnormalities of the corpus callosum, and seizure phenotypes in patients with microdeletions of 1q43q44. Hum. Genet. 2012; 131: 145–156.

Boland E, Clayton-Smith J, Woo VG, McKee S, Manson FDC, Medne L, et al. Mapping of deletion and translocation breakpoints in 1q44 implicates the serine/threonine kinase AKT3 in postnatal microcephaly and agenesis of the corpus callosum. Am. J. Hum. Genet. 2007; 81: 292–303.

Buffart TE, Israeli D, Tijssen M, Vosse SJ, Mrsić A, Meijer GA, et al. Across array comparative genomic hybridization: a strategy to reduce reference channel hybridizations. Genes. Chromosomes Cancer 2008; 47: 994–1004.

Caliebe A, Kroes HY, van der Smagt JJ, Martin-Subero JI, Tönnies H, van 't Slot R, et al. Four patients with speech delay, seizures and variable corpus callosum thickness sharing a 0.440 Mb deletion in region 1q44 containing the HNRPU gene. Eur. J. Med. Genet. 2010; 53: 179–185.

Chung BK, Eydoux P, An Karnebeek CD, Gibson WT. Duplication of AKT3 is associated with macrocephaly and speech delay. Am. J. Med. Genet. A. 2014; 164: 1868–1869.

Corum DG, Tsichlis PN, Muise-Helmericks RC. AKT3 controls mitochondrial biogenesis and autophagy via regulation of the major nuclear export protein CRM-1. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. 2014; 28: 395–407.

De Vries BB, Knight SJ, Homfray T, Smithson SF, Flint J, Winter RM. Submicroscopic subtelomeric 1qter deletions: a recognisable phenotype? J. Med. Genet. 2001; 38: 175–178.

Easton RM, Cho H, Roovers K, Shineman DW, Mizrahi M, Forman MS, et al. Role for Akt3/protein kinase Bgamma in attainment of normal brain size. Mol. Cell. Biol. 2005; 25: 1869–1878.

Gai D, Haan E, Scholar M, Nicholl J, Yu S. Phenotypes of AKT3 deletion: a case report and literature review. Am. J. Med. Genet. A. 2015; 167A: 174–179.

Hamshere ML, Walters JTR, Smith R, Richards AL, Green E, Grozeva D, et al. Genome-wide significant associations in schizophrenia to ITIH3/4, CACNA1C and SDCCAG8, and extensive replication of associations reported by the Schizophrenia PGC. Mol. Psychiatry 2013; 18: 708–712.

Hill AD, Chang BS, Hill RS, Garraway LA, Bodell A, Sellers WR, et al. A 2-Mb critical region implicated in the microcephaly associated with terminal 1q deletion syndrome. Am. J. Med. Genet. A. 2007; 143A: 1692–1698.

Hoebeeck J, van der Luijt R, Poppe B, De Smet E, Yigit N, Claes K, et al. Rapid detection of VHL exon deletions using real-time quantitative PCR. Lab. Investig. J. Tech. Methods Pathol. 2005; 85: 24–33.

Insolera R, Shao W, Airik R, Hildebrandt F, Shi S-H. SDCCAG8 regulates pericentriolar material recruitment and neuronal migration in the developing cortex. Neuron 2014; 83: 805–822.

Jovanovic L, Delahunt B, McIver B, Eberhardt NL, Grebe SKG. Optimising restriction enzyme cleavage of DNA derived from archival histopathological samples: an improved HUMARA assay. Pathology (Phila.) 2003; 35: 70–74.

Kenedy AA, Cohen KJ, Loveys DA, Kato GJ, Dang CV. Identification and characterization of the novel centrosome-associated protein CCCAP. Gene 2003; 303: 35–46.

Krijgsman O, Israeli D, van Essen HF, Eijk PP, Berens MLM, Mellink CHM, et al. Detection limits of DNA copy number alterations in heterogeneous cell populations. Cell. Oncol. Dordr. 2013; 36: 27–36.

Lall M, Thakur S, Puri R, Verma I, Mukerji M, Jha P. A 54 Mb 11qter duplication and 0.9 Mb 1q44 deletion in a child with laryngomalacia and agenesis of corpus callosum. Mol. Cytogenet. 2011; 4: 19.

Lopes F, Barbosa M, Ameur A, Soares G, de Sá J, Dias AI, et al. Identification of novel genetic causes of Rett syndrome-like phenotypes. J. Med. Genet. 2016; 53: 190–199.

Mankinen CB, Sears JW, Alvarez VR. Terminal (1)(q43) long-arm deletion of chromosome no. 1 in a three-year-old female. Birth Defects Orig. Artic. Ser. 1976; 12: 131–136.

de Munnik SA, García-Miñaúr S, Hoischen A, van Bon BW, Boycott KM, Schoots J, et al. A de novo non-sense mutation in ZBTB18 in a patient with features of the 1q43q44 microdeletion syndrome. Eur. J. Hum. Genet. EJHG 2014; 22: 844–846.

Nagamani SCS, Erez A, Bay C, Pettigrew A, Lalani SR, Herman K, et al. Delineation of a deletion region critical for corpus callosal abnormalities in chromosome 1q43-q44. Eur. J. Hum. Genet. EJHG 2012; 20: 176–179.

Nakatani K, Sakaue H, Thompson DA, Weigel RJ, Roth RA. Identification of a human Akt3 (protein kinase B gamma) which contains the regulatory serine phosphorylation site. Biochem. Biophys. Res. Commun. 1999; 257: 906–910.

Orellana C, Roselló M, Monfort S, Oltra S, Quiroga R, Ferrer I, et al. Corpus callosum abnormalities and the controversy about the candidate genes located in 1q44. Cytogenet. Genome Res. 2009; 127: 5–8.

Otto EA, Hurd TW, Airik R, Chaki M, Zhou W, Stoetzel C, et al. Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy. Nat. Genet. 2010; 42: 840–850.

Perlman SJ, Kulkarni S, Manwaring L, Shinawi M. Haploinsufficiency of ZNF238 is associated with corpus callosum abnormalities in 1q44 deletions. Am. J. Med. Genet. A. 2013; 161A: 711–716.

Poot M, Kas MJ. Antisense may make sense of 1q44 deletions, seizures, and HNRNPU. Am. J. Med. Genet. A. 2013; 161A: 910–912.

Poot M, Kroes HY, Hochstenbach R. AKT3 as a candidate gene for corpus callosum anomalies in patients with 1q44 deletions. Eur. J. Med. Genet. 2008; 51: 689–690.

Schaefer E, Zaloszyc A, Lauer J, Durand M, Stutzmann F, Perdomo-Trujillo Y, et al. Mutations in SDCCAG8/NPHP10 Cause Bardet-Biedl Syndrome and Are Associated with Penetrant Renal Disease and Absent Polydactyly. Mol. Syndromol. 2011; 1: 273–281.

Shimojima K, Okamoto N, Suzuki Y, Saito M, Mori M, Yamagata T, et al. Subtelomeric deletions of 1q43q44 and severe brain impairment associated with delayed myelination. J. Hum. Genet. 2012; 57: 593–600.

Thierry G, Bénéteau C, Pichon O, Flori E, Isidor B, Popelard F, et al. Molecular characterization of 1q44 microdeletion in 11 patients reveals three candidate genes for intellectual disability and seizures. Am. J. Med. Genet. A. 2012; 158A: 1633–1640.

van Bon BWM, Koolen DA, Borgatti R, Magee A, Garcia-Minaur S, Rooms L, et al. Clinical and molecular characteristics of 1qter microdeletion syndrome: delineating a critical region for corpus callosum agenesis/hypogenesis. J. Med. Genet. 2008; 45: 346–354.

Vermeesch JR, Melotte C, Froyen G, Van Vooren S, Dutta B, Maas N, et al. Molecular karyotyping: array CGH quality criteria for constitutional genetic diagnosis. J. Histochem. Cytochem. Off. J. Histochem. Soc. 2005; 53: 413–422.

Wang D, Zeesman S, Tarnopolsky MA, Nowaczyk MJM. Duplication of AKT3 as a cause of macrocephaly in duplication 1q43q44. Am. J. Med. Genet. A. 2013; 161A: 2016–2019.

SUB-CHAPTER 2.3

Phenotypic and functional consequences of haploinsufficiency of genes from exocyst and retinoic acid pathway due to a recurrent microdeletion of 2p13.2

Wen et al. Orphanet Journal of Rare Diseases 2013, 8:100
http://www.ojrd.com/content/8/1/100

ORPHANET JOURNAL OF RARE DISEASES

RESEARCH Open Access

Phenotypic and functional consequences of haploinsufficiency of genes from exocyst and retinoic acid pathway due to a recurrent microdeletion of 2p13.2

Jiadi Wen[1†], Fátima Lopes[2,3†], Gabriela Soares[4], Sandra A Farrell[5], Cara Nelson[6], Ying Qiao[1], Sally Martell[1], Chansonette Badukke[1], Carlos Bessa[2,3], Bauke Ylstra[7], Suzanne Lewis[8], Nina Isoherranen[6], Patricia Maciel[2,3*] and Evica Rajcan-Separovic[1*]

Abstract

Background: Rare, recurrent genomic imbalances facilitate the association of genotype with abnormalities at the "whole body" level. However, at the cellular level, the functional consequences of recurrent genomic abnormalities and how they can be linked to the phenotype are much less investigated.

Method and results: We report an example of a functional analysis of two genes from a new, overlapping microdeletion of 2p13.2 region (from 72,140,702-72,924,626). The subjects shared intellectual disability (ID), language delay, hyperactivity, facial asymmetry, ear malformations, and vertebral and/or craniofacial abnormalities. The overlapping region included two genes, *EXOC6B* and *CYP26B1*, which are involved in exocytosis/Notch signaling and retinoic acid (RA) metabolism, respectively, and are of critical importance for early morphogenesis, symmetry as well as craniofacial, skeleton and brain development. The abnormal function of *EXOC6B* was documented in patient lymphoblasts by its reduced expression and with perturbed expression of Notch signaling pathway genes *HES1* and *RBPJ*, previously noted to be the consequence of *EXOC6B* dysfunction in animal and cell line models. Similarly, the function of *CYP26B1* was affected by the deletion since the retinoic acid induced expression of this gene in patient lymphoblasts was significantly lower compared to controls (8% of controls).

Conclusion: Haploinsufficiency of CYP26B1 and EXOC6B genes involved in retinoic acid and exocyst/Notch signaling pathways, respectively, has not been reported previously in humans. The developmental anomalies and phenotypic features of our subjects are in keeping with the dysfunction of these genes, considering their known role. Documenting their dysfunction at the cellular level in patient cells enhanced our understanding of biological processes which contribute to the clinical phenotype.

Keywords: 2p13 deletion, EXOC6B, CYP26B1, Developmental delay, Cranial/skeletal anomalies

* Correspondence: pmaciel@ecsaude.uminho.pt; eseparovic@cw.bc.ca
†Equal contributors
[2]Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, Portugal
[1]Child and Family Research Institute, Department of Pathology, University of British Columbia, Vancouver, BC, Canada
Full list of author information is available at the end of the article

Background

Chromosomal abnormalities involving the 2p13 chromosomal region were detected previously in subjects with developmental delay using traditional chromosome testing [1,2]. Abnormal features present in at least 50% of cases included abnormal head size/shape, nose, ears, chest and vertebral, digital and genital anomalies. The involvement of the *CYP26B1* gene was suggested in a subject with a *de novo* inversion involving 2p13 and 2q34, who had a Klippel-Fiel anomaly (vertebral fusion of cervical spine), psychomotor retardation, speech limitation, facial asymmetry, ear abnormalities, and scoliosis [3,4]. Disruption of *EXOC6B* and its fusion with *TNS3* from 7p12 due to a translocation t(2;7)(p13;p12), was considered to be a possible cause of the intellectual disability, ADHD and congenital abnormalities (renal malformation, microcephaly, long bone diaphyseal broadening) in a subject reported by Borsani *et al.* [5].

CYP26B1 (cytochrome P450, family 26, subfamily B, polypeptide 1) is one of the three CYP26 gene isoforms (*CYP26B1*, *CYP26A1* and *CYP26C1*) which encode the cytochrome-P450 enzymes that catabolize retinoic acid (RA) [6]. RA is the principal active metabolite of vitamin A and is an essential component of cell-cell signaling during vertebrate organogenesis [7]. Too little or too much RA causes the human malformation syndrome associated with vitamin A deficiency (VAD) or RA embryopathy, which includes craniofacial (e.g. ears, eyes, facial asymmetry), central nervous, musculoskeletal, and urogenital abnormalities [8-12]. Homozygous knockout of *CYP26B1* in animals has been associated with postnatal mortality, abnormal craniofacial, limb and gonadal development [13,14], while a conditional deletion (controlled expression) of *CYP26B1* resulted in a less severe phenotype. The difference in the phenotypes was attributed to differing levels of activated retinoid signaling [15]. Although deletions of the whole *CYP26B1* gene have not yet been reported in humans, homozygous point mutations of *CYP26B1* were reported in two families, resulting in prenatal and early postnatal lethality, skeletal and craniofacial abnormalities, fusion of long bones, calvarial bone hypoplasia and sutural defects, resulting in craniostenosis and brachycephaly [16]. *CYP26B1* mutations in both families led to a significantly attenuated ability to metabolize exogenously applied retinoic acid, confirming the impact of dysfunction of this gene on RA metabolism. Previous studies have shown that *CYP26B1* is responsible for ATRA (all-trans retinoic acid) clearance in several tissues [17]. For example in T-cells, *CYP26B1* is the only CYP26 enzyme up-regulated by ATRA and its expression regulates retinoic acid dependent signals in T-cells [18]. In addition, in many human cell lines *CYP26B1* mRNA is inducible by retinoic acid treatment while in rodents cyp26b1 expression correlates with dietary vitamin A intake [19-22].

EXOC6B (exocyst complex component 6B) encodes a protein homologous to Sec15 in *Saccharomyces cervisiae*. It belongs to a multiprotein complex (exocyst) required for targeted secretion (exocytosis) which is crucial for cell polarity, growth and communication [23]. In *Drosophila*, sec15 promotes Notch signaling, through specific vesicle trafficking of delta ligand, and has a role in asymmetric division of sensory precursors and neuronal fate determination [24]. *Drosophila* neurons with sec15 mutations show loss of synaptic specificity and mislocalization of proteins known to affect synaptic specificity in photoreceptors [25]. Recently, Guichard *et al.* showed reduced transcription of Notch signaling effectors *HES1* and *RBPJ* in knockdown *Drosophila* sec15 and in a human brain microendothelial cell line with abnormal *EXOC6*B function [26]. Notch signaling plays a pivotal role in the regulation of many fundamental cellular processes, such as acquisition of specific fates in context-dependent manner, differentiation and lineage decisions during embryonic development, neurogenesis, as well as morphogenesis involving regulation of left-right asymmetry [27,28]. Perturbations of the Notch pathway have been reported in human developmental disorders which demonstrate a variety of symptoms, including ID and/or skeletal abnormalities (e.g. Allagile syndrome and spondylocostal dysostosis [29,30]).

We report the first description of the clinical and functional consequences of hemizygous deletion of the two genes, *CYP26B1* and *EXOC6B,* located in chromosome region 2p13, and which are respectively, involved in retinoic acid and Notch signaling pathways of critical importance to normal human fetal development.

Materials and methods

Whole genome array CGH analysis

Genomic DNA was extracted from peripheral blood using the PUREGENE DNA Isolation Kits (Gentra, Minneapolis, MN) for Subject 1 and Citogene® DNA isolation kit (Citomed, Portugal) for Subject 2**.** For Subject 1, Agilent 105 K array (version 4.0, June 2006, Agilent Technologies, CA, USA) analysis was performed as previously described.[25] CNV selection was done by Agilent DNA Analytics (version 3.5.14, Agilent Technologies) using the ADM-2 algorithm (cutoff 6.0), followed by a filter to select regions with three or more adjacent probes and a minimum average log2 ratio + 0.25 [31]. The deletion of Subject 1 was *de novo,* as determined by FISH using probe RP11-91 F23.

For Subject 2, the aCGH analysis was performed using a CGH Agilent 180 K custom array design accessible through the gene expression omnibus (GEO) accession number GPL15397. Previously published protocol was

used [32]. Image analysis was performed using the across-array methodology described previously [33]. CGH data was analyzed using Nexus Copy Number 5.0 software with FASST Segmentation algorithm. The deletion was determined to be de novo using the same array for both parents.

For both cases, the array design, database consultation and comparative analysis was performed using genome build 36.1/HG18.

Functional studies

Immortalized EBV transformed lymphoblastoid cell lines (LCL) from Subject 1 were cultured in RPMI media containing 10% FBS, 50 units/mL penicillin, 50 μg/mL streptomycin. Cells were maintained in a humidified 37°C incubator with 5% CO_2. Control cells were obtained from healthy subjects, the majority of which aged 28–50 years.

RNA expression

Total RNA was extracted using an RNeasy Plus Mini kit (Qiagen) from Subject 1-derived LCLs and whole blood collected in Tempus tubes. Aliquots (~500 ng) of the total RNA extracts prepared from were subsequently reverse-transcribed into cDNA using GeneAmp Gold RNA PCR Core Kit (Applied Biosystems).

The expression of *EXOC6B* and *CYP26B1* was first assessed in RNA extracted from control whole blood and control LCLs by real-time qPCR using the ABI PRISM 7300 Sequence Detection System (Perkin-Elmer Applied Biosystems). The specific nucleotide sequences for primers of *EXOC6B* and *CYP26B1* were as follows: *EXOC6B* Forward 5'-GAC CTC ATT GCC TTT CTT CGT A-3', Reverse 5'-CAA GCT GAC ATA CAC GCT GT-3'(mapped to exons 18–19); *CYP26B1* primer set 1; Forward 5'-ACA CAG GGC AAG GAC TAC T-3', Reverse: 5'-GCA TAG GCC GCA AAG ATC A-3'(mapped to exon 4–5); *CYP26B1* primer set 2: Forward: 5'-CTA CCT GGA CTG CGT CAT CA-3', Reverse: 5'-CCC GGA TGC TAT ACA TGA CA-3' (mapped to exon 5–6). *EXOC6B* had a detectable transcript, while no transcript was detected with *CYP26B1* primers for exons 4–5 and exons 5–6 in control whole blood, nor transformed lymphoblasts.

Real-time qPCR was performed for *EXOC6B* and its downstream genes *HES1* and *RBPJ* using SYBR® Green PCR Master Mix (Applied Biological Materials Inc). Primer sequences were as follows: *HES1* Forward: 5'-GAG CAC AGA AAG TCA TCA AAG C-3' (mapped to exons 1–3), Reverse: 5'-TTC CAG AAT GTC CGC CTT C-3'; *RBPJ* Forward: 5'-GGG ATA GGA AAT AGT GAC CAA GA-3', reverse: 5'-GTG CTT TCG CTT GTC TGA GT-3' (mapped to exons 7–8).

Quantification of expression level of *EXOC6B, HES1* and *RBPJ* was performed in comparison to actin. All PCR reactions were performed in triplicate, with the mean of $2^{-\Delta\Delta Ct}$ values [34] being used to determine mRNA levels for Subject 1 in comparison to the mean expression level for three controls. Significance was calculated using the Student's t test (VassarState Statistical Computation website). Values were considered statistically significant with a p-value of <0.05.

Protein expression

LCLs were washed three times in PBS and incubated in 1 mL RIPA buffer (Thermo Scientific, USA), supplemented with Halt Phosphatase Inhibitor Cocktail (Thermo Scientific, USA) for 15 minutes on ice. Cell lysates were centrifuged at 10,000 × g for 15 minutes at 4°C, and the supernatants were used for Western blotting. Cell lysate protein concentrations were determined using a DC™ Protein Assay (BioRad laboratories, USA). Western blots containing cell-lysate aliquots (~30 μg) were prepared and immunoblotted using a polyclonal antibody directed against human RBPJ or a monoclonal antibody against human HES1 (Abcam, Cambrige, MA, USA). Three antibodies against human EXOC6B were tested in this study (Sec15B(C-14) (goat polycolonal antibody):sc-34375, Santa Cruz Biotechnology,Inc. Anti-EXOC6B antibody (rabbit polycolonal antibody): ab116383, Abcam Inc. Anti-EXOC6B antibody (mouse polycolonal antibody): H00023233-B01P, Novus Biologicals) but did not provide a clear or visible band at the expected size, amenable to interpretation. To standardize the amounts of protein loaded into each lane, the blots were reprobed with a monoclonal antibody directed against human β-actin (NovusBiologicals, Littleton, USA). The ECL Western Blotting system was used to detect the amount of each antibody bound to antigen and the resultant photographic films were analyzed by UV densitometry (GE Healthcare Life Sciences, Pittsburgh, USA). The absorbance values obtained for HES1 and RBPJ were then normalized relative to the corresponding β-actin absorbance value. The average of HES1 or RBPJ protein expression was obtained from three independent replicates from the Subject and the control samples. P values for two tailed Student test were considered statistically significant if <0.05.

CYP26B1 expression upon ATRA induction

Cells from seven control subjects (aged 2–42 years) and Subject 1 were plated in 6-well plates at a density of 400,000 cells/mL in 3 mL and treated in triplicate with 1 μM *all-trans* retinoic acid (ATRA) or an equal volume of vehicle (ethanol) in triplicate and incubated in the dark for 72 hours. Following treatment, cells were pelleted and resuspended in 1 mL Tri-Reagent Solution (Life Technologies, Grand Island, NY) followed by repeated pipetting to lyse the cells. RNA was then isolated according to manufacturers instructions and RNA

quantity and quality were determined using a NanoDrop 2000c spectrophotometer (Thermo Scientific, Waltham, MA) by measuring the absorbance at 260 nm and 280 nm. Complimentary DNA (cDNA) was generated from 1 μg total RNA by reverse transcription using the TaqMan reverse transcription reagents kit (Life Technologies, Grand Island, NY), as previously described [17,35]. TaqMan Gene Expression Master Mix, PCR primers, and fluorescent probes were obtained from Applied Biosystems (Life Technologies, Grand Island, NY). Probes were labelled with the 5' reporter dye FAM (*CYP26A1* and *CYP26B1*) or VIC (GAPDH). Primer probe pairs used were: *CYP26A1* (Hs00219866_m1), *CYP26B1* (Hs00175627_m1), and the endogenous control *GAPDH* (RefSeq: NM_002046.3). Quantitative real-time PCR was conducted on a StepOnePlus Real-Time PCR instrument (Applied Biosystems, Foster City, CA) as previously described [17,35]. The fold-increase in *CYP26A1* and *CYP26B1* mRNA was calculated using the $\Delta\Delta C_t$ method (fold difference = $2^{-\Delta\Delta Ct}$) by comparing the ATRA-treated cells to the vehicle treated controls. Differences between controls and patient were tested by t-test using GraphPad Prism 5.0. Values were considered statistically significant with a p-value <0.05. Grubb's test was used to determine if any of the subjects met the criteria of an outlier.

The use of tissues was approved by the Committee for Ethical Review of Research involving Human Subjects, University of British Columbia. Written informed consent was obtained from the patients' parents for the publication of this report and any accompanying images.

Results

Clinical description

Subject 1

This male was seen by a geneticist as a neonate because of dysmorphic features, including an asymmetric crying face with a right facial nerve paralysis, a dysplastic right ear, brachycephaly (Figures 1 and 2) and mild contractures of the knees and elbows (which disappeared by five months). The pregnancy was complicated by maternal nausea and vomiting, with only a five pound maternal weight gain. Birth weight was on the 10th centile. A 2 mm secundum atrial septal defect closed in early infancy. His parents are not consanguineous. A paternal great-uncle has intellectual disability while another paternal great-uncle's adult son has

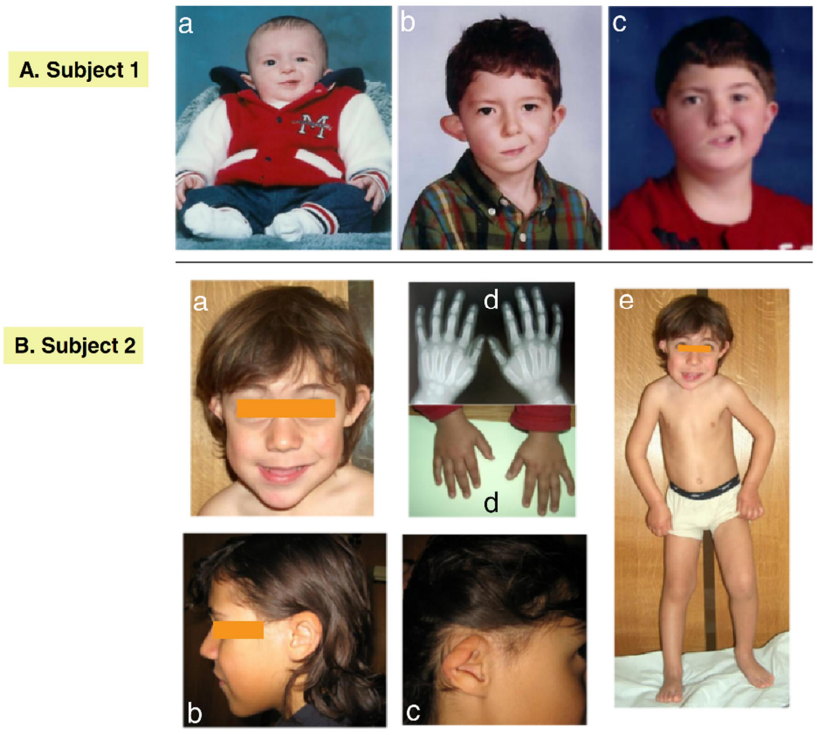


Figure 1 Clinical presentation of Subject 1 and Subject 2. A. Subject 1 at age 2, 7, and 13 **(a-c)**, demonstrating the facial appearance, with asymmetry, right facial nerve palsy, dysplastic and prominent ears. By age 13, the right ear was surgically corrected. **B**. Subject 2 demonstrating **a)** asymmetry of the jaw, **b)** left ear with thick helix, **c)** right ear (post surgery): small, dysplastic, cup-shaped, anteverted, and hypoplastic lobule, **d)** hands with slightly tapering fingers and **e)** normal body proportions except short neck.

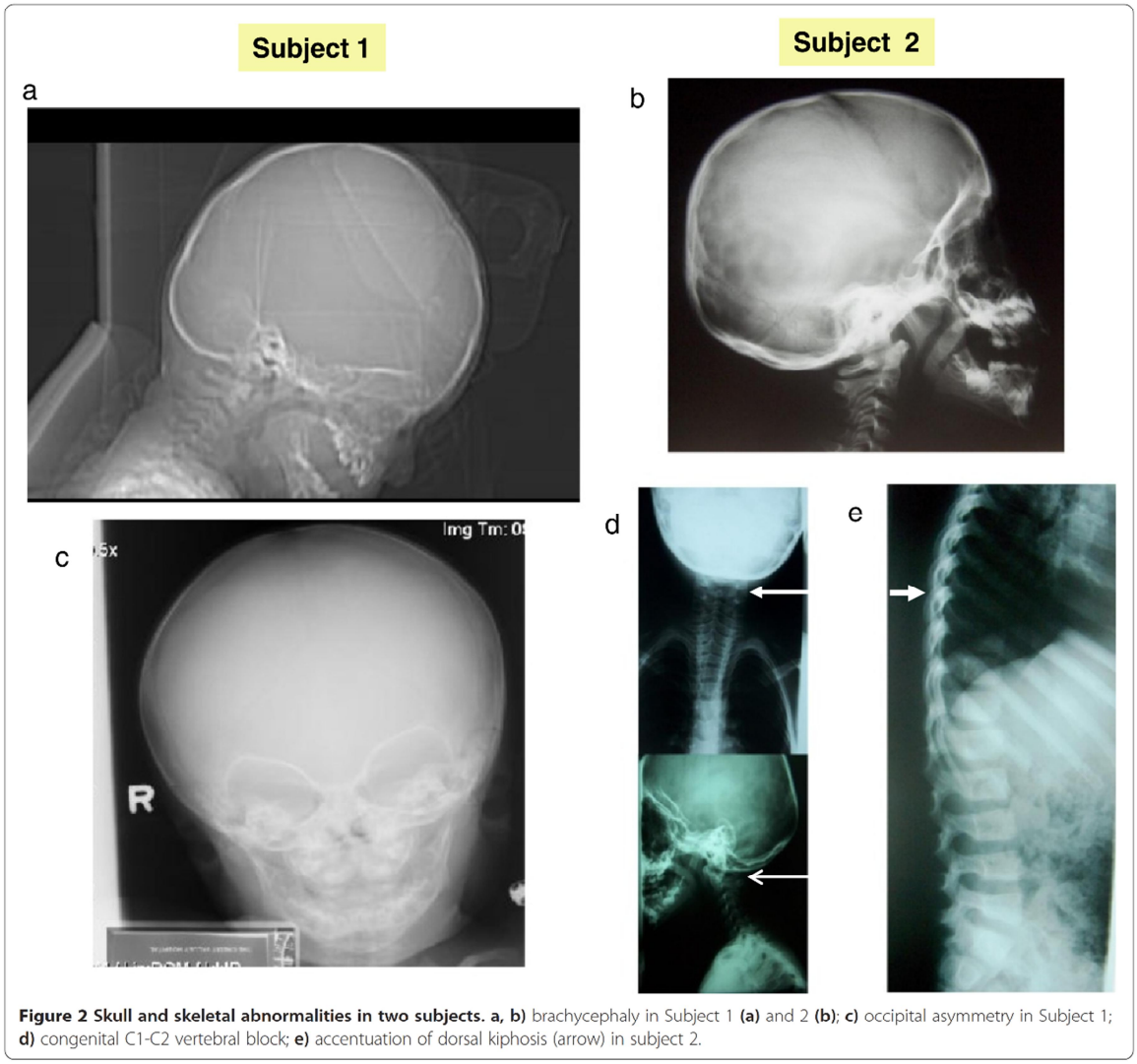


Figure 2 Skull and skeletal abnormalities in two subjects. a, b) brachycephaly in Subject 1 **(a)** and 2 **(b)**; **c)** occipital asymmetry in Subject 1; **d)** congenital C1-C2 vertebral block; **e)** accentuation of dorsal kiphosis (arrow) in subject 2.

"autistic like features" at age 40. The remainder of the family history is unremarkable.

As an infant, he had difficulty transitioning to solid foods, with gagging and vomiting. He walked at 13 months. At 22 months, he had only a few words. By three years, he was not using any words regularly although he would use a word for a few days and then stop. With most activities, he was not able to stay on track, being quite hyperactive and distractible. His clinical diagnosis of autism was confirmed by an assessment using ADOS at age 2 ½ years. At 13 years of age, his academic functioning was somewhat variable, with mathematics being comparable to the skill level of a ten year old, while the level for other subjects was comparable to an eight year old.

At 22 months, height was above the 50th centile, weight was below the 10th centile, and OFC was on the 3rd centile. He had brachycephaly, which had been seen in infancy. The metopic suture, which was open at 7 months, was closed at 22 months. The left ear was low set, while the right was dysplastic, prominent, with a thin helix and an abnormal crural fold. His face appeared to be asymmetric, due to the right facial nerve palsy. Eye examination was normal except for right eyelid paralysis, secondary to the right facial nerve palsy. The remainder of the physical examination was unremarkable.

Chromosome testing was normal, including FISH at 22q11.2. A CT scan of the brain was normal. Skull radiographs at age 7 months showed a shortened anterior-posterior diameter relative to width. There was slight

asymmetry of the appearance of the orbits, with the left side a little larger and slightly more prominent superiorlaterally (Figure 2). No obvious sutural stenosis was evident. At age 14, a C-spine radiograph was normal.

Subject 2

The Subject is a 9 year old boy referred for a genetics consultation because of delayed milestones and dysmorphic faces. His parents are non-consanguineous and he has one healthy sister. His maternal uncle died at the age of 3 months, of unexplained sudden death. No other family history of developmental problems or congenital anomalies was reported. At the age of 3 years, he had two episodes that raised the suspicion of absence seizures, but the EEG was normal. Delayed development was noted, especially in language. He spoke few words with no sentences. Mild motor delay also was described, with unaided walking starting at 16 months.

At four years, his growth parameters were normal (weight: 75th percentile; height: 50-75th percentile and OFC: 50-75th percentile). At five years he had a global developmental quotient of 53.4 on the Griffiths Mental Developmental Scale evaluation. His I.Q. at the age of 7 years based on the Wechsler Intelligence Scale was 73 (borderline mental retardation), with a verbal scale of 66 and performance scale of 89. The proband was dysmorphic with a triangular face, brachycephaly, hypertelorism, upslanting palpebral fissures, thin lips, hypertrophic gums, a pointed chin and short neck (Figure 1). He had abnormal ears (asymmetric, dysplastic and low-set). The right ear was small, cup-shaped, anteverted and the lobule was hypoplastic. The left ear was bigger than the right, with a thick helix. Asymmetry of the jaw was noted, with the left side longer than the right side. His fingers were slightly tapering. Radiographs revealed congenital C1-C2 vertebral fusion, with accentuation of dorsal kyphosis (Figure 2).

Subject 2 displayed stereotypies, aggressive behavior, hyperactivity (including jumping if agitated) and attention deficit. Brain MRI, echocardiogram, abdominal ultrasound, ophthalmology and ENT examinations were normal. Radiographs of upper and lower limbs, pelvis and rib cage did not reveal abnormalities. Chromosome, FISH for the 22q 11.2 region, molecular testing for Fragile X and metabolic studies (plasma aminoacids, urine organic acids, CDT, creatine and guanidinoacetic acid in urine, 7-dehidrocholesterol, lactate, pyruvate and ammonia) were normal.

The clinical description of the two subjects is summarized in Additional file 1: Table S1.

Array CGH

For Subject 1, aCGH revealed a 0.78 Mb de novo deletion at chromosome region 2p13.2-13.3 (72,140,702-72,924,626), containing 2 genes. Subject 2 had a 4 Mb *de novo* deletion detected at chromosome region 2p13.1-p13.3 (chr2:70,748,414-74,840,026), containing 62 genes (Figure 3A and 3B). The region of overlap between both cases is 0.78 Mb, located between the 72,140,702-72,924,626 genomic positions and encompasses the *CYP26B1* and *EXOC6B* genes.

Functional studies of EXOC6B and notch effector genes HES1 and RBPJ in subject 1

Significant reduction of RNA expression for all three genes (*EXOC6*B, *RBPJ* and *HES1*) was detected in lymphoblasts from Subject 1 in comparison to three controls (Figure 4A). Protein expression for both HES1 and RBPJ was also significantly reduced in his lymphoblasts (Figure 4B and 4C), however, EXOC6B protein level could not be assessed due to a number of non-specific bands or a single band of an unexpected size in controls obtained with three commercial EXOC6 antibodies.

We noted that in control lymphoblasts, the baseline level of *HES1* and *RBPJ* expression was significantly higher than *EXOC6*, possibly due to the fact that EBV viral protein EBNA2 induces *RBPJ* expression [36] (Figure 4A). To eliminate the effect of EBV transformation on RBPJ and HES expression, whole blood was used from Subject 1 and the controls. Similarly to the lymphoblasts, RNA expression of *EXOC6* and *RBPJ* was significantly reduced in the Subject's whole blood in comparison to a control. However, *HES1* expression was significantly higher in whole blood of Subject 1 than in the control bloods (Figure 5). Protein from whole blood was not available for assessment of RBPJ and HES1 expression levels.

ATRA induction of *CYP26B1* expression

EBV transformed lymphoblasts from seven controls and Subject 1 were treated with 1 μM ATRA for 3 days. To obtain a quantitative comparison of the ATRA induction of *CYP26B1* in Subject 1 LCLs versus LCLs from controls, the fold difference in *CYP26B1* mRNA induction with ATRA treatment was calculated for Subject 1 and controls in comparison to the vehicle treated cells. ATRA induction of CYP26B1 mRNA in LCLs from one control subject (56.7 ± 3.7 fold) was determined to be an outlier by Grubb's test from the other controls and was excluded from further analysis. The *CYP26B1* mRNA fold-induction from the six remaining control cells was averaged to show the range in induction in control LCLs (Figure 6A). *CYP26B1* mRNA induction with ATRA treatment was significantly less in Subject 1 compared to averaged control cells (1.9 ± 1.5 fold in subject and 12.9 ± 7.3 fold in controls,) (Figure 6A).

The relatively minimal expression of *CYP26B1* in ATRA-treated cells from Subject 1 raised the possibility of compensatory expression of *CYP26A1* to regulate retinoic acid levels. ATRA-treatment resulted in minimal

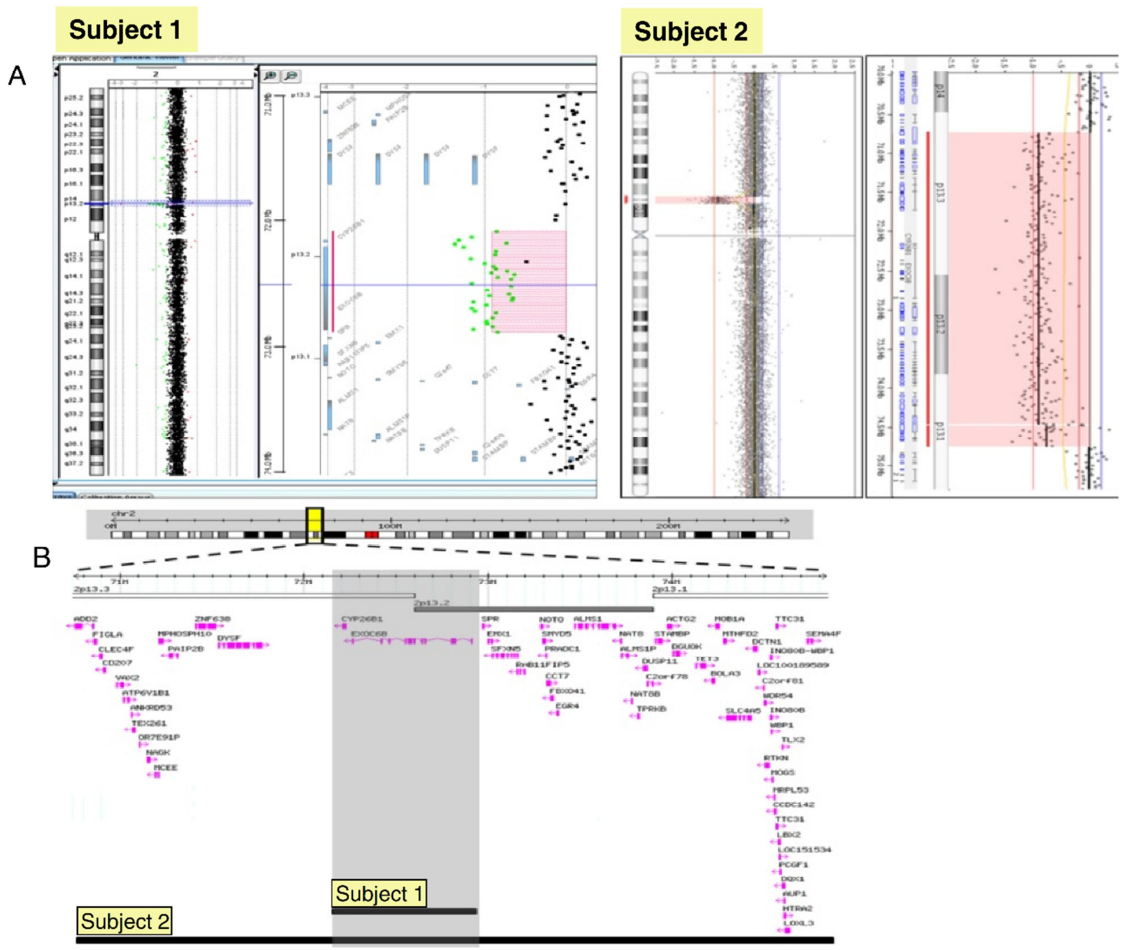


Figure 3 Array CGH result for Subject 1 and 2. A. Whole genome array profiles in Subject 1 and 2 showing microdeletions of 2p13; **B**. Gene content for the microdeletions in Subject 1 and 2.

induction of *CYP26A1* mRNA in all LCLs which is in agreement with the lack of *CYP26A1* induction and overall low to undetectable expression in blood [18]. *CYP26A1* induction was not significantly different in control LCLs (3.4 ± 3.4 fold) compared to Subject 1 LCLs (0.9 ± 0.9 fold) as shown in Figure 6B. Thus, *CYP26A1* was not induced to a greater extent in response to ATRA treatment in cells with a deletion of *CYP26B1*. To exclude the possibility that *CYP26A1* expression was higher in Subject 1 cells compared to control cells prior to ATRA treatment, we compared the ΔCt values of the vehicle-treated cells and found no differences ($P > 0.05$). Hence, it does not appear that *CYP26A1* mRNA expression is upregulated to compensate for reduced *CYP26B1* expression in Subject 1 LCLs. While upregulation of *Cyp26a1* has been observed in *Cyp26b1*$^{-/-}$ mice at specific developmental stages and tissues during mouse organogenesis, overall, the expression patterns and function of *Cyp26a1* and *Cyp26b1* are considered non-overlapping during mouse development [37-39].

Discussion

We report a 2p13.1-p13.3 microdeletion observed in two subjects with clinical effects on the cognitive function (ID and language delay), behaviour (hyperactivity), and development of the craniofacies (facial asymmetry, unusually shaped and asymmetric ears and brachycephaly). Skull and vertebral bone abnormalities included slight asymmetry of the appearance of the orbits and delayed closure of the metopic suture in Subject 1 and congenital C1-C2 vertebral fusion, and accentuation of dorsal kyphosis in Subject 2. Previous cases with cytogenetic deletions/

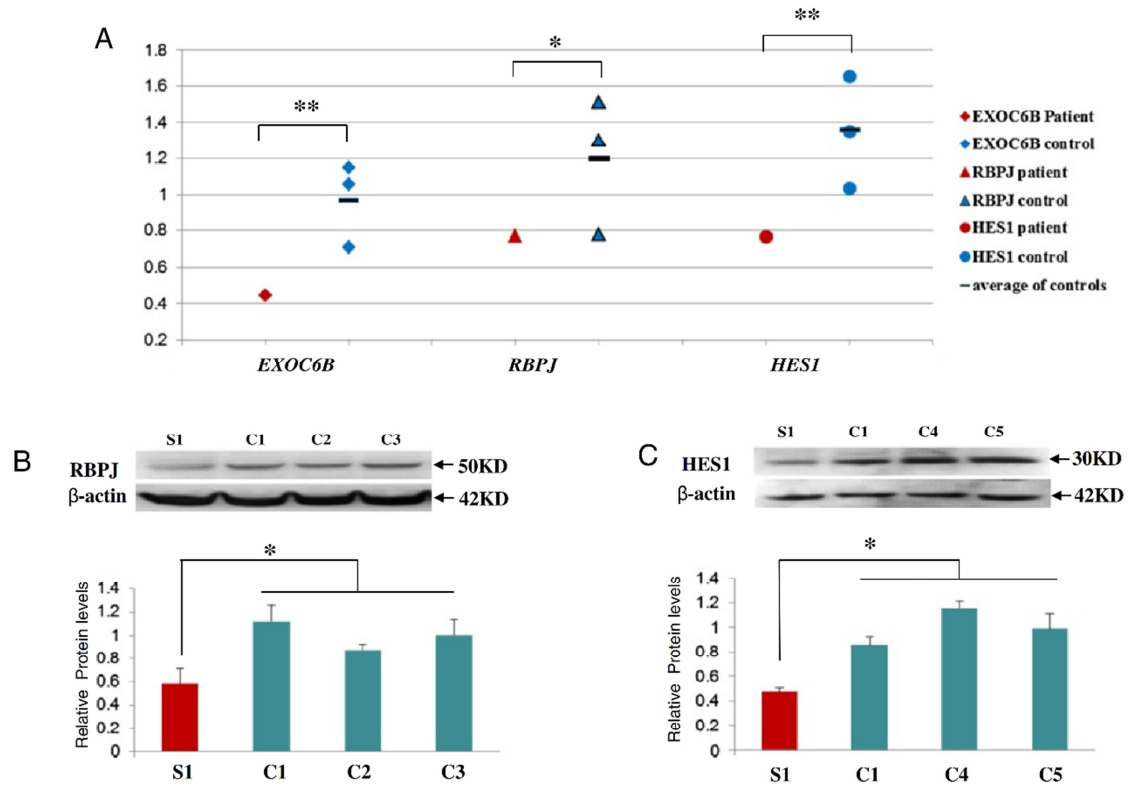


Figure 4 mRNA and protein expression in Subject 1 and control lymphoblast cells. A. Scatter graph of mRNA expression of *EXOC6, RPJB* and *HES1*in lymphoblasts. Values for *EXOC6, RPJB* and *HES1* mRNA expression were normalized to the corresponding ß-actin mRNA levels. The results derived from at least three sets of samples and the mean level of each sample is represented in the graph, horizontal bar is an average value between three controls. The difference between the study sample and the average of the controls has been evaluated by student t-test (*, $p < 0.05$; **, $P < 0.01$). **B**. RBPJ protein expression. **C**. HES1 protein expression. C1-C5 indicate control samples from the lymphoblasts, S1 indicate study sample from Subject1. Cell lysates were analyzed by SDS-PAGE and immunoblotting with membrane probed for RBPJ, HES1 and ß-actin, Values for RBPJ and HES1 expression were normalized to the corresponding ß-actin levels. The results derived from at least three sets of samples and the mean value of each sample is represented in the bar graph. The difference between Subject1 and the average of three controls has been evaluated by student t-test (*, $p < 0.05$).

disruptions of this region all had developmental delay and the majority had head/facial anomalies, ear and skeletal abnormalities [1,2].

The overlapping deleted region of our two cases encompasses two genes, *EXOC6B* and *CYP26B1*, both of which showed altered function in whole blood and/or LCLs from Subject 1. The reduced expression of *EXOC6B* in LCLs and whole blood could be the cause of the observed perturbed Notch signaling (i.e. *HES1* and *RBPJ* expression change), considering the similar effect of *EXOC6B* knockout on Notch signaling in *Drosophila*.[20] The reduction of RNA expression of *EXOC6B* and *RBPJ* was concordant between LCLs and whole blood, and the expression of the *RBPJ* protein also was reduced in Subject 1 lymphoblasts. However, *HES1* had reduced RNA and protein expression in LCLs and increased RNA expression in whole blood. The discrepancy in the pattern of abnormal expression of *HES1* in different cell types could be due to the difference in the Notch signaling pathway in lymphoblasts which contain dedifferentiated B cells vs. whole blood which contains multiple differentiated cell lineages. Cell-specific over or underexpression of *Hes1* has been reported in different regions of the brain in *Rbpj* knockout mice [40].

EXOC6B germline mutations or deletions have yet not been reported in humans. A homozygous mutation in *EXOC6B's* paralogue *EXOC6A* has been reported in mice with hemoglobin-deficit (hbd) due to defective iron transport in the endocytosis cycle [41] while haploinsufficiency of *EXOC6A* due to a 0.3 Mb microdeletion at 10q23.33, was reported in a family with nonsyndromic bi- and unilateral optic nerve aplasia [42]. Interestingly, this microdeletion also included two CYP genes, *CYP26A1*

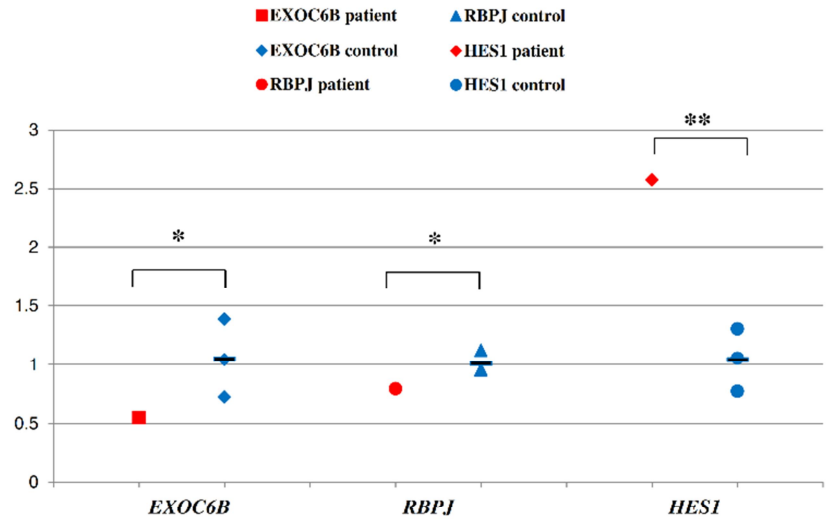


Figure 5 mRNA expression of *EXOC6, RPJB* and *HES1* in whole blood of Subject 1 and controls. Values for *EXOC6, RPJB* and *HES1* mRNA expression in whole blood were normalized to the corresponding ß-actin mRNA levels. The results derived from at least three sets of samples and the mean level of each sample is represented in the scatter graph, horizontal bar is an average value between three controls. The difference between subject 1 and the average of the controls has been evaluated by student t-test (*, $p < 0.05$; **, $P < 0.01$).

and *CYP26C1*. In contrast to very little information on genetic defects of *EXOC6B* and their role in disease, genetic abnormalities of *HES1* and *RBPJ* have been associated with a number of developmental defects in vertebrates. Homozygous *Hes1* and *Rbpj* knockout mice showed severe developmental defects and lethality [43,44], while *Rbpj* heterozygous knockout mice demonstrated learning deficits [29]. In humans, increased expression of *HES1* was reported in Down syndrome [45] and recent exome sequencing studies revealed heterozygous mutations in *RBPJ* and reduced expression of *HES1* in two families with Adams-Oliver syndrome, associated with congenital cutis aplasia, terminal limb abnormalities (asymmetric shortening of the hands and feet in one of the families) and a range of cognitive functioning (from intellectual disability to normal) [46].

The deletion of *CYP26B1* gene in both our Subjects is also likely to contribute to their abnormal phenotype, based on abnormal RA metabolism in Subject 1, as evidenced by significantly attenuated induction of *CYP26B1* expression with ATRA in comparison to controls. To the best of our knowledge, there are no reports on the effect of *CYP26B1* gene haploinsufficiency in humans. Previously, in two other families, two different homozygous mutations of *CYP26B1* have been reported, resulting in lethality, skeletal and craniofacial abnormalities, including fusion of long bones, calvarial bone hypopasia and craniosynostosis [16]. In one

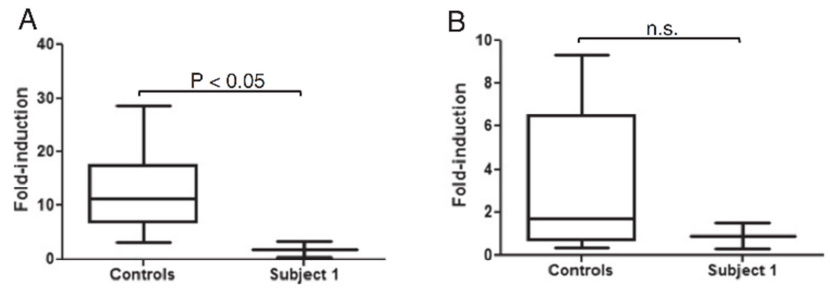


Figure 6 Induction of *CYP26B1* and *CYP26A1*. A. Fold induction of *CYP26B1* mRNA expression. *CYP26B1* was induced significantly more (*$P < 0.05$) in the control cells as compared to the Subject 1 (S1) cells with a chromosomal deletion of *CYP26B1* after treatment with ATRA. **B**. Induction of *CYP26A1* mRNA expression in ATRA-treated cells. There was no significant difference (n.s.; $P > 0.05$) in *CYP26A1* induction between the control cells and Subject 1 cells.

of the families with the hypomorphic mutation, brachycephaly and wide sagittal sutures were noted. The two mutation-bearing constructs had attenuated ability to metabolize ATRA (36% and 86%) [16]. In our Subjects, the decreased RA catabolism and increased sensitivity to RA, as a consequence of *CYP26B1* deletion could explain features similar to those noted by Laue *et al.* [16] (e.g. brachycephaly for both subjects, and for Subject 1, the delayed closure of the metopic suture) and in general compromise the craniofacial, skeletal development and neuronal functioning. With regard to the later phenotype, it is of interest that Subject 1 had an asymmetric crying face as a consequence of a right facial nerve palsy which previously was associated with RA exposure or early embryonic insult [11,47]. The phenotype of the subjects is in agreement with developmental effects of CYP26B1 deletion in experimental animals [37]. CYP26B1 null mice have craniofacial abnormalities, exhibit abnormal ear development and other bone and cartilage deformities [13]. Interestingly, the truncated limbs observed in CYP26B1$^{-/-}$ mice were absent in the patients. Similarly, in the zebrafish CYP26B1 deletion has been shown to result in overall defective craniofacial cartilage development with smaller head, severely decreased number of vagal branchiomotor neurons and defective or absent jaw cartilage [48].

It is intriguing to note the lethal phenotype in two subjects reported by Laue *et al.* [16] due to homozygous mutations of *CYP26B*1 in comparison to the survival and developmental abnormalities in our subjects with hemizygous *CYP26B1* deletion. The presence of one normal copy of the gene in each of our subjects and the efficiencies of the other two remaining RA catabolizing CYP26 genes (*CYP26A1* and *CYP26C1*), possible compensatory changes in other proteins known to regulate retinoic acid, such as RALDH [8,49,50] and environmental influences, such as diet and pharmacological treatment, [6,51] all could have an effect on the phenotypes. Phenotypic variability also was noted for carriers of *CYP26A1, CYP26C1* and *EXOC6A* deletions within one family [42], ranging from normal vision, to uni- and bilateral optic nerve hypoplasia and variable levels of cognitive functioning (normal to impaired).

The combined effect of deletion of both *EXOC6B* and *CYP26B1* on Notch and RA signaling, and consequently the phenotype, also should be considered in our subjects. Interaction of RA and Notch signaling in determination of left/right asymmetry and segmentation has been reported by Echeverri and Oates [52] who noted the requirement of *Rbpj* function for expression of RA catabolizing enzyme *Cyp26a1* which in turn, is needed for left/right symmetric cyclic gene expression. Vermot *et al.* [53] demonstrated that reduced levels of RA were associated with abnormal *Hes1* expression and asymmetry in mouse embryos, while Castella *et al.* [54] showed that addition of RA raises the level of *HES1* protein expression in *in vitro* cell culture.

Subject 2 had additional phenotypic features not noted in Subject 1, which might be explained by the larger size of the 2p13.1-13.3 deletion, interaction of its integral genes and genetic background. For example, seizure-like episodes were not noted in Subject 1, but were present in Subject 2 and a Decipher subject, # 257412, with developmental delay, whose deletion of 2p12-13.3 was 6.8 Mb (70,889,254-77,746,500), and his features included coarse faces, a flat malar region, myopathic hypotonia and prominent ears. The cervical fusion and spine abnormality are unique for Subject 2 and are of interest, considering the report of a subject with Klippel-Fiel anomaly, who had a balanced inversion involving chromosome 2p13 and congenital fusion of the cervical spine, impairment of hearing, psychomotor retardation, speech limitation, short stature, spinal asymmetry and scoliosis [3]. The genes disrupted/deleted by this chromosome rearrangement are unknown, however, it was speculated that *CYP26B1* was involved based on the similar vertebral phenotype observed in zebrafish with *Cyp26b1* mutations [4].

Our report is unique as it provides a new insight into the phenotypic and functional consequences of hemizygous deletion of two genes implicated in Notch and Retinoic acid signaling. It also supports the previous cell line and animal model based observation of exocyst role in Notch signaling. Further studies of exocyst complex function in patient cells would be of interest for understanding of its role in human disease.

Additional file

Additional file 1: Table S1. Clinical features of Subject 1 and 2.

Competing interests

The authors have no competing interests to declare.

Authors' contribution

JW, ERS, CN, NI, FL and PM: Conception, design, analysis and interpretation of data. Drafting the article or revising it critically for important intellectual content. SAF, GS, SL, CB, YQ, BY and SM: Analysis and interpretation of data. Drafting the article or revising it critically for important intellectual content. All authors read and approved the final manuscript.

Authors' information

Patricia Maciel and Evica Rajcan-Separovic are co-corresponding authors.

Acknowledgments

We would like to sincerely acknowledge the families of the presented cases for participation in the genetic studies and for allowing this publication. This work was supported by funding from the Canadian Institutes for Health Research (CIHR) (RT-64217; PI: MESL and MOP 74502; PI: ERS), Autism Speaks (PI: MESL) and Fundação para a Ciência e Tecnologia (PIC/IC/83026/2007. MES and ERS are Career Investigators with the Michael Smith Foundation for Health Research. The experiments were supported in part by a US National institutes of Health grant R01 GM081569 (NI and CN).

This study makes use of data generated by the DECIPHER Consortium. A full list of centres which contributed to the generation of the data is available from http://decipher.sanger.ac.uk and via email from decipher@sanger.ac.uk. Funding for the project was provided by the Wellcome Trust.

Author details
[1]Child and Family Research Institute, Department of Pathology, University of British Columbia, Vancouver, BC, Canada. [2]Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, Portugal. [3]ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal. [4]Center for Medical Genetics Dr. Jacinto Magalhães, National Health Institute Dr Ricardo Jorge, Porto, Portugal. [5]Genetics, Trillium Health Partners, Credit Valley Hospital Site, Mississauga, ON, Canada. [6]Department of Pharmaceutics, School of Pharmacy, University of Washington, Seattle, WA, USA. [7]Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands. [8]Child and family Research Institute, Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada.

Received: 21 February 2013 Accepted: 3 July 2013
Published: 10 July 2013

References

1. Lacbawan FL, White BJ, Anguiano A, Rigdon DT, Ball KD, Bromage GB, Yang X, DiFazio MP, Levin SW: **Rare interstitial deletion (2)(p11.2p13) in a child with pericentric inversion (2)(p11.2q13) of paternal origin.** *Am J Med Genet* 1999, **87:**139–142.
2. Prasher VP, Krishnan VH, Clarke DJ, Maliszewska CT, Corbett JA: **Deletion of chromosome 2 (p11-p13): case report and review.** *J Med Genet* 1993, **30:**604–606.
3. Papagrigorakis MJ, Synodinos PN, Daliouris CP, Metaxotou C: **De novo inv (2)(p12q34) associated with Klippel-Feil anomaly and hypodontia.** *Eur J Pediatr* 2003, **162:**594–597.
4. Laue K, Janicke M, Plaster N, Sonntag C, Hammerschmidt M: **Restriction of retinoic acid activity by Cyp26b1 is required for proper timing and patterning of osteogenesis during zebrafish development.** *Development* 2008, **135:**3775–3787.
5. Borsani G, Piovani G, Zoppi N, Bertini V, Bini R, Notarangelo L, Barlati S: **Cytogenetic and molecular characterization of a de-novo t(2p;7p) translocation involving TNS3 and EXOC6B genes in a boy with a complex syndromic phenotype.** *Eur J Med Genet* 2008, **51:**292–302.
6. Ross AC, Zolfaghari R: **Cytochrome P450s in the regulation of cellular retinoic acid metabolism.** *Annu Rev Nutr* 2011, **31:**65–87.
7. Duester G: **Retinoic acid synthesis and signaling during early organogenesis.** *Cell* 2008, **134:**921–931.
8. Lee LM, Leung CY, Tang WW, Choi HL, Leung YC, McCaffery PJ, Wang CC, Woolf AS, Shum AS: **A paradoxical teratogenic mechanism for retinoic acid.** *Proc Natl Acad Sci USA* 2012, **109:**13668–13673.
9. Rothman KJ, Moore LL, Singer MR, Nguyen US, Mannino S, Milunsky A: **Teratogenicity of high vitamin A intake.** *N Engl J Med* 1995, **333:**1369–1373.
10. Loureiro KD, Kao KK, Jones KL, Alvarado S, Chavez C, Dick L, Felix R, Johnson D, Chambers CD: **Minor malformations characteristic of the retinoic acid embryopathy and other birth outcomes in children of women exposed to topical tretinoin during early pregnancy.** *Am J Med Genet A* 2005, **136:**117–121.
11. Sarici D, Akin MA, Kurtoglu S, Uzum K, Kiraz A: **Asymmetric Crying Face in a Newborn with Isotretinoin Embryopathy.** *Pediatr Dermatol* 2012, **21:**1–2.
12. Wise M, Emmanouil-Nikoloussi EN, Moxham BJ: **Histological examination of major craniofacial abnormalities produced in rat foetuses with a variety of retinoids.** *Eur J Anat* 2007, **11:**17–26.
13. Maclean G, Dolle P, Petkovich M: **Genetic disruption of CYP26B1 severely affects development of neural crest derived head structures, but does not compromise hindbrain patterning.** *Dev Dyn* 2009, **238:**732–745.
14. Pennimpede T, Cameron DA, MacLean GA, Li H, Abu-Abed S, Petkovich M: **The role of CYP26 enzymes in defining appropriate retinoic acid exposure during embryogenesis.** *Birth defects research* 2010, **88:**883–894.
15. Dranse HJ, Sampaio AV, Petkovich M, Underhill TM: **Genetic deletion of Cyp26b1 negatively impacts limb skeletogenesis by inhibiting chondrogenesis.** *J Cell Sci* 2011, **124:**2723–2734.
16. Laue K, Pogoda HM, Daniel PB, van Haeringen A, Alanay Y, von Ameln S, Rachwalski M, Morgan T, Gray MJ, Breuning MH, *et al*: **Craniosynostosis and multiple skeletal anomalies in humans and zebrafish result from a defect in the localized degradation of retinoic acid.** *Am J Hum Genet* 2011, **89:**595–606.
17. Topletz AR, Thatcher JE, Zelter A, Lutz JD, Tay S, Nelson WL, Isoherranen N: **Comparison of the function and expression of CYP26A1 and CYP26B1, the two retinoic acid hydroxylases.** *Biochem Pharmacol* 2012, **83:**149–163.
18. Takeuchi H, Yokota A, Ohoka Y, Iwata M: **Cyp26b1 regulates retinoic acid-dependent signals in T cells and its expression is inhibited by transforming growth factor-beta.** *PLoS One* 2011, **6:**e16089.
19. White JA, Ramshaw H, Taimi M, Stangle W, Zhang A, Everingham S, Creighton S, Tam SP, Jones G, Petkovich M: **Identification of the human cytochrome P450, P450RAI-2, which is predominantly expressed in the adult cerebellum and is responsible for all-trans-retinoic acid metabolism.** *Proc Natl Acad Sci USA* 2000, **97:**6403–6408.
20. Wu L, Ross AC: **Acidic retinoids synergize with vitamin A to enhance retinol uptake and STRA6, LRAT, and CYP26B1 expression in neonatal lung.** *J Lipid Res* 2010, **51:**378–387.
21. Ross AC, Cifelli CJ, Zolfaghari R, Li NQ: **Multiple cytochrome P-450 genes are concomitantly regulated by vitamin A under steady-state conditions and by retinoic acid during hepatic first-pass metabolism.** *Physiol Genomics* 2011, **43:**57–67.
22. Zolfaghari R, Cifelli CJ, Lieu SO, Chen Q, Li NQ, Ross AC: **Lipopolysaccharide opposes the induction of CYP26A1 and CYP26B1 gene expression by retinoic acid in the rat liver in vivo.** *Am J Physiol* 2007, **292:**G1029–G1036.
23. Heider MR, Munson M: **Exorcising the exocyst complex.** *Traffic (Copenhagen, Denmark)* 2012, **13:**898–907.
24. Jafar-Nejad H, Andrews HK, Acar M, Bayat V, Wirtz-Peitz F, Mehta SQ, Knoblich JA, Bellen HJ: **Sec15, a component of the exocyst, promotes notch signaling during the asymmetric division of Drosophila sensory organ precursors.** *Dev Cell* 2005, **9:**351–363.
25. Mehta SQ, Hiesinger PR, Beronja S, Zhai RG, Schulze KL, Verstreken P, Cao Y, Zhou Y, Tepass U, Crair MC, Bellen HJ: **Mutations in Drosophila sec15 reveal a function in neuronal targeting for a subset of exocyst components.** *Neuron* 2005, **46:**219–232.
26. Guichard A, McGillivray SM, Cruz-Moreno B, van Sorge NM, Nizet V, Bier E: **Anthrax toxins cooperatively inhibit endocytic recycling by the Rab11/Sec15 exocyst.** *Nature* 2010, **467:**854–858.
27. Borggrefe T, Oswald F: **The Notch signaling pathway: transcriptional regulation at Notch target genes.** *Cell Mol Life Sci* 2009, **66:**1631–1646.
28. Kato Y: **The multiple roles of Notch signaling during left-right patterning.** *Cell Mol Life Sci* 2011, **68:**2555–2567.
29. Costa RM, Honjo T, Silva AJ: **Learning and memory deficits in Notch mutant mice.** *Curr Biol* 2003, **13:**1348–1354.
30. Penton AL, Leonard LD, Spinner NB: **Notch signaling in human development and disease.** *Semin Cell Dev Biol* 2012, **23:**450–457.
31. Fan YS, Jayakar P, Zhu H, Barbouth D, Sacharow S, Morales A, Carver V, Benke P, Mundy P, Elsas LJ: **Detection of pathogenic gene copy number variations in patients with mental retardation by genomewide oligonucleotide array comparative genomic hybridization.** *Hum Mutat* 2007, **28:**1124–1132.
32. Krijgsman O, Israeli D, van Essen HF, Eijk PP, Berens ML, Mellink CH, Nieuwint AW, Weiss MM, Steenbergen RD, Meijer GA, Ylstra B: **Detection limits of DNA copy number alterations in heterogeneous cell populations.** *Cellular oncology (Dordrecht)* 2013, **36:**27–36.
33. Buffart TE, Israeli D, Tijssen M, Vosse SJ, Mrsic A, Meijer GA, Ylstra B: **Across array comparative genomic hybridization: a strategy to reduce reference channel hybridizations.** *Genes Chromosomes Cancer* 2008, **47:**994–1004.
34. Schmittgen TD, Livak KJ: **Analyzing real-time PCR data by the comparative C(T) method.** *Nat Protoc* 2008, **3:**1101–1108.
35. Tay S, Dickmann L, Dixit V, Isoherranen N: **A comparison of the roles of peroxisome proliferator-activated receptor and retinoic acid receptor on CYP26 regulation.** *Mol Pharmacol* 2010, **77:**218–227.
36. Zimber-Strobl U, Strobl LJ: **EBNA2 and Notch signalling in Epstein-Barr virus mediated immortalization of B lymphocytes.** *Semin Cancer Biol* 2001, **11:**423–434.
37. White RJ, Schilling TF: **How degrading: Cyp26s in hindbrain development.** *Dev Dyn* 2008, **237:**2775–2790.
38. Abu-Abed S, MacLean G, Fraulob V, Chambon P, Petkovich M, Dolle P: **Differential expression of the retinoic acid-metabolizing enzymes CYP26A1 and CYP26B1 during murine organogenesis.** *Mech Dev* 2002, **110:**173–177.
39. Yashiro K, Zhao X, Uehara M, Yamashita K, Nishijima M, Nishino J, Saijoh Y, Sakai Y, Hamada H: **Regulation of retinoic acid distribution is required for**

proximodistal patterning and outgrowth of the developing mouse limb. *Dev Cell* 2004, **6**:411–422.
40. Shi M, Hu ZL, Zheng MH, Song NN, Huang Y, Zhao G, Han H, Ding YQ: **Notch-Rbpj signaling is required for the development of noradrenergic neurons in the mouse locus coeruleus.** *J Cell Sci* 2012, **125**:4320–4332.
41. Lim JE, Jin O, Bennett C, Morgan K, Wang F, Trenor CC 3rd, Fleming MD, Andrews NC: **A mutation in Sec15l1 causes anemia in hemoglobin deficit (hbd) mice.** *Nat Genet* 2005, **37**:1270–1273.
42. Meire F, Delpierre I, Brachet C, Roulez F, Van Nechel C, Depasse F, Christophe C, Menten B, De Baere E: **Nonsyndromic bilateral and unilateral optic nerve aplasia: first familial occurrence and potential implication of CYP26A1 and CYP26C1 genes.** *Mol Vis* 2011, **17**:2072–2079.
43. Oka C, Nakano T, Wakeham A, de la Pompa JL, Mori C, Sakai T, Okazaki S, Kawaichi M, Shiota K, Mak TW, Honjo T: **Disruption of the mouse RBP-J kappa gene results in early embryonic death.** *Development* 1995, **121**:3291–3301.
44. Ishibashi M, Ang SL, Shiota K, Nakanishi S, Kageyama R, Guillemot F: **Targeted disruption of mammalian hairy and Enhancer of split homolog-1 (HES-1) leads to up-regulation of neural helix-loop-helix factors, premature neurogenesis, and severe neural tube defects.** *Genes Dev* 1995, **9**:3136–3148.
45. Fischer DF, van Dijk R, Sluijs JA, Nair SM, Racchi M, Levelt CN, van Leeuwen FW, Hol EM: **Activation of the Notch pathway in Down syndrome: cross-talk of Notch and APP.** *FASEB J* 2005, **19**:1451–1458.
46. Hassed SJ, Wiley GB, Wang S, Lee JY, Li S, Xu W, Zhao ZJ, Mulvihill JJ, Robertson J, Warner J, Gaffney PM: **RBPJ mutations identified in two families affected by Adams-Oliver syndrome.** *Am J Hum Genet* 2012, **91**:391–395.
47. Sjogreen L, Kiliaridis S: **Facial palsy in individuals with thalidomide embryopathy: frequency and characteristics.** *J Laryngol Otol* 2012, **126**:902–906.
48. Reijntjes S, Rodaway A, Maden M: **The retinoic acid metabolising gene, CYP26B1, patterns the cartilaginous cranial neural crest in zebrafish.** *Int J Dev Biol* 2007, **51**:351–360.
49. Abu-Abed S, Dolle P, Metzger D, Wood C, MacLean G, Chambon P, Petkovich M: **Developing with lethal RA levels: genetic ablation of Rarg can restore the viability of mice lacking Cyp26a1.** *Development* 2003, **130**:1449–1459.
50. Niederreither K, Abu-Abed S, Schuhbaur B, Petkovich M, Chambon P, Dolle P: **Genetic evidence that oxidative derivatives of retinoic acid are not involved in retinoid signaling during mouse development.** *Nat Genet* 2002, **31**:84–88.
51. Ross AC: **Retinoid production and catabolism: role of diet in regulating retinol esterification and retinoic Acid oxidation.** *J Nutr* 2003, **133**:291S–296S.
52. Echeverri K, Oates AC: **Coordination of symmetric cyclic gene expression during somitogenesis by Suppressor of Hairless involves regulation of retinoic acid catabolism.** *Dev Biol* 2007, **301**:388–403.
53. Vermot J, Gallego Llamas J, Fraulob V, Niederreither K, Chambon P, Dolle P: **Retinoic acid controls the bilateral symmetry of somite formation in the mouse embryo.** *Science (New York, NY)* 2005, **308**:563–566.
54. Castella P, Sawai S, Nakao K, Wagner JA, Caudy M: **HES-1 repression of differentiation and proliferation in PC12 cells: role for the helix 3-helix 4 domain in transcription repression.** *Mol Cell Biol* 2000, **20**:6170–6183.

doi:10.1186/1750-1172-8-100
Cite this article as: Wen *et al.*: Phenotypic and functional consequences of haploinsufficiency of genes from exocyst and retinoic acid pathway due to a recurrent microdeletion of 2p13.2. *Orphanet Journal of Rare Diseases* 2013 **8**:100.

Supplementary Table S1 - Clinical features of Subject 1 and 2

Feature		Subject 1	Subject 2
Genomic data	**Start-End (bp)(hg18)**	2p13.2-13.3 (chr 2: 72,140,702-72,924,626)	2p13.1-p13.3 (chr 2: 70,748,414- 74,840,026)
	Size (Mb)	0,78	4,1
	Type	Deletion	Deletion
	Origin	de novo	de novo
	Gene contents	CYP26B1, EXOC6B	62 genes
Sex		Male	Male
Age at clinical examination		infant to 14-year-old	9-year-old
Family history of DD		Yes (paternal side)	No
Developmental delay	**Delay in language**	Yes	Yes
	Delay in motor function	Yes	Yes
Dysmorphisms	**head shape/size**	Brachycephaly, microcephaly	Brachycephaly
	Ears	The left ear was low set, while the right was dysplastic, prominent, with a thin helix and an abnormal crural fold.	Abnormal ears (asymmetric, dysplastic and low-set). The right ear was small, cup-shaped, anteverted and the lobule was hypoplastic. The left ear was bigger than the right, with a thick helix.
	Face	Unilateral facial paralysis, slightly broad nasal tip. His face appeared to be asymmetric, due to the right facial nerve palsy. Eye examination was normal except for right eyelid paralysis, secondary to the right facial nerve palsy	Asymmetry of the jaw was noted, with the left side longer than the right side. Triangular face, hypertelorism, up-slanting palpebral fissures, thin lips, hypertrophic gums, a pointed chin.
	Neck	Normal	Short
Behaviour		diagnosed as having an autistic spectrum disorder at age 2 1/2 ; quite hyperactive and distractible	stereotypies, aggressive behavior and hyperactivity and attention deficit
Extremities		at birth mild contractures of the knees and elbows (which disappeared by five months	slightly tapering fingers
Seizure		No	Yes, absence seizures, but EEG normal
Clinical test	**Karyotype**	Normal	Normal
	X-ray	a shortened AP diameter of the skull relative to the width. There was slight asymmetry of the appearance of the orbits on the left side looked a little larger and slightly more prominent superolaterally.	brachycephalic skull with hypoplasia of facial bones and mild shortening of lower jaw . Congenital C1-C2 vertebral fusion, with accentuation of dorsal kyphosis

The contribution of 7q33 interstitial deletions for intellectual disability

The contribution of 7q33 interstitial deletions for intellectual disability

Fátima Lopes, Fátima Torres, Sally Ann Lynch, Arminda Jorge, Susana Sousa, João Silva, Paula Rendeiro, Purificação Tavares, Ana Maria Fortuna, Patricia Maciel

Abstract

Copy number variations at 7q33 cytoband are very rarely described in the literature and almost all of the cases comprise large deletions affecting more than just the q33 segment. We report seven patients (two families with two siblings and their affected mother and one unrelated patient) with neurodevelopmental delay associated with copy number variations in 7q33 alone. All the patients presented mild to moderate intellectual disability, dysmorphic features and a behavioral phenotype. One family presents a small duplication affecting only *CALD1* and *AGBL3* gene while the other four patients carry two larger deletions encompassing *EXOC4*, *CALD1, AGBL3* and *CNOT4*. This collection of patients adds to the previous reports of 7q33 copy number variations and suggests that the *CALD1* gene contributes for intellectual disability. This work helps to refine the phenotype and narrow the minimal critical region involved in 7q33 CNVs. Comparison with similar cases, expression and functional studies should help us clarify the relevance of the deleted genes for intellectual disability.

Introduction

Interstitial deletions in 7q are a rare event and, consequently, poorly characterized. Specifically, there are only 10 reports in the literature of interstitial CNVs involving 7q33. Two cases are deletions (7.6 Mb and 7 Mb) derived from chromosomal translocations (Malmgren et al., 2005; Yue et al., 2005); one case is a small deletion (100Kb) affecting only two genes (Mitchell *et al.*, 2012); seven cases show large deletions ranging from citoband 7q32 until 7q35 (Nielsen *et al.*, 1979; Stallard and Juberg, 1981; Verma *et al.*, 1992; Rossi *et al.*, 2008; Petrin *et al.*, 2010; Ponnala and Dalal, 2011; Dilzell *et al.*, 2015). A deletion affecting 7q33 only was reported as an abstract but gave no reference to the deletion size and genes affected (Ponnala and Dalal, 2011). In addition, a 100 Kb deletion affecting only the *SLC35B4* gene was reported in a patient described to have PHACE syndrome (Mitchell *et al.*, 2012). The two most recent reports in the literature regarding interstitial 7q deletions describe genomic losses in a patient with ID, language delay and microcephaly (Kale and Philip, 2016) and in a patient with ID and dysmorphisms (Dilzell *et al.*, 2015). Not surprisingly, given the variable sizes of the deletion in all the reported cases, there is a widely variable phenotypic presentation in these, which is most likely due to the large number of genes involved in these variants. A summary of these reports is present in table I.

Considering solely the 7q33 citoband it is possible to identify several interesting genes that may account for the intellectual disability/developmental delay (ID/DD) phenotype are present there. Among the most promising ones are *EXOC4* (exocyst complex component 4), which encodes component of the exocyst complex involved in the vesicle docking and fusion in brain (Hsu *et al.*, 1996), *CNOT4* (CCR4-NOT transcription complex, subunit 4), encoding a subunit of the CCR4/NOT complex, known to play a role in transcription regulation (Mersman *et al.*, 2009) and *CALD1* (Caldesmon 1) wich encodes a caldesmon protein that works in the stabilization of actin filaments playing a role in axon extension (Morita *et al.*, 2012).

In this report we describe the clinical and genetic findings of seven patients with 7q33 copy number variations (CNVs) and extend the phenotypic spectrum of 7q33 interstitial CNVs. We also propose *CALD1* major contribution for the ID phenotype of these patients.

Table XI - Summary of the reports with CNVs affecting the 7q33 citoband.

Publication	Affected cases (n)	CNV	Description	Size	Genes affected	Phenotype	Inheritance
Malmgren at al., 2015	3	Trans + del	ins(6;7)(p25;q33q34) Der(7) carriers: 7q33-q34 deletion	Del 7.4-7.6 Mb	Up to 68	DD/ID (variable degree), growth retardation, recurrent infections, facial dysmorphisms (long philtrum, thin upper lip, bulbous nose, large mouth, hypertelorism, dysmorphic ears)	Inherited (familial translocation leading to del/dup)
Yue et al., 2005	1	Trans + del	t(7;10)(q33;q23) Der(7) carriers: 7q34-q35 deletion	Del 7Mb	Up to 31; PTEN-EXOC4 gene fusion	DD, macrocephaly, hypotonia, scoliosis, feeding problems, recurrent infections, speech delay, eyes dysmorphisms	*De novo*
Nielsen et al., 1979	3	Trans + del	der(7)ins(13;7) (q32;q32q34) Der(7): 7q32-q34 deletion	ND	ND	ID, growth retardation, hypertelorism, facial dysmorphisms (bulbous nose, large mouth, large ears)	Inherited (familial translocation leading to del/dup)
Ponnala and Dalal, 2011	1	Trans + del	t(7;14)(q33;q32.3) Der (7): 7q33-qter deletion	ND	ND	DD, absent speech, microcephaly, facial dysmorphisms (prominent eyes, arched eyebrows, malformed ears, bulbous nose)	Not maternal
Stallard and Juberg, 1981	1	Del	7q31-q34 deletion	ND	ND	ID, growth retardation, facial dysmorphisms (long philtrum, thin upper lip, bulbous nose, dysmorphic ears)	*De novo*
Verma et al., 1992	1	Del	7q33-q35 deletion	ND	ND	ID, growth retardation, motor retardation, poor eye contact, recurrent infections, conductive deafness, cleft palate	*De novo*
Rossi et al., 2008	1	Del	7q33-q35 deletion	12 Mb	80	DD, autism, primary amenorrhea, neonatal seizures, sleep difficulties, poor language, truncal obesity, facial dysmorphisms (sunken eyes, hypertelorism, bulbous nose, long philtrum, large mouth)	*De novo*
Petrin et al., 2010	1	Del	7q33-q35 deletion	10 Mb	ND	DD, language delay, mild cerebellar and cerebral atrophy. Language: severe fluency disorder characterized by stuttering and cluttering.	*De novo*
Mitchell et al., 2012	1	Del	7q33 deletion	100 Kb	SLC35B4	PHACE syndrome; brain anomalies (dysplastic right superior vermis, absent inferior vermis, hypoplastic right dural venous sinus, proliferating hemangiomas, aberrant circle of Willis), necrotizing enterocolitis (surgery required), non-viable small intestine, died at 2 months	ND
Dilzell et al., 2015	1	Del	7q33-q35 deletion	9.92 Mb	64	ID, recurrent infections, obesity, self-injury behavior, facial dysmorphisms (small ears, large mouth, smooth philtrum, thin upper lip, hypertelorism, bulbous nose, short neck)	Not maternal

Methods

Molecular karyotyping

Genomic DNA was extracted from peripheral blood using Citogene® DNA isolation kit (Citomed, Portugal) for patients 1, 2 and 3; QIAsymphony SP apparatus (QIAGEN GmbH, Germany) for patient 5, 6 and 7. The aCGH analysis was performed using: a CGH Agilent 180K custom array design accessible through the gene expression omnibus GEO accession number GL15397 for patients 1, 2 and 3 (according to the previously published protocol and the across-array methodology [Krijgsman et al., 2013; Buffart et al., 2008]; Agilent 44K oligo (median probe spacing 40 000 bp) for patient 4; Affymetrix CytoScan 750K platform (750.000 markers distributed throughout the genome, with a medium resolution of 8-20Kb) for patients 5 and 7. aCGH data was analyzed using Nexus Copy Number 5.0 software with FASST Segmentation algorithm for patients 1, 2 and 3; DNA Analytics v4.0.76 for patient 4; Analysis Suite (ChAS 3.0) software for patients 5 and 7.

Quantitative PCR confirmations

Primers for qPCR were designed using Primer3Plus software (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and taking into account standard recommendations for qPCR primer development (Jovanovic *et al.*, 2003). A set of primers was designed for exon 10 of the *CNOT4* gene (NM_001008225) and for exon 4 of the *CALD1* gene (NM_033138). The reference genes used were *SDC4* (NM_002999) and *ZNF80* (NM_007136) localized in the 20q12-q13 and 3p12 regions, respectively (primers designed for this purpose are listed in supplementary table I). qPCR reactions were carried out in a 7500-FAST Real Time PCR machine (Applied Biosystems) using Power SYBR Green® (Applied Biosystems). The specificity of each of the reactions was verified by the generation of a melting curve for each of the amplified fragments. The primer efficiency was calculated by the generation of a standard curve fitting the accepted normal efficiency percentage. Quantification was performed as described elsewhere (Hoebeeck *et al.*, 2005). Ct values obtained for each test were analyzed in DataAssist ™ software (Applied Biosystems, Foster City, CA, USA).

Results

Clinical description

Patients 1, 2 and 3

The proband of the first family (patient 1) is a male who was evaluated at 12 years of age for psychomotor delay, ID and dysmorphic features. Parents are non-consanguineous and the delivery was uncomplicated, with normal growth parameters. At the time of the first consultation he had short stature, weight was in the 25th centile, and OFC was in the 75th centile. The patient is currently 24 years old. Evaluation with the Wechsler Intelligence Scale for Children (third edition) was performed at childhood and showed a full scale IQ of 42, associated with behavioral changes such as aggressiveness, hyperactivity and disinhibition. He is dysmorphic with bulbous and snub nose (with concave root of the nose), downslanting palpebral fissures, epicanthic folds, deep set eyes, thin upper lip, poor dental implantation and narrow cleft palate, dysplastic and posteriorly rotated ears, and prognatism (figure 1A). Additionally, he also has bushy eyebrows, spiky hair with a frontal cowlick and two whirlwind at forehead. The hands present light membranous syndactyly and feet with brachydactyly, sandal gap and fetal pads. Brain MRI detected a peri-vascular space enlargement while echocardiogram and abdominal ultrasound retrieved no abnormalities.

Patient 2 (patient 1´s sister) was observed for the first time at 19 years of age. Pregnancy and delivery were uncomplicated with apparently normal development. At the time of the clinical evaluation she presented short stature, weight was in the 95th centile, and OFC was in the 75th centile. She presented several dysmorphic features, similar to the brother's: snub nose with a concave root, bushy eyebrows, spiky hair with a frontal cowlick and two whirlwind at forehead, deep set eyes, epicanthic folds, thin upper lip and poor dental implantation) (figure 1B). She also has a short neck, narrow palate and small dysplastic ears, posteriorly rotated. Abnormalities of the hands and feet included light membranous syndactyly and brachydactyly, respectively. CT scanning, echocardiogram and abdominal ultrasound showed no abnormalities. Evaluation with the Wechsler Intelligence Scale for Children (third edition) showed a full scale IQ of 62. Currently she is 29 years old. Concerning behavior, she presents aggressiveness (similar to her brother) and disinhibition.

Patient 3 is the mother of patient 1 and 2. She has some clinical features similar to the daughter, such as facial dysmorphic features (milder) and brachydactyly (figure 1C). She has mild ID,

although no formal neuropsychological evaluation was performed; she didn´t complete the 4th grade of school but she has the ability to do household chores.

Patient 4

Patient 4 was born at term to unrelated parents. He was noted to be dysmorphic at birth and was admitted to the hospital because of respiratory grunting. He had feeding problems early on. At the four months of age a right inguinal hernia was detected. He was noted to have a wide open anterior fontenelle at eight months. Otitis media developed and a congenital meatal stenosis required meatoplasty at age four years. An evaluation at 10 years old revealed that he weighed 39.75kg (centile 75), heighted 139.8 cm (centile 50) and had an OFC of 57.2 cm (all within normal parameters) (figure 1D). Currently he has hypertelorism, and is short sighted. Behavioral issues were noticed at four years of age and he was referred to Psychiatry. His attention span was poor. He had aggressive outbursts, unpredictable behavior and used bad language. He also presented a low frustration threshold, was impulsive and with oppositional behavior. Currently he has poor peer relationships (has no friends), he still has odd habits regarding feeding (concerns about bacteria on food) and is preoccupied with germs, death, bugs and smells. He had a diagnosis of attention deficit and hyperactive disorder (ADHD) and developmental dyspraxia at age 11 years. He also has a tendency to be disinhibited.

Patients 5, 6 and 7

The proband of this family (patient 5) was evaluated at 11 years of age. Parents are non-consanguineous and delivery was uncomplicated, with normal somatometric parameters (at birth and now). He currently presents moderate ID (IQ=54), associated with behavioral alterations (opposition). He doesn´t have significant facial dysmorphisms besides strabismus.

Patient 6 (patient 5´s brother) is a 8 years old boy with mild ID (IQ level not available) and aggressive behavior. He presents normal weight, height and OFC (at birth and currently) and doesn't have significant facial dysmorphisms. He is short sighted (myopia).

Their parents were described as having learning difficulties at school. The mother (here referred as patient 7) has a documented ID (IQ level not available) and a psychiatric disorder. Although the father was not formally evaluated in the consultation by the responsible physician, he is described as not healthy. In fact, due to their health conditions, the siblings currently live in an institution, since the parents' do not have the ability to take care of them. The facial appearance of patients 5 and 6 is presented in figure 1E and 1F).

A clinical comparison between the cases is presented in table II.

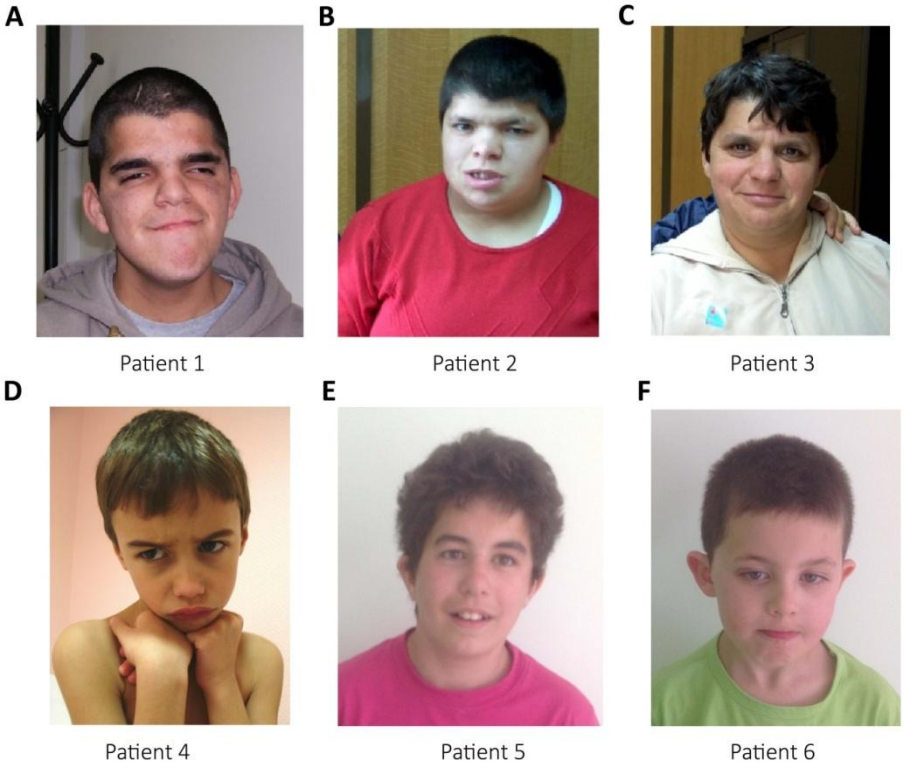


Figure 37 – Facial features of the patients.

Molecular findings

aCGH in patient 1 and 3 revealed a maternally inherited 2.08 Mb deletion at chromosome region 7q33 (chr7:133,176,651-135,252,871, hg19) containing 15 genes (according to the Decipher database). A qPCR assay for the *CNOT4* gene was designed and used for validation and determination of the copy number of the region in the sister and both parents, confirming the presence of only one copy of the segment in the patient, sister and mother. The father presented two copies for the analyzed segment.

Table XII – Clinical summary of the patients

Clinical feature	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7
Gender	♂	♀	♀	♂	♂	♂	♀
Consanguinity	No	No	No	No	No	No	No
Age at presentation/evaluation	12y	19y	Adult	Neonate	11y	8y	Adult
ID	Moderate	Mild	Mild	Mild	Moderate	Mild	Mild
Stature	Short stature	Short stature	ND	Short stature			ND
EEG/Seizures	Seizures (just 1 episode)	No	No	No	No	No	No
Cerebral MRI	Enlargement of peri-vascular spaces	Normal CT scanning	NP	NP	NP	NP	NP
Hypotonia	Yes	No	ND	Yes	No	No	ND
Behaviour phenotype	Aggressiveness, hyperactivity and disinhibition	Aggressiveness, Disinhibition	ND	OCD; Emotional lability; Aggressiveness; Low frustration threshold; Impulsive; Disinhibition	Opposition behaviour	Aggressiveness	ND
Dysmorphisms							
Eyes/ Ophtalmological examination	Strabismus; Epicanthus; Sunken eyes	Epicanthus; Sunken eyes	ND	Downslanting palpebral; Hypertelorism; Normal vision	Strabismus	Myopia	ND

(Cont.)

Nose, Mouth and teeth	Bulbous nose; Thin upper lip; Open mouth; Poor and crowded dental implantation; High and thin palate	Bulbous nose; Thin upper lip; Open mouth; Poor and crowded dental implantation;	Bulbous nose; Thin upper lip	Wide mouth	NA	NA	NA
Forehead, Chin and neck	Whirlwind on the forehead; Prognathism	Whirlwind on the forehead; Short neck	NA	Prominent forehead	NA	NA	NA
Ears/ Audition	Small; Dysplastic	Small; Unilateral hypoacusia	NA	Otitis media in infancy	NA	NA	NA
Hands and feet	Hands: light membranous syndactyly; Feet: sandal gap and fetal pads	Hands: light membranous syndactyly; Feet: sandal gap and fetal pads	ND	Small feet; Arthrogryposis	NA	NA	NA
Abdomen and genitalia	Inguinal hernia	ND	ND	Right inguinal hernia; Chordae of penis	NA	NA	ND
Family history	Family history of ID (maternal uncle with ID, dysmorphisms and epilepsy; second grade cousin (paternal) with ID	Family history of ID (maternal uncle with ID, dysmorphisms and epilepsy; second grade cousin (paternal) with ID	Family history of ID (brother with ID, dysmorphisms and epilepsy)	None	Family history of ID (brother and mother have ID)	Family history of ID (brother and mother have ID)	ND

NA: not available; ND: not described

Patient 4 was found to carry a *de novo* 3.04 Mb deletion at chromosome region 7q33 (chr7:132,766,730-135,802,894, hg19) containing 21 genes (according to the Decipher database).

Patient 5 presented a 216 kb maternal duplication at 7q33 region (chr7:134,598,205-134,807,358, hg19) containing 3 genes (*CALD1*, *AGBL3* and *C7orf49*). A qPCR assay for the *CALD1* gene was designed and used for validation of the copy number of the region in the patient and both parents, and for the determination the copy number variation in his brother, confirming the presence of 3 copies of the fragment in patients 5, 6 and 7 (mother). The father presented two copies for the analyzed fragment (a result concordant with the aCGH).

A comparison between the molecular alterations present in the reported patients is present in figure 2 and table III.

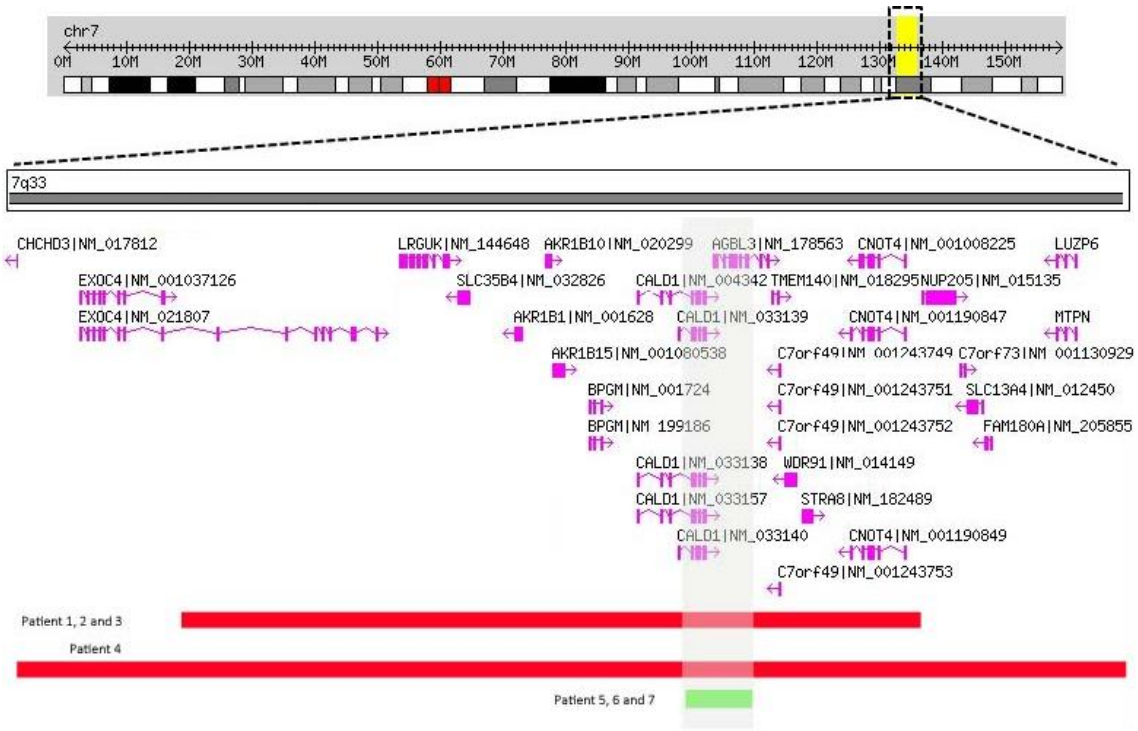


Figure 38 - Schematic representations and overlap of the CNVs found in the patients. A 3 Mb genomic portion of the cytoband 7q33 is shown. RefSeq genes present within the genomic region (in pink; transcriptional direction represented by the arrows) are shown. Shaded in grey is the overlapping deleted region for all the patients. Individual red horizontal bars represent deletions. In each CNV corresponding patient is indicated the.

Table XIII – Summary and comparison for the molecular findings present in the patients.

Clinical feature	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Patient 6	Patient 7
Gender	♂	♀	♀	♂	♂	♂	♀
Consanguinity	No	No	No	No	No	No	No
Molecular karyotyping (aCGH)	Agilent 180K	-	-	Agilent 44K	CytoScan 750K	-	CytoScan 750K
Confirmation	qPCR	qPCR	qPCR	NP	qPCR	qPCR	qPCR
CNV size	2,07Mb	-	-	3,04Mb	216Kb	-	-
Interval coordinates (Hg19)	chr7:133,176,651-135,252,871	-	-	chr7:132,766,430-135,802,894	chr7:134,598,205-134,807,358	-	-
Heritability	Maternal inherited	Maternal inherited	-	*de novo*	Maternal inherited	Maternal inherited	-
Genes affected	*AGBL3, AKR1B1, AKR1B10, AKR1B15, BPGM, C7orf49,* ***CALD1***, ***CNOT4***, ***EXOC4***, *LRGUK, NUP205, SLC35B4, STRA8, TMEM140, WDR91*	***CNOT4*** (performed by qPCR; presumably the same ones as patient 1)	***CNOT4*** (performed by qPCR; presumably the same ones as patient 1)	*AGBL3, AKR1B1, AKR1B10, AKR1B15, BPGM, C7orf49, C7orf73,* ***CALD1***, *CHCHD3,* ***CNOT4***, ***EXOC4***, *FAM180A, LRGUK, MTPN, NUP205, SLC13A4, LUZP6, STRA8, TMEM140, WDR91, SLC35B4*	***CALD1***, *AGBL3, C7orf49*	***CALD1*** (performed by qPCR; presumably the same ones as patient 5)	***CALD1***, *AGBL3, C7orf49*

Discussion

Several interstitial deletions of chromosome 7q have been described in the recent past ranging from 7,6 Mb to 13.8 Mb in size all (Malmgren *et al.*, 2005; Rossi *et al.*, 2008; Petrin *et al.*, 2010; Dilzell *et al.*, 2015; Kale and Philip, 2016).

In this work, we report seven patients with 7q33 CNVs, all affecting at least the *CALD1* gene. Of these, four patients present two large deletions also affecting the *EXOC4* and *CNOT4* genes. Considering that patient 1 and 2 are siblings and present the same deletion and very similar phenotypes, the main comparison should be made with patient 4. From the behavioral phenotype comparison patients 1, 2 and 4 display aggressive behavior, disinhibition and hyperactivity. Patients 1 and 2 also present some overlapping facial dysmorphisms with those of a patient previously described by Dilzell and colleagues - bulbous nose, thin upper lip, philtrum anomalies, small ears and low posterior hairline (Dilzell *et al.*, 2015). The deletions´ overlap for these four patients is determined by the deletion of patients 1, 2 and 3 (meaning a 2.08 Mb overlap, including a total of 15 genes shared by both deletions). An analysis for several constraints metrics of all the 21 genes affected in patient 4 (the largest deletion among the patients) is presented in table III. We can see that *EXOC4* and *CNOT4* are two of the genes with the lowest haploinsufficiency score. These genes are also affected in the smaller deletion present in patient 1, 2 and 3.

EXOC4 (EXOCYST COMPLEX COMPONENT 4) is one of the common genes deleted among the first four patients. *EXOC4* is the human homologous of Sec8 in yeast. *EXOC4*/Sec8 is a member of the exocyst complex, broadly expressed in rat brain, localized in the synapses and which plays a role in neurotransmitter release (Hsu *et al.*, 1996). Sec8 was described to be involved in the directional movement of AMPA-type glutamate receptors towards synapses, promoting the membrane communication between polarized cells, as well as in the delivery of NMDA (N-methyl-D-aspartate) receptors to the cell surface in neurons through the interaction of Sec8 PZD domain with synapse-associated protein 102 (SAP102) (Sans *et al.*, 2003; Gerges *et al.*, 2006). Sec8 was also described to bind to postsynaptic density protein-95 (PSD-95) (Riefler *et al.*, 2003).

In the literature, Yue colleagues reported a patient with DD and macrocephaly who presented a *de novo* translocation t(7;10)(q33;q23), together with a paternal 7Mb deletion at 7q33. The authors hypothesized that the phenotype might arise due to the resulting EXOC4-PTEN fusion protein and/or haploinsufficiency of the disrupted genes (Yue *et al.*, 2005). The patient had some clinical features in common with the four patients reported here: he also presented ID, delayed

speech, hypotonia and facial dysmorphisms. Unfortunately a picture is not available in order to make a comparison with the present cases (Thomas Haaf and Susan Holder, personal communication).

The heterozygous deletion of this gene is thus common to four of the present patients and to the patient reported by Yue and colleagues. At this point we can only hypothesize that *EXOC4* haploinsufficiency in our patients can result in neurotransmission and synaptic impairment, and thus contribute to ID in these patients. However, we cannot disregard that the deletions present in patients 1, 2, 3 and 4 encompass other interesting genes.

One of those is the *CNOT4* (CCR4-NOT transcription factor complex, subunit 4) gene which encodes a protein that belongs to the conserved Ccr4-Not complex, involved in biological processes such as transcription regulation, mRNA degradation, histone methylation and DNA repair (Collart, 2003; Kruk *et al.*, 2011; Grönholm *et al.*, 2012). The disruption of the proper methylation state of several genes has been shown to be associated with several neurodevelopmental disorders (see (Rudenko and Tsai, 2014) for revision). In yeast, the *CNOT4* homolog Not4 functions as an E3 ubiquitin ligase and controls the level of Jhd2, the yeast ortholog of *JARID1C* (Mersman *et al.*, 2009). This is interesting since mutations in *JARID1C* (lysine-specific demethylase 5C) were reported in patients with X-linked ID, revealing that the correct expression of this protein is essential for correct neuronal function (Abidi *et al.*, 2009; Ounap *et al.*, 2012; Brookes *et al.*, 2015). Mersman and colleagues (2009) demonstrated that in yeast JARID1C protein levels are also regulated by CNOT4 via a polyubiquitin-mediated degradation process (Mersman *et al.*, 2009). More recently, Not4 was also described to be involved in the regulation of JAK/STAT pathway-dependent gene expression, an important pathway involved in organogenesis and immune and stress response in *Drosophila* (Grönholm *et al.*, 2012, p. 4). The International Mouse Phenotyping Consortium reports that mice carrying a homozygous intragenic deletion in Cnot4 present pre-weaning lethality (with complete penetrance), while the heterozygous mice have an abnormal caudal vertebrae morphology, hematopoiesis and immune system defects ('Cnot4 MGI Mouse Gene Detail - MGI:1859026 - CCR4-NOT transcription complex, subunit 4', n.d.). No mention is made to CNS or cognitive deficits nor craniofacial features in these mice. However, the literature reports revealing its E3 ubiquitin ligase activity (UPS function being a common theme in neurodevelopmental genetics) and the functional connection to other known ID-causative genes further reinforce the possible contribution of *CNOT4* for the phenotype in patients 1, 2, 3 and 4.

CALD1 encodes for the caldesmon protein, which is widely expressed, including in the nervous system. Caldesmon is an actin-linked regulatory protein that binds and stabilizes actin filaments, and regulates actin-myosin interaction playing an important role in cell motility regulation (Lin et al., 2009). Since caldesmon has numerous functions in cell motility (such as migration, invasion, and proliferation), executed through the reorganization of the actin cytoskeleton (Mayanagi et al., 2011), its alteration is likely to have a functional contribution for ID pathogenesis, as this is a common biological theme linking many ID causative genes. Caldesmon overexpression induced by excess glucocorticoids was described to lead to altered patterns of neuronal radial migration trough the reorganization of the cytoskeleton and impact on nervous system structure and function (Fukumoto et al, 2009; Mayanagi T, 2008). Caldesmon is an important regulator of axon development (Morita *et al.*, 2012)] and may also play a role in synaptogenesis, synaptic plasticity and dendritic arborization (Sobue and Fukumoto, 2010). An important point to determine is whether the duplication occurs in *cis* or *trans* in patients 5, 6 and 7. If the duplication region is inserted in *cis* the rearrangement might lead to the ablation of the protein expression in the mutated allele, causing a similar effect as that of the deletion present in the rest of the patients here described.

In all the patients, the *AGBL3* (ATP/GTP binding protein-like 3) gene, which encodes a cytosolic carboxypeptidase (CCP3) belonging to a group of enzymes that catalyze posttranslational removal of acidic amino acids tails from tubulin (Tort *et al.*, 2014), is also affected. The deglutamylation of tubulin plays an important role in regulation of the microtubule cytoskeleton, of known relevance for neurons; in fact, it was shown that the control of the length of the polyglutamate side chains linked to tubulin is critical for neuronal survival (Rogowski *et al.*, 2010), which would make this gene a possible contributor to the patients' phenotype. However, although tubulin is a key protein in regulation of the microtubule cytoskeleton and this is of known relevance for neurons (Rogowski *et al.*, 2010) there is not enough evidence that *AGBL3* does have a function in cytoskeleton regulation or in neurons. In fact, according to the GTEx portal, *AGBL3* has very low expression in most of the tissues in human, with only the testis presenting a slightly higher expression at the mRNA level (table III). For these reasons its contribution for the phenotype remains undetermined.

Beside the analysis of the candidate genes in the 7q33 affected region is also important to take into account the patients described in Decipher, with deletions and duplications that partially overlap the 7q33 affected region, summarized in table IV. Regarding the deletions, there are two

patients (Decipher 280233 and 331287) with small inherited deletions affecting only the *EXOC4* gene. Even though for patient 331287 the submitters classified it as likely pathogenic, the phenotypic description of the progenitor is not provided. Concerning the duplications, there are two Decipher patients (255520 and 251768) carrying duplications affecting *EXOC4*, inherited from normal parents. As mentioned before, in these cases it is important to determine if the duplicated region is located in *cis* or *trans* in order to fully understand the impact of the duplication in the expression of the contained genes. For this reason, the inherited duplications in Decipher cases 255520 and 251768 should be interpreted with caution. Nevertheless, these four Decipher cases raise doubts about the straightforward contribution of *EXOC4* for the NDD phenotype, leaving *CNOT4* and *CALD1* as more promising candidates.

Conclusion

This chapter presents the first seven patients with interstitial 7q33 CNVs and suggest that *EXOC4*, *CNOT4* and *CALD1* genes are likely contributing for ID and behavioral phenotype. Copy number variations in this region should be considered with care when present in patients with ID and behavioral alterations. Further studies (such as the genes' expression in peripheral blood cells) need to be performed in order to better understand the contribution of each gene within this region to the phenotype.

Table XIV - OMIM entrance, haploinsufficiency score and constrain metrics for the genes deleted in patient 4 (the largest deletion).

7q33	**List of all the genes affected in P4**	***AGBL3, AKR1B1, AKR1B10, AKR1B15, BPGM, C7orf49, C7orf73, CALD1, CHCHD3, CNOT4, EXOC4, FAM180A, LRGUK, MTPN, NUP205, SLC13A4, LUZP6, STRA8, TMEM140, WDR91, SLC35B4***							
Gene	**Morbid gene**	**OMIM**	**% HI score**	**DDG2P**	**ClinVar**	**Constraint Metrics**			
						Synonymous (z)	**Missense (z)**	**LoF (pLI)**	**CNV (z)**
AGBL3	No	-	60-70%	-	10dels/11dups	-	-	-	-
AKR1B1	No	-	30-30%	-	9dels/10dups	-0.32	0.27	0	-2.25
AKR1B10	No	-	70-80%	-	9dels/11dups	-0.28	-0.27	0	-4.12
AKR1B15	No	-	80-90%	-	9dels/11dups	0.02	-1.03	0	-3.36
BPGM	Yes	222800, Erythrocytosis due to bisphosphoglycerate mutase deficiency, AR	20-30%	-	11dels/11dups/1SNV	0.16	0.77	0.13	0.5
C7orf49	No	-	80-90%	-	10dels/11dups	-0.13	-0.36	0.34	0.56
C7orf73	No	-	20-30%	-	11dels/11dups	-	-	-	-
CALD1	No	-	20-30%	-	10dels/11dups	1.02	-0.14	1	0.73
CHCHD3	No	-	0-10%	-	9dels/13dups	0.18	0.15	0.04	-0.13
CNOT4	No	-	0-10%	-	11dels/12dups	0.14	3.38	1	0.81
EXOC4	No	-	0-10%	-	18dels/18dups/1SNV	-0.09	-0.27	0	-1.74
FAM180A	No	-	60-70%	-	11dels/11dups	-0.26	-0.33	0.34	1.16
LRGUK	No	-	70-80%	-	10dels/12dups	0.6	-1.63	0	-1.4
MTPN	No	-	10-20%	-	11dels/11dups	0.57	2.05	0.75	0.98
NUP205	Yes	616893, ?Nephrotic syndrome, type 13	10-20%	-	11dels/12dups/1SNV	-0.77	0.87	1	0.18
SLC13A4	No	-	40-50%	-	11dels/11dups	0.64	2.16	0.92	-0.96
LUZP6	No	-	80-90%	-	11dels/11dups	-	-	-	-
(Cont.)									
STRA8	No	-	50-60%	-	10dels/11dups	1.42	0.74	0	0.51
TMEM140	No	-	80-90%	-	10dels/11dups	-0.01	-0.05	0.04	-
WDR91	No	-	40-50%	-	10dels/11dups	0.7	1.12	0	0.51
SLC35B4	No	-	20-30%	-	9dels/12dups	-1.1	0.44	0	0.04

OMIM: Online Mendelian Inheritance in Man; HI score: Haploinsufficiency Score index - high ranks (e.g. 0-10%) indicate a gene is more likely to exhibit haploinsufficiency, low ranks (e.g. 90-100%) indicate a gene is more likely to NOT exhibit haploinsufficiency (retrieved from Decipher); LoF: Loss of function; CNVs: copy number variations; z: Z score is the deviation of observed counts from the expected number for one gene (positive Z scores = gene intolerance to variation, negative Z scores = gene tolerant to variation) (retrieved from ExAC); pLI: probability that a given gene is intolerant of loss-of-function variation (pLI closer to one = more intolerant the gene is to LoF variants, pLI >= 0.9 is extremely LoF intolerant) (retrieved from ExAC); del – deletion; dup – duplication; SNV – single nucleotide variant; ins – insertion; indel – insertion/deletion

Table XV – Summary of the Decipher patients with relatively small and overlapping CNVs.

Decipher number	CNV	Size	Genes affected*	Inheritance	Pathogenicity	Index phenotype	Parent phenotype
280233	del	178Kb	*EXOC4*	paternal	ND	ID	ND
253613	del	45Kb	*EXOC4*	ND	ND	ID, autism, speech delay, hypotonia, obesity, puberty delay, limb abnormalities (short foot and tapered finger)	ND
262735	del	259Kb	*EXOC4*	ND	ND	ID, behavioral abnormalities, hypotonia, atopic dermatitis	ND
271567	del	160Kb	*EXOC4*	ND	ND	ND	ND
273272	del	139Kb	*AKR1B1*	ND	ND	ID	ND
333171	del	121Kb	*EXOC4*	ND	ND	Behavioral abnormality, language impairment	ND
338702	del	468Kb	*EXOC4*	ND	ND	Behavioral abnormality, delayed speech and language development	ND
331287	del	585Kb	*EXOC4*	maternal	Likely pathogenic	Developmental delay	ND
267399	del	123Kb	*EXOC4, LRGUK*	ND	ND	ND	ND
328659	del	2.6Mb	*AGBL3, AKR1B1, AKR1B10, AKR1B15, BPGM, C7orf49, C7orf73,* ***CALD1****, CHRM2, CNOT4, FAM180A, LUZP6, MTPN, NUP205, SLC13A4, STRA8, TMEM140, WDR91*	*De novo*	Likely pathogenic (partially explaining part of the phenotype)	ID, psychosis	ND
282285	dup	487Kb	*EXOC4, LRGUK, SLC35B4*	maternal	Uncertain (has a *de novo* pathogenic larger del in chr9)	Autism	ND
305865	dup	346Kb	*EXOC4, LRGUK, SLC35B4*	ND	Uncertain	Autism, global developmental delay	ND
255520	dup	719Kb	*CHCHD3, EXOC4*	Inherited from normal parent	ND	ND	Healthy
251768	dup	828Kb	*AKR1B1, AKR1B10, AKR1B15, BPGM, EXOC4, LRGUK, SLC35B4*	Inherited from normal parent	ND	ID, hypotonia, brachydactyly, sparse hair, synophrys, abnormal dental morphology, high and narrow palate, open mouth, microcephaly, strabismus, large hears, heart defects (atrial and ventricular septal defect, coarctation of aorta)	Healthy
256271	dup	1Mb	*CHCHD3, EXOC4, PLXNA4*	*De novo*	ND	ND	ND

ND: not determined; *genes affected in the Decipher patient

References

Abidi F, Holloway L, Moore CA, Weaver DD, Simensen RJ, Stevenson RE, et al. Novel human pathological mutations. Gene symbol: JARID1C. Disease: mental retardation, X-linked. Hum. Genet. 2009; 125: 344.

Brookes E, Laurent B, Õunap K, Carroll R, Moeschler JB, Field M, et al. Mutations in the intellectual disability gene KDM5C reduce protein stability and demethylase activity. Hum. Mol. Genet. 2015

Buffart TE, Israeli D, Tijssen M, Vosse SJ, Mrsić A, Meijer GA, et al. Across array comparative genomic hybridization: a strategy to reduce reference channel hybridizations. Genes. Chromosomes Cancer 2008; 47: 994–1004.

Collart MA. Global control of gene expression in yeast by the Ccr4-Not complex. Gene 2003; 313: 1–16.

Dilzell K, Darcy D, Sum J, Wallerstein R. Deletion of 7q33-q35 in a Patient with Intellectual Disability and Dysmorphic Features: Further Characterization of 7q Interstitial Deletion Syndrome. Case Rep. Genet. 2015; 2015: 131852.

Gerges NZ, Backos DS, Rupasinghe CN, Spaller MR, Esteban JA. Dual role of the exocyst in AMPA receptor targeting and insertion into the postsynaptic membrane. EMBO J. 2006; 25: 1623–1634.

Grönholm J, Kaustio M, Myllymäki H, Kallio J, Saarikettu J, Kronhamn J, et al. Not4 enhances JAK/STAT pathway-dependent gene expression in Drosophila and in human cells. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. 2012; 26: 1239–1250.

Hoebeeck J, van der Luijt R, Poppe B, De Smet E, Yigit N, Claes K, et al. Rapid detection of VHL exon deletions using real-time quantitative PCR. Lab. Investig. J. Tech. Methods Pathol. 2005; 85: 24–33.

Hsu SC, Ting AE, Hazuka CD, Davanger S, Kenny JW, Kee Y, et al. The mammalian brain rsec6/8 complex. Neuron 1996; 17: 1209–1219.

Jovanovic L, Delahunt B, McIver B, Eberhardt NL, Grebe SKG. Optimising restriction enzyme cleavage of DNA derived from archival histopathological samples: an improved HUMARA assay. Pathology (Phila.) 2003; 35: 70–74.

Kale T, Philip M. An Interstitial Deletion at 7q33-36.1 in a Patient with Intellectual Disability, Significant Language Delay, and Severe Microcephaly. Case Rep. Genet. 2016; 2016: 6046351.

Krijgsman O, Israeli D, van Essen HF, Eijk PP, Berens MLM, Mellink CHM, et al. Detection limits of DNA copy number alterations in heterogeneous cell populations. Cell. Oncol. Dordr. 2013; 36: 27–36.

Kruk JA, Dutta A, Fu J, Gilmour DS, Reese JC. The multifunctional Ccr4-Not complex directly promotes transcription elongation. Genes Dev. 2011; 25: 581–593.

Malmgren H, Malm G, Sahlén S, Karlsson M, Blennow E. Molecular cytogenetic characterization of an insertional translocation, ins(6;7)(p25;q33q34): deletion/duplication of 7q33-34 and clinical correlations. Am. J. Med. Genet. A. 2005; 139: 25–31.

Mersman DP, Du H-N, Fingerman IM, South PF, Briggs SD. Polyubiquitination of the demethylase Jhd2 controls histone methylation and gene expression. Genes Dev. 2009; 23: 951–962.

Mitchell S, Siegel DH, Shieh JTC, Stevenson DA, Grimmer JF, Lewis T, et al. Candidate locus analysis for PHACE syndrome. Am. J. Med. Genet. A. 2012; 158A: 1363–1367.

Morita T, Mayanagi T, Sobue K. Caldesmon Regulates Axon Extension through Interaction with Myosin II. J. Biol. Chem. 2012; 287: 3349–3356.

Nielsen KB, Egede F, Mouridsen I, Mohr J. Familial partial 7q monosomy resulting from segregation of an insertional chromosome rearrangement. J. Med. Genet. 1979; 16: 461–466.

Ounap K, Puusepp-Benazzouz H, Peters M, Vaher U, Rein R, Proos A, et al. A novel c.2T > C mutation of the KDM5C/JARID1C gene in one large family with X-linked intellectual disability. Eur. J. Med. Genet. 2012; 55: 178–184.

Petrin AL, Giacheti CM, Maximino LP, Abramides DVM, Zanchetta S, Rossi NF, et al. Identification of a microdeletion at the 7q33-q35 disrupting the CNTNAP2 gene in a Brazilian stuttering case. Am. J. Med. Genet. A. 2010; 152A: 3164–3172.

Ponnala R, Dalal A. Partial monosomy 7q. Indian Pediatr. 2011; 48: 399–401.

Riefler GM, Balasingam G, Lucas KG, Wang S, Hsu S-C, Firestein BL. Exocyst complex subunit sec8 binds to postsynaptic density protein-95 (PSD-95): a novel interaction regulated by cypin (cytosolic PSD-95 interactor). Biochem. J. 2003; 373: 49–55.

Rogowski K, Dijk J Van, Magiera MM, Bosc C, Deloulme J, Bosson A, et al. A Family of Protein-Deglutamylating Enzymes Associated with Neurodegeneration. Cell 2010; 143: 564–578.

Rossi E, Verri AP, Patricelli MG, Destefani V, Ricca I, Vetro A, et al. A 12Mb deletion at 7q33-q35 associated with autism spectrum disorders and primary amenorrhea. Eur. J. Med. Genet. 2008; 51: 631–638.

Rudenko A, Tsai L-H. Epigenetic modifications in the nervous system and their impact upon cognitive impairments. Neuropharmacology 2014; 80: 70–82.

Sans N, Prybylowski K, Petralia RS, Chang K, Wang Y-X, Racca C, et al. NMDA receptor trafficking through an interaction between PDZ proteins and the exocyst complex. Nat. Cell Biol. 2003; 5: 520–530.

Sobue K, Fukumoto K. Caldesmon, an actin-linked regulatory protein, comes across glucocorticoids. Cell Adhes. Migr. 2010: 185–189.

Stallard R, Juberg RC. Partial monosomy 7q syndrome due to distal interstitial deletion. Hum. Genet. 1981; 57: 210–213.

Tort O, Tanco S, Rocha C, Bièche I, Seixas C, Bosc C, et al. The cytosolic carboxypeptidases CCP2 and CCP3 catalyze posttranslational removal of acidic amino acids. Mol. Biol. Cell 2014; 25: 3017.

Verma RS, Conte RA, Sayegh SE, Kanjilal D. The interstitial deletion of bands q33-35 of long arm of chromosome 7: a review with a new case report. Clin. Genet. 1992; 41: 82–86.

Xu Z, Geng Q, Luo F, Xu F, Li P, Xie J. Multiplex ligation-dependent probe amplification and array comparative genomic hybridization analyses for prenatal diagnosis of cytogenomic abnormalities. Mol. Cytogenet. 2014; 7: 84.

Yue Y, Grossmann B, Holder SE, Haaf T. De novo t(7;10)(q33;q23) translocation and closely juxtaposed microdeletion in a patient with macrocephaly and developmental delay. Hum. Genet. 2005; 117: 1–8.

Cnot4 MGI Mouse Gene Detail - MGI:1859026 - CCR4-NOT transcription complex, subunit 4 [Internet]. [cited 2017 Apr 14] Available from: http://www.informatics.jax.org/marker/MGI:1859026

Supplementary Table I – Primers used for quantitative real-time PCR confirmation of copy number changes.

Gene	RefSeq	Primer Forward 5'→3'	Primer location	Primer Reverse 5'→3'	Amplicon size (bp)
CNOT4	NM_001190850.1	CACCGAGCGGTTTATAATTCA	Exon 10	AGACCTGTGTTGTGCTGTGG	164
CALD1	NM_033139.3	GAATGACGATGATGAAGAGGAG	Exon 4	ACAGTACCTGTTCTGGGCATTC	139
ZNF80	NM_007136	GCTACCGCCAGATTCACACT	Exon 1	AATCTTCATGTGCCGGGTTA	182
SDC4	NM_002999	ACCGAACCCAAGAAACTAGA	Exon 4	GTGCTGGACATTGACACCT	101

Variant Rett syndrome in a girl with a pericentric X-chromosome inversion leading to epigenetic changes and overexpression of the *MECP2* gene

Int. J. Devl Neuroscience 46 (2015) 82–87

Contents lists available at ScienceDirect

International Journal of Developmental Neuroscience

journal homepage: www.elsevier.com/locate/ijdevneu

Variant Rett syndrome in a girl with a pericentric X-chromosome inversion leading to epigenetic changes and overexpression of the *MECP2* gene

José Pedro Vieira[a,1], Fátima Lopes[b,c,1], Anabela Silva-Fernandes[b,c], Maria Vânia Sousa[a], Sofia Moura[b,c], Susana Sousa[b,c], Bruno M. Costa[b,c], Mafalda Barbosa[d,e], Bauke Ylstra[f], Teresa Temudo[g], Teresa Lourenço[a], Patrícia Maciel[b,c,*]

[a] Neurology Department, Hospital Dona Estefânia, Centro Hospitalar de Lisboa Central, Portugal
[b] Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, Portugal
[c] ICVS/3B's – PT Government Associate Laboratory, Braga/Guimarães, Portugal
[d] Department of Genetics and Genomic Sciences, The Mindich Child Health & Development Institute, The Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, NY, USA
[e] Instituto Gulbenkian de Ciência, Oeiras, Portugal
[f] Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands
[g] Neuropediatrics Department, Centro Hospitalar do Porto, Portugal

ARTICLE INFO

Article history:
Received 18 May 2015
Received in revised form 3 July 2015
Accepted 15 July 2015
Available online 10 August 2015

Keywords:
Epigenetics
Epilepsy
Intellectual disability
Neurodevelopment

ABSTRACT

Rett syndrome is a neurodevelopmental disorder caused by mutations in the *MECP2* gene. We investigated the genetic basis of disease in a female patient with a Rett-like clinical. Karyotype analysis revealed a pericentric inversion in the X chromosome −46,X,inv(X)(p22.1q28), with breakpoints in the cytobands where the *MECP2* and *CDKL5* genes are located. FISH analysis revealed that the *MECP2* gene is not dislocated by the inversion. However, and in spite of a balanced pattern of X inactivation, this patient displayed hypomethylation and an overexpression of the *MECP2* gene at the mRNA level in the lymphocytes (mean fold change: 2.55 ± 0.38) in comparison to a group of control individuals; the expression of the *CDKL5* gene was similar to that of controls (mean fold change: 0.98 ± 0.10). No gains or losses were detected in the breakpoint regions encompassing known or suspected transcription regulatory elements. We propose that the de-regulation of *MECP2* expression in this patient may be due to alterations in long-range genomic interactions caused by the inversion and hypothesize that this type of epigenetic de-regulation of the *MECP2* may be present in other RTT-like patients.

1. Introduction

Rett syndrome (RTT) is a neurodevelopmental disorder characterized by an early regression followed by stagnation in growth and development, leading to mental retardation, stereotypies and motor apraxia that may be associated with epilepsy and with dysautonomic features, including disturbed breathing, sleep and gastrointestinal motility (Hagberg et al., 1983; Rett, 1966). RTT is caused by mutations in the *MECP2* gene, encoding a methyl-CpG binding protein thought to be involved in transcriptional regulation (Amir et al., 1999; Hoffbuhr et al., 2001; Philippe et al., 2006). Mutations in this gene are present in more than 90% of patients that fulfill the diagnostic criteria, and with lesser frequency in patients with so-called variant forms of the disease, that do not comply with all these criteria (Amir et al., 1999; Hoffbuhr et al., 2001; Philippe et al., 2006).

The most frequent mutations in *MECP2* are point mutations, small insertions or deletions. The genotype of patients without one of these mutations has also been investigated and some cases with gross rearrangements involving the *MECP2* gene were found, usually large deletions. In boys, *MECP2* duplication is the rare cause of a severe encephalopathy with recurrent infections, severe mental retardation, microcephaly and epilepsy (Ramocki et al., 2009; Van Esch et al., 2005). *MECP2* duplication has also been described in a

* Corresponding author at: Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, 4710-057 Braga, Portugal. Tele/Fax: +351 253 604824/20.
E-mail address: pmaciel@ecsaude.uminho.pt (P. Maciel).
[1] Both authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ijdevneu.2015.07.010
0736-5748/

girl with a preserved-speech variant of RTT and in another with nonspecific mental retardation (Ariani et al., 2004; Makrythanasis et al., 2010). Recently Bijlsma et al. reviewed *MECP2* duplication in eleven females; five were from their study and six had been previously reported. Intellectual disability was a prominent feature while hand stereotypies, epilepsy, dysmorphism and recurrent infections were less frequent (Bijlsma et al., 2012). Overexpression of normal MeCP2 has also been shown to cause a progressive neurological disorder in a mouse model, showing that the nervous system requires a very precise dosage of this protein for proper functioning (Collins et al., 2004).

Here, we describe a patient with a RTT-like clinical presentation in whom we found no mutations in the *MECP2* gene by direct sequencing, nor changes in exon dosage. Karyotype analysis, however, revealed a pericentric inversion in the X chromosome 46,X,inv(X)(p22.1q28). This inversion is accompanied by epigenetic changes in the form of decreased DNA methylation levels within the 5′ regulatory region of *MECP2* gene, and led to a significantly increased expression of the gene, in spite of its normal dosage, through a mechanism that may involve relatively distant regions of the X chromosome.

2. Materials and methods

2.1. Patient assessment

The proband was ascertained and assessed through the Hospital Dona Estefânia (Centro Hospitalar de Lisboa Central) and *MECP2* study was requested at the Molecular Diagnostic Service at Life and Health Sciences Research Institute (ICVS) – School of Health Sciences, University of Minho, Braga, Portugal. Informed consent was obtained from the child's parents for blood sampling and genetic analyses. Genomic DNA was extracted from peripheral blood using the Puregene DNA isolation kit (Gentra, Minneapolis, MN).

2.2. MECP2 mutation detection

The coding region and exon-intron boundaries of the *MECP2* gene were amplified by PCR and sequenced on an automated DNA-sequencer (ABI-3130 Genetic Analyzer, Applied Biosystems, Life technologies).

Copy-number analysis of the *MECP2* gene was performed by multiplex ligation-dependent probe amplification (MLPA) (MRC-Holland, Amsterdam, Netherlands) using SALSA P015 kit and quantitative PCR analysis (qPCR). For MLPA, PCR reactions were carried out using C1000 – Thermal Cycler (BioRad) thermocycler according to the manufacturer's instructions. Products were separated in a 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) using ROX-500 as internal pattern. The data was analyzed for copy-number differences with Coffalyser software (MRC-Holland, Amsterdam, Netherlands). Primers for qPCR were designed using Primer3Plus software (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and taking into account standard recommendations for qPCR primer development (Jovanovic et al., 2003). Four sets of primers were designed for the *MECP2* gene (ENSG00000169057) within the coding region of exon one, two and three and in the 3′UTR non-coding region of exon four. The reference genes used were *SDC4* (ENSG00000124145) and *ZNF80* (ENSG00000174255) localized in the 20q12-q13 and 3p12 regions, respectively. qPCR reactions were carried out in a 7500-FAST Real Time PCR machine (Applied Biosystems, Foster City, CA, USA) using Power SYBR Green® (Applied Biosystems). The specificity of each of the reactions was verified by the generation of a melting curve for each of the amplified fragments. The primer efficiency was calculated by the generation of a standard curve fitting the accepted normal efficiency percentage. Quantification was performed as described elsewhere (Hoebeeck et al., 2005). Ct values obtained for each test were analyzed in DataAssist™ software (Applied Biosystems, Foster City, CA, USA).

2.3. X chromosome inactivation

The X-inactivation assay (XCI) was carried out by analysis of the HUMARA (human androgen receptor) locus, which is a highly polymorphic CAG repeat with a very near Hhal enzyme restriction site, according to the described protocol (Jovanovic et al., 2003). The products were separated on an automated DNA-sequencer (ABI-3130 Genetic Analyzer, Applied Biosystems, Life technologies) the allele peak height serving as the semi-quantitative measure of the amount of PCR product amplified from each allele. The samples were defined as having a deviated XCI pattern if they presented a corrected allele ratio out of the 0.33–3 range.

2.4. Conventional karyotyping and FISH

Standard karyotpe analysis was performed using cells cultured from peripheral blood leukocytes, and FISH analysis was undertaken using two distinct probes covering the Xq28 region: RP4-671D9 (153286964–153386210) and RP11-617G06 (153110141–153283527) (BlueGnome, UK).

2.5. Molecular karyotyping

The aCGH analysis was performed on a human genome CGH Agilent 180K custom array designed by the Low Lands Consortium (LLC, Dr. Klas Kok) in order to be used in the analysis of children with ID/DD (AMADID:023363; Agilent, Santa Clara, CA). DNA labeling was performed using the ENZO Labeling Kit for Oligo Arrays (Enzo Life Sciences, Inc.). As reference DNA we used Kreatechís MegaPoll Reference DNA (Kreatech Diagnostics). Arrays were then hybridized using the Agilent SurePint G3Human CGH Microarray Kit. Data was extracted with the Agilent Feature Extraction (FE) Software v10.5 using default settings for CGH hybridizations. Image analysis was performed using the across-array methodology described previously (Buffart et al., 2008). CGH data was analyzed using Nexus Copy Number 6.0 software with FASST2 Segmentation algorithm. The custom 1 M Agilent array was designed using the eArray web tool (https://earray.chem.agilent.com/earray/). The maximal amount of probes available was selected for the breakpoint regions (Xq22.1 and Xq28). DNA labeling, hybridization and data extraction were performed in the same way as for aCGH 180K (described above). CGH data was analyzed using Nexus Copy Number 6.0 software with the FASST2 Segmentation algorithm.

2.6. mRNA expression analysis

Total RNA was isolated from leucocytes using TRIZOL (Invitrogen, Calrsbad, CA, USA) according to the manufacturer's protocol. First-strand cDNA, synthesized using iScript™ cDNA Synthesis Kit (Bio-rad), was amplified by quantitative reverse-transcriptase PCR (qRT-PCR) using QuantiTect SYBR Green PCR Kit (Quiagen). Human *MECP2*, *CDKL5* and *HPRT* primers were used for relative quantification analysis. *MECP2* (Forward: 5′-GGCGCTCCATCATCCGTGAC-3′; Reverse: 5′-CTGGGGATTGATCAAATACACAT-3′), *CDKL5* (Forward: 5′-CATTCACTGTCTGCACCTCAC-3′; Reverse: 5′-GTTCTCCAGGTCGGGGTGAC-3′) and *HPRT* (Forward: 5′-CTAATTATGGACAGGGACTGAACG-3′; Reverse: 5′-CAGTCATAGGAATGGATCTATCA-3′). The expression levels of the genes were normalized to the *HPRT* gene and relative quantification was used to determine the fold change difference between *MECP2*, *CDKL5* and *HPRT*, using the DDCT method, as described before (Pfaffl, 2001). The comparison between the case study

($n = 1$) and controls ($n = 6$) is presented, with the respective *t*-test analysis for replicate 1. This difference was reproduced for the three replicates.

2.7. Methylation analysis

For epigenetic analysis the EpiTect Methyl II PCR Primer Assay for Human *MECP2* was used where a region of the CpG island 106 (ChrX: 153362568–153363473) was assessed (CpG Island 115358) EPHS115358-1A (Quiagen®) in the index case, mother and 4 healthy female controls. Briefly, 250 ng of DNA from each sample was digested for 6 h at 37 °C degrees. The samples were them amplified by real time PCR using the kit specific *MECP2* primers (EPHS115358-1A) with RT^2 qPCR SYBR Green ROX master mix in a 7500-FAST Real Time PCR machine (Applied Biosystems, Foster City, CA, USA). Ct values obtained for each sample were analyzed accordingly with the manufacturer's instructions in order to obtain the gene methylation status, as the percentage of unmethylated (UM) and methylated (M) DNA in the input genomic sequence. A *t*-test analysis was performed for the unmethylated DNA in replicate 1 of the assay where the case study ($n = 1$) is compared with the controls ($n = 4$). This difference was reproduced for the three replicates.

3. Results

3.1. Clinical description

The patient is the second child of healthy, non-consanguineous parents and has no family history of Rett syndrome, intellectual disability or epilepsy. Gestation was uneventful. Delivery, at 39 weeks, was complicated by a difficult forceps extraction. Apgar scores were 8 at 1 min and 10 at 5 min. Birth weight was 3755 g (75th centile), length 50 cm (50th centile) and head circumference 35.5 cm (75th centile). A brain computed tomography (CT) showed a parietal skull fracture and a small underlying epidural hematoma without parenchymatous brain lesions. In the first days, she remained clinically well, and a control CT scan showed total resolution of the hematoma. The child was first seen by us when she was forty days old, after a brief left-sided clonic partial motor seizure. The neurologic examination was normal. Electroencephalogram (EEG) showed focal paroxysmal activity. CT scan and magnetic resonance imaging (MRI) were normal. Phenobarbital was introduced and up to five months she experienced a few more seizures with a similar pattern. Subsequently she remained asymptomatic. The EEG, repeated at five months, was normal. At this time, however, psychomotor development was clearly in the delayed range. The patient had axial hypotonia and her hands infrequently reached the midline. At nine months she did not sit without support, was unable to reach out and pick objects and her lower limbs were slightly spastic. From twelve months on, RTT-like hand stereotypies appeared and became progressively more frequent. She did not acquire at this time any expressive or comprehensive language and did not point to objects. In the period between four and eighteen months head circumference went from the 75th centile to slightly below the 5th centile. Formal developmental evaluation was performed using the Ruth Griffiths scale at the age of four (45 months), revealing a mental age of 15 months and a Global Developmental Quotient of 34%.

At the age of six to seven she was walking with minimal support, hand movements were more coordinated and intentional, she held a pen and scratched, was able to point to some figures in a book and to say six to eight simple words. Currently, at the age of ten, her vocabulary is still limited to eight words, fine and gross motor skills remain at the same level and RTT-like hand stereotypies are still present, although less frequently than before. Teeth grinding is frequent. Hyperpnea, apnea, insomnia, unprovoked laughing/screaming and autistic features were not reported by the parents. There is no history of recurrent and/or severe infectious diseases. Recently, epilepsy relapsed with brief hypomotor seizures that were controlled with valproate and oxcarbazepine. EEG reveals posterior temporal spikes.

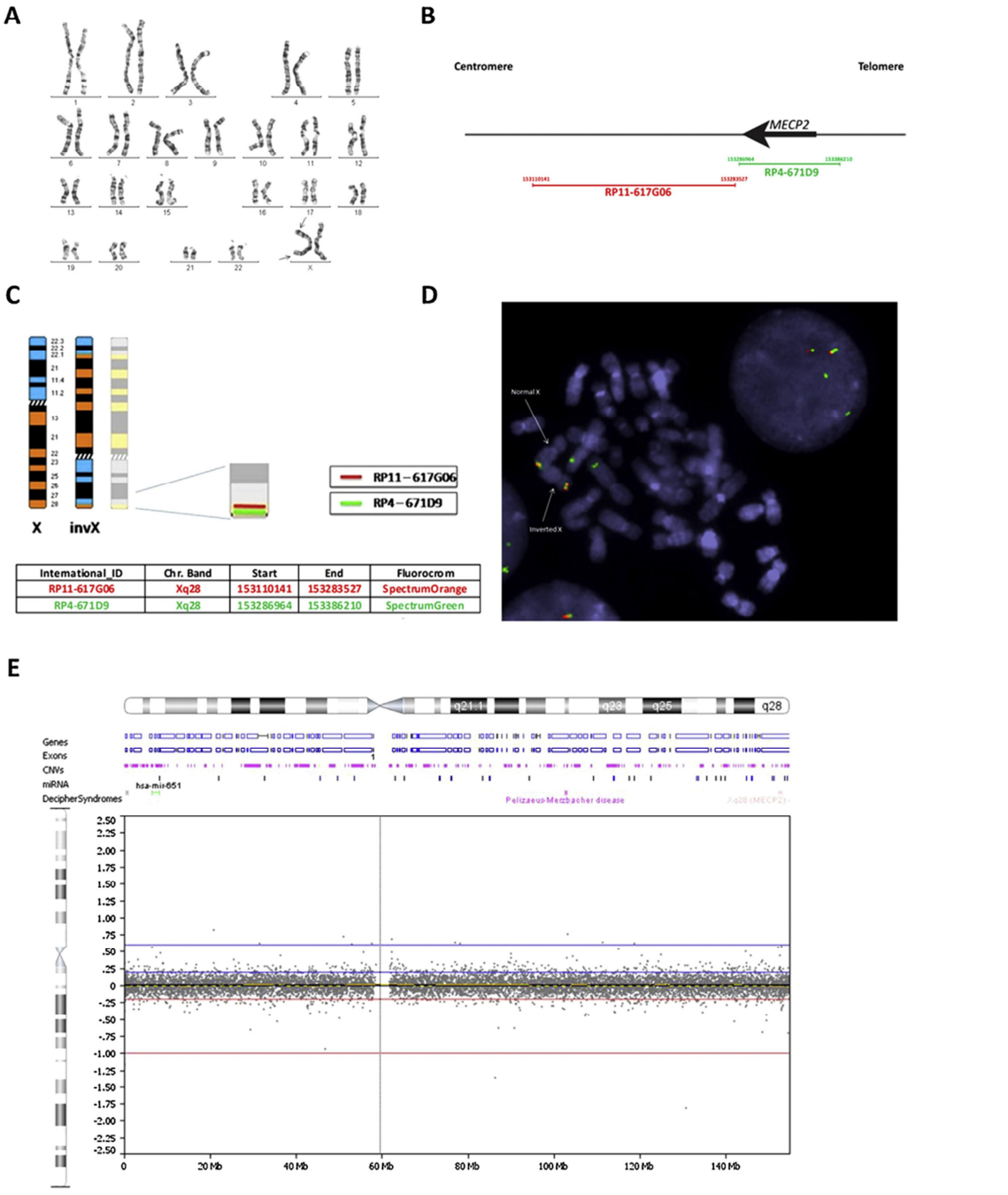
The child is generally happy, smiles quite often and has good visual interaction. She has a divergent strabismus, misalignment and deformity of the teeth, a slight dorsolumbar scoliosis and no dysmorphic features. Her feet are often cold and pale. Neurologic examination is remarkable for the microcephaly, expressionless face suggesting an extrapyramidal hypokinetic disorder, Rett-like hand movements, spastic hypertonia of lower limbs, dyspraxic gait and brisk reflexes with flexor plantar responses.

3.2. MECP2 mutation analysis and karyotyping

We found no variants in the coding region and intron-exon boundaries of this gene by direct Sanger sequencing, nor changes in exon dosage by MLPA and quantitative PCR analysis (qPCR) for the 4 exons. Karyotype analysis in cells cultured from peripheral blood leukocytes, however, revealed a pericentric inversion in the X chromosome 46,X,inv(X)(p22.1q28) in the patient, which was not present in the mother or father (Fig. 1A). Of notice, the *MECP2* gene is located in cytoband Xq28, and the *CDKL5* gene in Xp22.1. We then performed FISH to test the hypothesis that the pericentric inversion could have displaced the *MECP2* coding sequence or its nearby regulatory regions. We used two probes, one (RP4-671D9) covering the *MECP2* gene and 22998 bp of its 5′ region, and the second (RP11-G17G06) covering 173386 bp of 3′ downstream region of the *MECP2* gene (see Fig. 1, panel B and C). The analysis, however, revealed that the *MECP2* gene 5′, exonic, and 3′ regions had not been separated by the inversion. These results place the breakpoint at Xq28 more proximally than the location of the RP11-617G06 probe (153110141–153283527, hg19 coordinates), meaning that the breakpoint is not occurring within a distance of 176.883 Kb downstream the 3′ UTR of the *MECP2* gene. In order to clarify if the inversion lead to any loss or gain in the breakpoints we undertook aCGH analysis on a human genome CGH Agilent 180K custom array (Buffart et al., 2008) and on a high-resolution custom 1 M Agilent array designed so that the maximal amount of Agilent probes available was selected for the breakpoint regions (Xp22.1 and Xq28 where the *CDKL5* and *MECP2* genes are located, respectively) and at the citobands 14q12 (where the *FOXG1* gene is located). No significant CNVs were found in these regions in both analyses.

3.3. MECP2 mRNA expression and methylation status

We then investigated by qRT-PCR the expression levels of the *MECP2* (located at Xq28) and *CDKL5* (located at Xp22) genes. Intriguingly, the patient displayed an overexpression of the *MECP2* gene at the mRNA level in the lymphocytes (mean fold change for the three replicates: 2.55 ± 0.38), in comparison to a group of 6 control female individuals; in contrast, the expression of the *CDKL5* gene was similar to that of controls (mean fold change for the three replicates: 0.98 ± 0.10) (Fig. 2A). The X-inactivation assay (XCI), carried out by analysis of the HUMARA locus (Jovanovic et al., 2003), showed that the patient had a balanced pattern of X inactivation, with a ratio of 1/1. To further explore the mechanism leading to the overexpression of *MECP2*, we assessed the methylation status of a portion of the CpG island located in the 5′-regulatory region of the *MECP2* using the EpiTect® Methyl II PCR Assay (Quiagen®). These analyses showed decreased levels of



Intemational_ID	Chr. Band	Start	End	Fluorocrom
RP11-617G06	Xq28	153110141	153283527	SpectrumOrange
RP4-671D9	Xq28	153286964	153386210	SpectrumGreen

Fig. 1. Cytogenetic and molecular analysis. (A) Karyotype of the patient; (B) schematic representation of the position of the FISH probes relative to the *MECP2* gene; (C) Scheme of the results of the FISH analysis of the patient, undertaken using two distinct probes targeting the Xq28 region: RP4-671D9 (153286964–153386210) and RP11-617G06 (153110141–153283527) (coordinates in hg19); (D) FISH image (arrows indicate the inverted and normal X chromosomes); (E) 180K custom array overview of the X chromosome showing absence of CNVs.

methylation at the assessed CpG site (Fig. 2C), a pattern that is likely associated with the increased *MECP2* gene expression observed in lymphocytes, and which we speculate may also be happening in the nervous system. These findings lead us to speculate that *MECP2* overexpression is the most likely disease causing mechanism in this patient.

4. Discussion

Rett syndrome (RTT) results in a large fraction of patients (more than 90%) from mutations in the *MECP2* gene in patients that fulfill the diagnostic criteria and, in lesser frequency, in patients with so-called variant forms of the disease, that do not comply with all

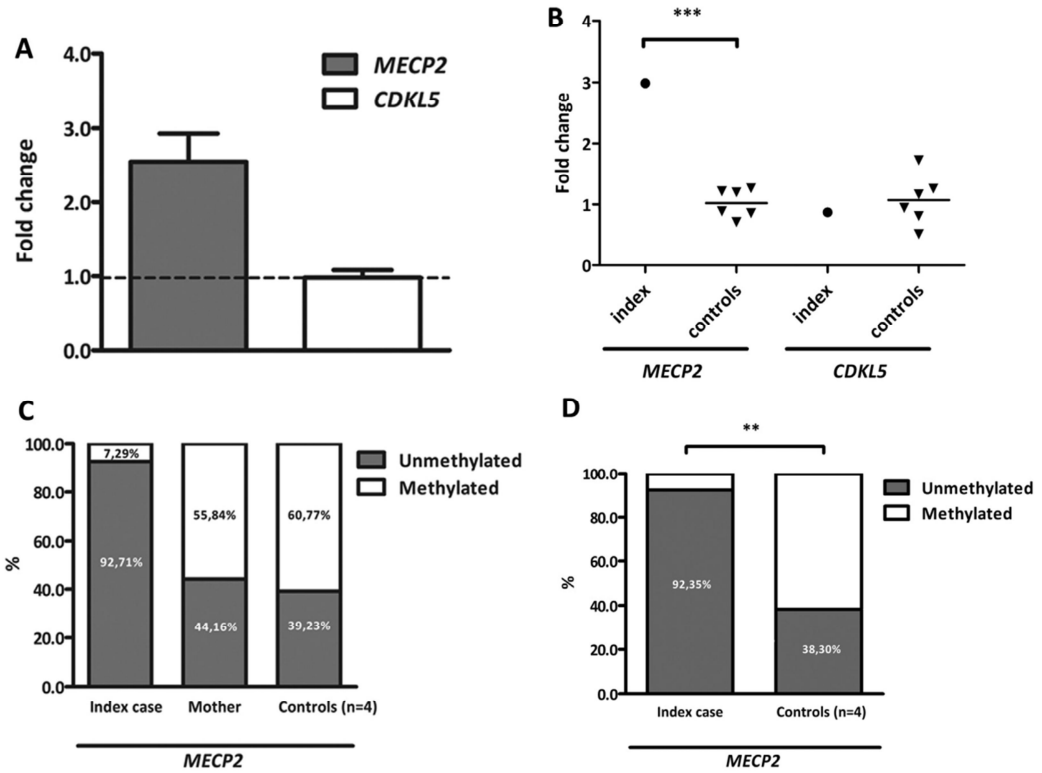


Fig. 2. *MECP2* and *CDKL5* mRNA expression levels and methylation status of MECP2. (A) qRT-PCR analysis revealed an upregulation of *MECP2* mRNA levels (mean fold change for the three replicates: 2.55 ± 0.38) in the lymphocytes of the index case in comparison to a group of control female individuals ($n = 4$) (the *HPRT* gene was used as a housekeeping gene); (B) Statistical analysis for MECP2 and CDKL5 mRNA expression between the case study ($n = 1$) and controls ($n = 6$) is presented, with the respective t-test analysis for replicate one ($p = 0.000532$; *** $p < 0.001$). This difference was significant for the three replicates. (C) Analysis of the DNA methylation status of the part of the CpG island 106 (ChrX: 153362568–153363473, hg19 coordinates) in the index patient, mother and controls ($n = 4$) using the EpiTect Methyl II PCR Primer Assay for Human *MECP2* (mean values for the three replicates). (D) Statistical analysis (*t*-test) for percentage of unmethylated DNA for replicate one where the case study ($n = 1$) is compared with the controls ($n = 4$) ($p = 0.002152$; **$p < 0.05$). This difference was significant for the three replicates.

these criteria (Amir et al., 1999; Hoffbuhr et al., 2001; Philippe et al., 2006).

Although our patient does not meet the diagnostic criteria for classic RTT (because she fulfills only two of the four major criteria), her development was normal in the first months of life and later a severe retardation emerged, she has acquired microcephaly, neurologic signs in lower limbs, cold and pale feet, scoliosis, bruxism and, importantly, a persistent and prolonged history of RTT-like hand stereotypies, which are among the most striking clinical features of RTT. With this combination of signs and symptoms she also does not strictly meet diagnostic criteria for variant RTT as published in 2001 (Hagberg et al., 2002) or, more recently, in 2010 (Neul et al., 2010); she does, however, meet the criteria for variant RTT according to the early Hagberg classification (Hagberg and Skjeldal, 1994). The main difference between later and former diagnostic criteria is that the 2001 and 2010 criteria emphasize regression (which is questionable in this case) as a key diagnostic feature of RTT. Therefore, and given this phenotypic resemblance, we were encouraged to study the *MECP2* gene.

Although the *MECP2* gene itself was not mutated nor involved in the chromosomal rearrangement, its expression was altered, and this deregulation might underlie the neurodevelopmental phenotype observed. This phenotype differs from that of male patients with *MECP2* duplications (Van Esch, 2012). Heterozygous females with *MECP2* duplication usually show extreme to complete skewing of X-chromosome inactivation and are asymptomatic or have mainly neuropsychiatric disorders, but severe mental retardation has been reported (Bijlsma et al., 2012). We were not able to determine the mechanism through which the pericentric inversion is affecting the methylation and the transcriptional regulation of the *MECP2* gene. aCGH analysis, provided no evidence for deletions or duplications encompassing other genes described as mutated in RTT-like phenotypes or a known or suspected regulatory element that could be related with *MECP2* altered expression. Taking these findings in consideration, we hypothesize that the structural alteration of the X chromosome, even without significantly large CNVs, could be disrupting chromosomal domain organization and long-range genomic interactions (Maksimenko and Georgiev, 2014; Robyr et al., 2011), thus influencing *MECP2* epigenetic modifications, including DNA methylation, and consequently modifying its expression. This type of perturbation has been described in fibroblasts of patients with Down syndrome (Letourneau et al., 2014).

In summary, we report the case of a girl with a Rett syndrome-like clinical presentation with a pericentric inversion in the X chromosome 46,X,inv(X)(p22.1q28) that doesnít disrupt the *MECP2* gene but is accompanied by epigenetic changes in the form of decreased DNA methylation levels within the 5′ regulatory region of *MECP2* gene, leading to a significantly increased expression of the gene. We speculate that this de-regulation of *MECP2* expression may be a novel mechanism underlying disease in other RTT-like patients who test negative for *MECP2* mutations, a hypothesis that should be investigated in the near future.

Acknowledgments

We thank the patient and family for their participation, all the members of the Microarray Core Facility, Cancer Center Amsterdam at Vrije Universiteit Medical Center for technical help with the array design and CGC Genetics/Centro de Genética Clínica S.A. for karyotype and FISH analysis.

References

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., Zoghbi, H.Y., 1999. Rett syndrome is caused by mutations in X-linked MECP2: encoding methyl-CpG-binding protein 2. Nat. Genet. 23, 185–188.

Ariani, F., Mari, F., Pescucci, C., Longo, I., Bruttini, M., Meloni, I., Hayek, G., Rocchi, R., Zappella, M., Renieri, A., 2004. Real-time quantitative PCR as a routine method for screening large rearrangements in Rett syndrome: report of one case of MECP2 deletion and one case of MECP2 duplication. Hum. Mutat. 24, 172–177.

Bijlsma, E.K., Collins, A., Papa, F.T., Tejada, M.I., Wheeler, P., Peeters, E.a.J., Gijsbers, A.C.J., van de Kamp, J.M., Kriek, M., Losekoot, M., Broekma, A.J., Crolla, J.A., Pollazzon, M., Mucciolo, M., Katzaki, E., Disciglio, V., Ferreri, M.I., Marozza, A., Mencarelli, M.A., Castagnini, C., Dosa, L., Ariani, F., Mari, F., Canitano, R., Hayek, G., Botella, M.P., Gener, B., Mínguez, M., Renieri, A., Ruivenkamp, C.a.L., 2012. Xq28 duplications including MECP2 in five females: expanding the phenotype to severe mental retardation. Eur. J. Med. Genet. 55, 404–413.

Buffart, T.E., Israeli, D., Tijssen, M., Vosse, S.J., Mrsić, A., Meijer, G.A., Ylstra, B., 2008. Across array comparative genomic hybridization: a strategy to reduce reference channel hybridizations. Genes Chromosomes Cancer 47, 994–1004.

Collins, A.L., Levenson, J.M., Vilaythong, A.P., Richman, R., Armstrong, D.L., Noebels, J.L., David Sweatt, J., Zoghbi, H.Y., 2004. Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. Hum. Mol. Genet. 13, 2679–2689.

Hagberg, B., Aicardi, J., Dias, K., Ramos, O., 1983. A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. Ann. Neurol. 14, 471–479.

Hagberg, B.A., Skjeldal, O.H., 1994. Rett variants: a suggested model for inclusion criteria. Pediatr. Neurol. 11, 5–11.

Hagberg, B., Hanefeld, F., Percy, A., Skjeldal, O., 2002. An update on clinically applicable diagnostic criteria in Rett syndrome Comments to Rett Syndrome Clinical Criteria Consensus Panel Satellite to European Paediatric Neurology Society Meeting, Baden Baden, Germany, 11 September 2001. Eur. J. Paediatr. Neurol. EJPN Off. J. Eur. Paediatr. Neurol. Soc. 6, 293–297.

Hoebeeck, J., van der Luijt, R., Poppe, B., De Smet, E., Yigit, N., Claes, K., Zewald, R., de, J., ong, G.-J., De Paepe, A., Speleman, F., Vandesompele, J., 2005. Rapid detection of VHL exon deletions using real-time quantitative PCR. Lab. Investig. J. Tech. Methods Pathol. 85, 24–33.

Hoffbuhr, K., Devaney, J.M., LaFleur, B., Sirianni, N., Scacheri, C., Giron, J., Schuette, J., Innis, J., Marino, M., Philippart, M., Narayanan, V., Umansky, R., Kronn, D., Hoffman, E.P., Naidu, S., 2001. MeCP2 mutations in children with and without the phenotype of Rett syndrome. Neurology 56, 1486–1495.

Jovanovic, L., Delahunt, B., McIver, B., Eberhardt, N.L., Grebe, S.K.G., 2003. Optimising restriction enzyme cleavage of DNA derived from archival histopathological samples: an improved HUMARA assay. Pathology (Phila.) 35, 70–74.

Letourneau, A., Santoni, F.A., Bonilla, X., Sailani, M.R., Gonzalez, D., Kind, J., Chevalier, C., Thurman, R., Sandstrom, R.S., Hibaoui, Y., Garieri, M., Popadin, K., Falconnet, E., Gagnebin, M., Gehrig, C., Vannier, A., Guipponi, M., Farinelli, L., Robyr, D., Migliavacca, E., Borel, C., Deutsch, S., Feki, A., Stamatoyannopoulos, J.A., Herault, Y., van Steensel, B., Guigo, R., Antonarakis, S.E., 2014. Domains of genome-wide gene expression dysregulation in Down's syndrome. Nature 508, 345–350.

Makrythanasis, P., Moix, I., Gimelli, S., Fluss, J., Aliferis, K., Antonarakis, S.E., Morris, M.A., Béna, F., Bottani, A., 2010. De novo duplication of MECP2 in a girl with mental retardation and no obvious dysmorphic features. Clin. Genet. 78, 175–180.

Maksimenko, O., Georgiev, P., 2014. Mechanisms and proteins involved in long-distance interactions. Front. Genet. 5, 28.

Neul, J.L., Kaufmann, W.E., Glaze, D.G., Christodoulou, J., Clarke, A.J., Bahi-Buisson, N., Leonard, H., Bailey, M.E.S., Schanen, N.C., Zappella, M., Renieri, A., Huppke, P., Percy, A.K., RettSearch, C., onsortium, 2010. Rett syndrome: revised diagnostic criteria and nomenclature. Ann. Neurol. 68, 944–950.

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29, e45.

Philippe, C., Villard, L., De Roux, N., Raynaud, M., Bonnefond, J.P., Pasquier, L., Lesca, G., Mancini, J., Jonveaux, P., Moncla, A., Chelly, J., Bienvenu, T., 2006. Spectrum and distribution of MECP2 mutations in 424 Rett syndrome patients: a molecular update. Eur. J. Med. Genet. 49, 9–18.

Ramocki, M.B., Peters, S.U., Tavyev, Y.J., Zhang, F., Carvalho, C.M.B., Schaaf, C.P., Richman, R., Fang, P., Glaze, D.G., Lupski, J.R., Zoghbi, H.Y., 2009. Autism and other neuropsychiatric symptoms are prevalent in individuals with MeCP2 duplication syndrome. Ann. Neurol. 66, 771–782.

Rett, A., 1966. On a unusual brain atrophy syndrome in hyperammonemia in childhood. Wien. Med. Wochenschr. 1946 (116), 723–726.

Robyr, D., Friedli, M., Gehrig, C., Arcangeli, M., Marin, M., Guipponi, M., Farinelli, L., Barde, I., Verp, S., Trono, D., Antonarakis, S.E., 2011. Chromosome conformation capture uncovers potential genome-wide interactions between human conserved non-coding sequences. PloS One 6, e17634.

Van Esch, H., 2012. MECP2 duplication syndrome. Mol. Syndromol. 2, 128–136.

Van Esch, H., Bauters, M., Ignatius, J., Jansen, M., Raynaud, M., Hollanders, K., Lugtenberg, D., Bienvenu, T., Jensen, L.R., Gecz, J., Moraine, C., Marynen, P., Fryns, J.-P., Froyen, G., 2005. Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. Am. J. Hum. Genet. 77, 442–453.

CHAPTER 3

Whole exome sequencing as a tool for identifying disease causing genes

Identification of novel genetic causes of Rett syndrome-like phenotypes

JMG Online First, published on January 6, 2016 as 10.1136/jmedgenet-2015-103568

ORIGINAL ARTICLE

Identification of novel genetic causes of Rett syndrome-*like* phenotypes

Fátima Lopes,[1,2] Mafalda Barbosa,[3,4] Adam Ameur,[5] Gabriela Soares,[6] Joaquim de Sá,[7] Ana Isabel Dias,[8] Guiomar Oliveira,[9,10] Pedro Cabral,[11] Teresa Temudo,[12] Eulália Calado,[8] Isabel Fineza Cruz,[13] José Pedro Vieira,[8] Renata Oliveira,[7] Sofia Esteves,[1,2] Sascha Sauer,[14,15] Inger Jonasson,[5] Ann-Christine Syvänen,[16] Ulf Gyllensten,[5] Dalila Pinto,[3] Patrícia Maciel[1,2]

▸ Additional material is published online only. To view this file please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2015-103568)

For numbered affiliations see end of article.

Correspondence to
Professor Patricia Maciel, Life and Health Sciences Research Institute—School of Health Sciences, University of Minho, Campus de Gualtar, Braga 4710-057, Portugal; pmaciel@ecsaude.uminho.pt

FL and MB contributed equally.

Received 6 October 2015
Revised 17 November 2015
Accepted 6 December 2015

ABSTRACT

Background The aim of this work was to identify new genetic causes of Rett-like phenotypes using array comparative genomic hybridisation and a whole exome sequencing approach.

Methods and results We studied a cohort of 19 Portuguese patients (16 girls, 3 boys) with a clinical presentation significantly overlapping Rett syndrome (RTT). Genetic analysis included filtering of the single nucleotide variants and indels with preference for de novo, homozygous/compound heterozygous, or maternally inherited X linked variants. Examination by MRI and muscle biopsies was also performed. Pathogenic genomic imbalances were found in two patients (10.5%): an 18q21.2 deletion encompassing four exons of the *TCF4* gene and a mosaic UPD of chromosome 3. Variants in genes previously implicated in neurodevelopmental disorders (NDD) were identified in six patients (32%): de novo variants in *EEF1A2*, *STXBP1* and *ZNF238* were found in three patients, maternally inherited X linked variants in *SLC35A2*, *ZFX* and *SHROOM4* were detected in two male patients and one homozygous variant in *EIF2B2* was detected in one patient. Variants were also detected in five novel NDD candidate genes (26%): we identified de novo variants in the *RHOBTB2*, *SMARCA1* and *GABBR2* genes; a homozygous variant in *EIF4G1*; compound heterozygous variant in *HTT*.

Conclusions Network analysis reveals that these genes interact by means of protein interactions with each other and with the known RTT genes. These findings expand the phenotypical spectrum of previously known NDD genes to encompass RTT-like clinical presentations and identify new candidate genes for RTT-like phenotypes.

INTRODUCTION

Rett syndrome (RTT) is a severe neurodevelopmental disorder (NDD) affecting mostly girls, characterised by an apparently normal prenatal and perinatal period followed by a stagnation in development and a severe regression in language and motor skills.[1] RTT is clinically divided into classical and atypical forms of the disease.[2] The clinical diagnostic criteria for RTT can be revisited in table 1. Patients with RTT or RTT-like clinical presentation often present with severe intellectual disability (ID), autistic features and epilepsy, and their differential diagnosis includes Angelman syndrome, Pitt-Hopkins syndrome (PTHS) and some epileptic encephalopathies.[3–5]

Whole exome sequencing (WES), has had a major impact in medical practice, leading to the identification of several new genes involved in ID.[6–8] We used a genomic approach combining array comparative genomic hybridisation (aCGH) and WES to find genetic causes of disease in a group of RTT-like patients who tested negative for *MECP2* mutations and—whenever clinically appropriate—*CDKL5* mutations. We were able to detect pathogenic variants and very likely pathogenic variants that we believe can account for the RTT-like phenotype in 13 (68%) of these patients.

METHODS

Patients

We enrolled 19 patients (16 girls and 3 boys) with idiopathic neurodevelopmental phenotypes that clinically overlap with RTT and their unaffected parents (trios). The patients were selected from a previously established database of patients with idiopathic ID and confirmed as eligible by consulting with medical geneticists, paediatric neurologists and neurodevelopmental paediatricians, using the revised clinical criteria for RTT diagnosis.[1] We included patients meeting sufficient criteria for the diagnosis of Rett (classical or atypical)—except for documented regression, which was not considered mandatory. Exclusion criteria were also taken into account (tables 1 and 2; online supplementary data1).

Before enrolment all patients had undergone routine diagnostic workup, including brain MRI and metabolic screen. *MECP2* analysis was performed by Sanger sequencing and qPCR for all patients and *CDKL5* sequencing was undertaken for patients presenting early onset seizure variant. No patient presented with clearly congenital forms, hence *FOXG1* was not tested. Patients would only be enrolled in the study if their complementary exams had been normal or with abnormalities that could not clearly explain the phenotype.

Molecular analysis

For all patients included in this work an aCGH analysis was performed first, followed by WES (provided aCGH profile had been normal or inconclusive).

To cite: Lopes F, Barbosa M, Ameur A, *et al*. *J Med Genet* Published Online First: [*please include* Day Month Year] doi:10.1136/jmedgenet-2015-103568

Table 1 Clinical diagnostic criteria for Rett syndrome (adapted from Neul and colleagues[1])

Main criteria	Supportive criteria	Exclusion criteria	Required for classic RTT	Required for variant RTT
Partial/complete loss of acquired purposeful hand skills	Breathing disturbances (awake)	Brain injury secondary to trauma (perinatal or postnatal)	A period of regression followed by recovery or stabilisation	A period of regression followed by recovery or stabilisation*
Partial/complete loss of acquired spoken language	Bruxism (awake)	Neurometabolic disease	All main criteria and none exclusion criteria	At least 2 of the 4 main criteria
Gait abnormalities: Impaired (dyspraxic) or absent	Impaired sleep pattern	Severe infection that causes neurological problems	Supportive criteria are not required, although often present in typical RTT	At least 5 out of 11 supportive criteria
Stereotypical hand movements (wringing/squeezing, clapping/tapping, mouthing, washing/rubbing automatisms)	Abnormal muscle tone	Grossly abnormal psychomotor development in the first 6 months of life		
	Peripheral vasomotor disturbances			
	Scoliosis/kyphosis			
	Growth retardation			
	Small cold hands and feet			
	Inappropriate laughing/screaming spells			
	Diminished response to pain			
	Intense eye communication —'eye pointing'			

*Because *MECP2* mutations are now identified in some individuals prior to any clear evidence of regression, the diagnosis of 'possible' RTT should be given to children under 3 years old who have not lost any skills but otherwise have clinical features suggestive of RTT. These individuals should be reassessed every 6–12 months for evidence of regression. If regression manifests, the diagnosis should then be changed to definite RTT. However, if the child does not show any evidence of regression by 5 years, the diagnosis of RTT should be questioned.

Array comparative genomic hybridisation

aCGH analysis was performed using two different platforms: human genome CGH Agilent 180K custom array and Illumina HumanOmniExpress beadchip array (see online supplementary part 2; figure S2.1). All genomic coordinates are in build GRCh37/hg19.

Exome sequencing and variant detection

Exomes were enriched with Agilent's SureSelect All Human Exome V.4 Kit (51 Mb encompassing the exons of 20 965 genes), followed by AB SOLiD5500xl System sequencing (Life Technologies). Filtering of single nucleotide variants and indels is described in online supplementary data2. Preference was given to (1) de novo variants, (2) homozygous or compound heterozygous variants compatible with an autosomal recessive mode of transmission and (3) X linked variants. The impact of variants was predicted using in silico tools, namely SIFT,[9] PolyPhen2,[10] Mutation Assessor,[11] Mutation Taster,[12] PMut[13] and Condel.[14] Alignment for amino acid conservation among species was performed using the ClustalW2 webtool (http://www.ebi.ac.uk/Tools/msa/clustalw2/) (see online supplementary data2; figure S2.2).

Selection and interpretation of the variants

Candidate variants were validated by Sanger sequencing in the trios. Variants were selected for Sanger confirmation as described in online supplementary data2 and in figure S2.3. The variants selected for Sanger sequencing confirmation are described in online supplementary data1, table S1.19. The primers designed for this purpose are listed in online supplementary data2, table S2.2. The variants were classified according to the flow chart depicted in online supplementary figure S2.4, adapted from de Ligt and colleagues.[6]

Network analysis

We performed gene network analysis to: (1) verify if our candidate genes interacted among themselves and with the known RTT genes (*MECP2*, *CDKL5*, *FOXG1*), (2) study the topology of these interactions, (3) predict additional genes that may be involved in RTT if they are shown to interact with a large number of genes in the query set, (4) identify common biological themes by exploring functional enrichment analysis of Gene Ontology (GO) terms.

Network analysis was performed with GeneMANIA (V.3.1.2.7, http://www.genemania.org/).[15 16] Given a set of input genes, GeneMANIA finds related genes using a very large set of functional association data, including protein interactions, genetic interactions, pathway, coexpression, colocalisation, shared protein domain and predicted functional relationship. GeneMANIA also allows for functional enrichment analysis. For our analysis, the genes used as input were the already known RTT genes (*MECP, CDKL5, FOXG1*) as well as the genes selected as likely causing RTT-like phenotype in our cohort.

For additional details on the methodology of the gene network analysis performed using GeneMANIA (V.3.1.2.7) see online supplementary data2.[15 16]

RESULTS

Patients' clinical profile

We enrolled 19 patients (16 girls and 3 boys) with ages between 6 and 31 years (mean age 15.8±6.3 years), with idiopathic neurodevelopmental phenotypes that clinically overlapped with RTT, as well as their unaffected parents (trios). The patients were selected from a previously established database of patients with ID and confirmed as eligible using the revised clinical criteria for RTT diagnosis.[1] All patients had normal routine diagnostic workup, including brain MRI, which for RTT diagnosis purpose was classified as 'normal' if the alterations present were not a consequence of a perinatal or postnatal insult, neurometabolic disease or severe infection; this was the case for patients 1, 2, 6, 14, 15 and 17. Detailed description of the MRI findings for these patients can be found in the online supplementary data. Metabolic screen, *MECP2* analysis by Sanger sequencing

Table 2 Clinical characterisation of the cohort relative to RTT clinical diagnosis criteria

				Main criteria				Minor criteria												Comorbidities	
Proband_ID	Gender	Rett	Regression	Partial/complete loss of acquired hand skills	Partial/complete loss of spoken language	Gait abormalities	Stereotypical hand movements	Breathing disturbances	Bruxism when awake	Impaired sleep pattern	Abnormal muscle tone	Peripheral vasomotor disturbances	Scoliosis / Kyphosis	Growth retardation	Small cold hands/ feet	Laughing / Screaming spells	Diminished response to pain	Intense eye communi-cation	Epilepsy	Autism Spectrum Disorder
1	F	Classical	Y	Y	Y	Y	Y	Y	N	N	Y	N	Y	Y	Y	Y		Y	Y	Y
2	F	Classical	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y
3	F	Classical	Y	Y	Y	Y	Y	Y	N	N	Y	N	N	N	N	Y	N	N	Y	Y
4	F	Classical	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	N	Y	Y	N	Y	Y	Y
5	F	Atypical	Y	N	Y	Y	Y	Y	Y	N	Y	N	N	N	Y	Y	N	Y	N	N
6	F	Atypical	Y	Y	Y	Y	N	N	N	N	Y	N	Y	Y	Y	N	N	Y	N	Y
7	F	Atypical	Y	Y	N	N	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y
8	F	Atypical	Y	Y	Y	N	Y	N	N	N	Y	Y	Y	N	Y	Y	N	N	Y	Y
9	F	Atypical	N	Y	Y	N	Y	Y	Y	Y	Y	N	N	N	Y	Y	N	Y	N	Y
10	F	Like	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N
11	F	Like	N	N	Y	Y	Y	N	Y	Y	N	Y	N	Y	Y	N	Y	Y	Y	N
12	F	Like	N	N	Y	Y	Y	Y	Y	Y	N	Y	Y	N	Y	N	Y	Y	Y	N
13	F	Like	N	Y	Y	Y	N	N	Y	Y	Y	N	Y	N	N	Y	N	N	Y	N
14	M	Like	N	N	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N	Y	N	Y	Y
15	M	Like	N	Y	Y	Y	Y	N	N	Y	Y	N	N	N	N	Y	N	N	Y	Y
16	F	Like	N	N	Y	Y	Y	N		N		N	N		N	Y		N	Y	Y
17	F	Like	N	N	Y	N	Y	Y	Y	Y	Y	N	Y	Y	N	Y	Y	N	Y	Y
18	F	Like	N	N	N	Y	Y	N	Y	Y	N	Y	N	N	Y	N	Y	N	N	Y
19	M	Like	N	N	Y	Y	N	N	Y	N	N	Y	Y	N	Y	N	Y	Y	N	Y

Blank cells=information not provided.
F, Female; ID, intellectual disability; M, Male; N, No, Y, Yes.

and qPCR, as well as *CDKL5* sequencing (if early onset seizure variant) were performed for all patients.

The main clinical findings of the patients are summarised in table 2 (more detailed in online supplementary data1). We have classified the patients according to Neul and colleagues, as: (1) classical RTT, if they had had regression and presented all 4 main criteria; (2) atypical RTT, if they had had regression and met at least 2 of the 4 main criteria and 5 of the 11 minor criteria, and (3) RTT-like, if they met criteria for RTT (classical or atypical) except for documented regression. The most common comorbidities in this cohort were epilepsy, affecting 74% (14 of 19), and autism spectrum disorder (ASD) in 74% (14 of 19) of the patients. All patients were sporadic cases and no other relevant findings were reported in their family history, except where specified below. The clinical features of the cohort are summarised in table 2.

Global yield of genomic analysis

Using the combined aCGH and WES analysis we were able to detect genomic imbalances in 10.5% (2 out of 19) of the patients and single nucleotide variants in 58% (11 out of 19) patients (summary of the results in online supplementary table S1.1 and interpretation workflow in online supplementary figure S2.1). We detected a de novo 18q21.1 microdeletion encompassing the *TCF4* gene in patient 7, which was confirmed by qPCR, and a mosaic UPD of chromosome 3 in patient 16 (see online supplementary figures S1.1 and S1.2, respectively). For patients without diagnostic findings in the aCGH analysis, we performed WES analysis (see online supplementary figure S2.2–S2.4). A summary of the WES results is provided in tables 3 and 4: six variants in six genes previously described as associated with NDDs (table 3); likely pathogenic variants in five genes not described as associated with a RTT-like or ID related phenotype but which, due to their functions, may account for the disease in the patients (table 4).

Patients with CNVs causing RTT-like phenotypes

Patient 7 is a girl who had an apparently normal development up until the age of 4 months, when regression was noticed. Hand stereotypies were documented around 30 months; still, acquisition of some hand skills occurred around 6 years of age, but these were lost 3 years later. Currently the patient is an adult with moderate ID, epilepsy and autistic behaviour. Additionally, she presents with eight minor criteria for RTT diagnosis, including respiratory disturbances—more precisely hyperventilation. Overall, the patient was classified as atypical RTT. Microarrays revealed a de novo 0, 25 Mb microdeletion at chromosome 18q21.1 encompassing four exons of *TCF4* and the *MIR4529* gene (see online supplementary figure S1.1). Loss of function mutations and microdeletions affecting *TCF4* have been described in patients with PTHS.[19 20 34] The natural history of RTT and PTHS overlaps significantly, the latter being usually considered in the differential diagnosis of RTT. The fact that the patient is a girl, lacks dysmorphisms, started to show stereotypies around 2.5 years old and hyperpnoea at 7 years lead to the consideration of RTT as a first possibility.

Patient 16 is a 9 year-old girl who showed developmental stagnation at around 6 months, which coincided with appearance of West syndrome and deceleration of head growth. She has slowly acquired motor skills, with some purposeful grasp and ataxic gait. Severe ID and hand wringing raised the diagnostic hypothesis of RTT. aCGH analysis revealed an entire chromosome 3 with log R ratio (LRR)=0 and B Allele Frequency (BAF) split (0.3 and 0.6), compatible with mosaic UPD of chromosome 3 (see online supplementary figure S1.2). This abnormality occurred de novo and is predicted to be present in about 75% of the cells of the patient.[35] Only three of all the genes in chromosome 3 are predicted to be differently expressed according to parental lineage: *ALDH1L1* and *ZIC1* are paternally imprinted, and *HES1* is maternally imprinted.[36] When considering the possibility of a variant in heterozygosity in the mother/father being present in homozygosity in the child's cells with UPD, only a maternal missense variant in *SRGAP3* found in the WES analysis fits this hypothesis. Interestingly, *SRGAP3* encodes a Slit-Robo Rho GTPase activating protein that has been implicated in the pathogenesis of ID.[37 38]

Patients with variants in genes previously associated with similar phenotype

Patient 5 is a 5 year-old girl who regressed at around 8 months of life. Though the child presents with severe ID and no language, some psychomotor developmental milestones were attained, with tiptoe walking around 2 years of age. Hand stereotypies, intense eye communication, breathing disturbances (apnoea followed by hyperpnoea) and screaming spells prompted the clinical diagnosis of atypical RTT. Though severely microcephalic, the patient's brain MRI did not reveal any relevant morphological changes. Also, the patient does not have short stature. In this patient, a de novo c.C556T, p.(R186X) variant was detected in the *ZNF238* gene (see online supplementary figure S1.5).

Patient 17 is a 6 year-old girl who presented with seizures at 1 month of life and whose development was significantly delayed, with first words around 3 years and walking at 4 years of age. The patient also has hand stereotypies, bruxism and crying spells when awake, sleep problems, hyperpnoeas and apnoeas, and poor eye contact. In this patient a de novo c.

Table 3 List of patients with variants found in genes previously associated with neurodevelopmental phenotype

Proband	Gene	Location	NM number	cDNA	Protein	Related phenotype	Reference
5	*ZNF238*	chr1:244217659	NM_006352	c.C556T	p.(R186X)	ID; 1q43q44 deletion syndrome	17, 18
7	*TCF4*	chr18:52996207–53243605	Microdeletion encompassing 4 exons			Pitt-Hopkins syndrome	19, 20
8	*EIF2B2*	chr14:75470349	NM_014239	c.C380T	p.(A127V)	Leucoencephalopathy with vanishing white matter	21
14	*STXBP1*	chr9:130425592	NM_001032221	c.T538C	p.(C180R)	Early infantile epileptic encephalopathy	22, 23
15	*SLC35A2*	chrX:48762414	NM_001042498	c.G772A	p.(V258M)	Early onset epileptic encephalopathy; Congenital disorder of glycosylation type II	23, 24
17	*EEF1A2*	chr20:62127259	NM_001958	c.G274A	p.(A92T)	ID and epilepsy	6, 25
19	*SHROOM4*	chrX:50378637	NM_020717	c.C436T	p.(R146W)	Stocco dos Santos syndrome	26

ID, intellectual disability.

Table 4 List of patients with variants in candidate disease genes

Proband	Gene	Variant: genomic coordinates	NM number	cDNA	Protein	Functional impact prediction						Gene Function	KO/KD phenotype*	Expression in Human Neocortex†	References
						SIFT	PolyPhen2	MutAsse.	Condel	Pmut	MutTast				
2	*HTT*	chr4:3133374 chr4:3162034	NM_002111	c.C2108T c.C3779T	p.(P703L) p.(T1260M)	NP NP	P P	P NP	P NP	NP NP	P P	Ubiquitously expressed nuclear protein that regulates transcription; involved in vesicular traffic	Conditional mutants are small with progressive neurodegeneration	Moderate	27, 28
4	*SMARCA1*	chrX:128599594	NM_139035	c.G2897T	p.(G966V)	P	P	P	P	P	P	Chromatin remodelling; Wnt signalling; Interacts with *FOXG1*	Hemizygous male/ homozygous female KO show abnormal neuron proliferation and differentiation, increased brain and heart weight	Moderate	29, 30
9	*GABBR2*	chr9:101133817	NM_005458	c.G1699A	p.(A567T)	P	NP	P	NP	NP	P	γ-aminobutyric acid (GABA) type B receptor; Regulation of neurotransmitter release	Homozygous KO mice show clonic seizures, hyperactivity, anxiety.	High	31
11	*RHOBTB2*	chr8:22865220	NM_001160036	c.A1528G	p.(N510D)	P	P	NP	P	NP	NP	Rho GTPase family; Binds to CUL3	ND	High	32
11	*EIF4G1*	chr3:184038482	NM_182917	c.G602A	p.(R201H)	P	P	NP	P	NP	P	Recruitment of mRNA to the 40S ribosomal subunit	ND	High	33

*The Jackson laboratory, 2014.
†Allen Institute for Brain Science, 2004.
KO, knockout; KD, knockdown; ND, not described.

G274A, p.(A92T) substitution in the *EEF1A2* gene was found (see online supplementary figure S1.15).

Patient 8 is a 13 year-old girl and the first child of a healthy and non-consanguineous couple. She seemed to have a normal development up until 24 months when regression was noticed in language and hand manipulation, followed by the first signs of autistic behaviour-like impairment in social interactions around 30 months. Concomitant hand stereotypies, ataxia, rigidity and screaming spells led to the classification of the patient as RTT-like. The patient also presented continuous tremor but no spasticity or optic atrophy. Brain MRI performed when the patient was 3 years old did not reveal significant alterations. In this patient a homozygous recessive variant c.C380T, p.(A127V) in the *EIF2B2* gene was found (see online supplementary figure S1.7).

Patient 14 is a 12 year-old boy who presented with West syndrome at 1 month of age and severe developmental delay, in addition to poor eye contact, hyperventilation episodes, bruxism when awake, decreased response to pain and midline hand stereotypies. This RTT-like patient has a de novo c.T538C, p.(C180M) variant in the *STXBP1* gene (see online supplementary figure S1.13). A missense mutation in amino acid 180 was described in 2008.[39] Although variable,[40] the core phenotype of *STXBP1* mutations seems to include epilepsy within the 1st months of life.[41] Mutations in *STXBP1* can also cause ID without epilepsy[22] and it may actually be a relatively common cause of severe ID,[42] which highlights the role of *S1XBP1* in cognitive function alone. Recently, patients with *STXBP1* mutations were noted to have head and upper limb stereotypies (eg, hand flapping)[43] but midline hand stereotypies are reported for the first time in our study. While our manuscript was in preparation a de novo missense variant in *STXBP1* was reported in a girl with classical Rett diagnosis.[44]

Patient 15 is an 8 year-old boy whose development stagnated around 6 months; his psychomotor development was significantly delayed, language and gait never being attained. The patient also had an abnormal sleep pattern, inconsolable crying spells and hand and head stereotypies, as well as autism features; he was classified as RTT-like. WES revealed a c.G772A, p.(V258M) variant in the *SLC35A2* gene (see online supplementary figure S1.14). The patient's mother also carried the variant and presented with a random X inactivation. The variant was also present in the patient's skin biopsy. Transferrin isoelectric focusing analysis in the patient was normal. Interestingly, upon reanalysis of the clinical data, we found that the patient also had facial dysmorphisms, gastro-oesophageal reflux, epileptic encephalopathy (but not West syndrome), microcephaly and brain malformations (namely brain atrophy, thin corpus callosum and frontoparietal periventricular heterotopies), resembling other patients with *SLC35A2* variants.[23 24]

Patient 19 is 14 year-old boy with dyspraxic gait and no language, who acquired purposeful grasp only around 2 years of age. Eye pointing, kyphosis, peripheral vasomotor disturbances and small cold hand and feet lead to the classification as RTT-like. This patient carries two maternally inherited X linked variants: a c.C436T, p.(R146W) variant in *SHROOM4* and a c.G409A, p.(D137N) variant in *ZFX* (see online supplementary figures S1.17 and S1.18). Extended pedigree analysis reveals that patient 19 is an isolated case and both variants are inherited from the healthy mother, who has a random X inactivation pattern. *SHROOM4* encodes a regulator of cytoskeletal architecture and has been associated with X linked ID.[26]

Other variants in genes previously associated with NDDs but which weren't in accordance with the inheritance patterns described in the literature were also found in some cases (see online supplementary data1, tables S1.2–S1.18).

Patients with variants in genes possibly relevant for ID pathogenesis

For five of the patients enrolled in the study, likely pathogenic variants were found in functionally relevant and/or candidate ID genes.

Patient 2 is a 18 year-old girl who showed developmental regression around 6 months of life and 2 months later started to have partial complex seizures as well as lack of interest in interacting with the environment. This classical RTT patient meets all four main criteria in addition to eight supportive criteria. On neurological exam it was also observed that the patient had swallowing problems, dystonia and bradykinesia (but not rigidity) in addition to continuous manual stereotypies (but not chorea). Interestingly, brain MRI performed when the patient was approximately 5 years old showed significant striatum atrophy (especially in the caudate nuclei) as well as mild atrophy of the cortex and cerebellar vermis. WES revealed two compound heterozygous variants in the *HTT* gene (see online supplementary figure S1.3): a maternal c.C2108T, p.(P703L) and a paternal c.C3779T, p.(T1260M). The latter variant is described in single nucleotide polymorphism database (dbSNP) as a polymorphism (rs34315806) with a minor allele frequency of T=0.0276/138.

Patient 4 is a girl who has been classified as classical RTT. She had moderate developmental delay, superimposed by regression at around 5 years of age. Currently, at 25 years of age, the patient is severely autistic, non-verbal, can't walk, has lost purposeful hand use and has hand sterotypies, in addition to seven minor criteria (see details in table 2). She carries a de novo variant c.G2897T, p.(G966V) in *SMARCA1* (see online supplementary figure S1.4).

Patient 9 had developmental stagnation at 7 months followed by regression, currently (at age 19 years) presenting with severe ID. She was classified as an RTT variant in light of her absence of language, hand stereotypies and lack of hand use, breathing disturbances (hyperventilation), bruxism, abnormal sleep cycle, crying spells, autistic features, eye pointing and small feet. She has never had seizures. A de novo variant c.G1699A, p.(A567T) in the *GABBR2* gene was found (see online supplementary figure S1.8).

Patient 11 is a 6 year-old girl whose development stagnated at around 6–9 months, coinciding with the beginning of generalised epilepsy. Additional findings that lead to the classification of the patient as RTT-like include: hand stereotypies, intense eye communication, sleep problems, peripheral vasomotor disturbances, bruxism when awake, growth retardation, diminished response to pain and resting tremor. Her mother has resting tremor and is suspected of having psychiatric disease, possibly early onset dementia. The maternal grandmother is bedridden and demented. The father is also suspected of having psychiatric disease. The patient carries a de novo variant c.A1528G, p.(N510D) in *RHOBTB2* (see online supplementary figure S1.10). She also has a homozygous c.G602A, p.(R201H) variant in *EIF4G1* (see online supplementary figure S1.11).

Bioinformatic analysis of the interactions between the novel candidate genes and known RTT-like NDD associated genes

Phenotypical overlap between RTT and the patients in our cohort with variants in genes previously implicated in NDDs was observed (table 2 and figure 1). Network analysis using GeneMania revealed that our candidate genes interact with each other and with the already known RTT genes by means of

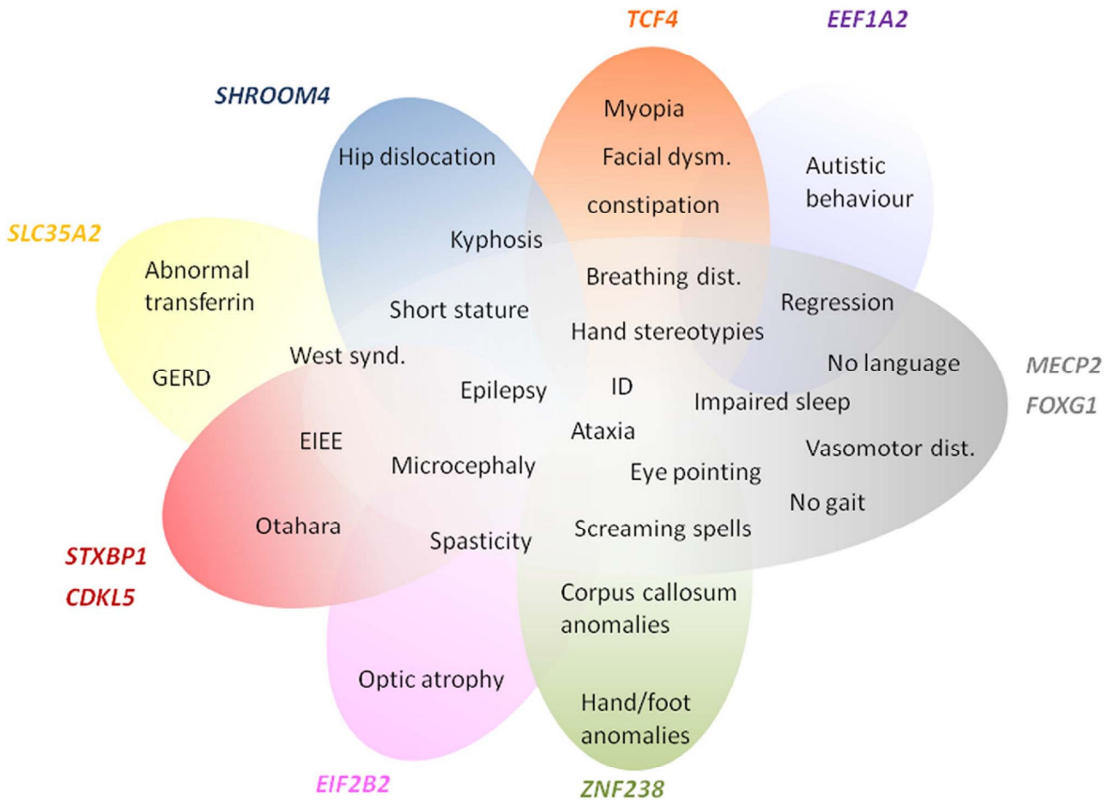


Figure 1 Scheme representing phenotype overlap of the genes already identified as causing neurodevelopmental disorders and the RTT phenotype. The phenotypes clearly blend, suggesting that the RTT spectrum may still be expanding.

protein (71%), predicted (15%) and genetic (3%) interactions, as well as coexpression (7%) and participating in a common pathway (2%) (figure 2).

DISCUSSION

In this study we identified a possible genetic cause of disease in eight RTT-like patients. Recently, two publications described the application of WES to smaller series of patients with features of RTT, where, if we exclude the variants found in *MECP2* and *FOXG1* genes, an yield of 27% was found.[45 46] However, important questions such as (1) what other genes may lead to an RTT-like similar phenotype and (2) which pathways and genetic mechanisms can lead to such a specific phenotype still remain unanswered. To try to clarify these questions we undertook genomic analysis by aCGH and WES in a group of 19 trios whose index presented NDD with RTT-like features achieving an yield of 37% (excluding a case of uniparental dissomy and variants found in candidate genes).

Regarding the WES results, in 6 of the 19 patients we detected variants in genes previously associated with overlapping neurodevelopmental phenotypes, namely: *SLC35A2*, *STXBP1*, *ZNF238*, *EEF1A2*, *EIF2B2* and *SHROOM4* (table 3). On the other hand, five patients present novel variants in genes with known function in the central nervous system and which, to the best of our knowledge, have not been clearly associated with ID in humans (*HTT*, *SMARCA1*, *GABBR2*, *RHOBTB2* and *EIF4G1*) (table 4). All these genes, taking into account their function, are good candidates to be disease-causing. *HTT* encodes a protein that directly interacts with MeCP2[27] and Mecp2-deficient mice have reduced expression of Htt in the entire brain, leading to a defect in the axonal transport of Bdnf.[28] RTT and Huntington disease seem to share features at a molecular level (nuclear interaction for transcriptional regulation and axonal trafficking through *BDNF*)[27] and neuropathological findings (striatum atrophy)[47] and clinical presentation (of compulsive movement disorder plus cognitive dysfunction).

Another of our candidates, *SMARCA1* (alias *SNF2L*), encodes a protein which was described to function antagonistically with Foxg1 in the regulation of brain size in mice.[29 48]

GABBR2 encodes a γ-aminobutyric acid type B receptor that is involved in neuronal activity inhibition through the regulation of neurotransmitter release.[31] Recently, de novo missense variants in *GABRR2* were identified in two different patients with infantile spasms.[49]

RHOBTB2 belongs to the Rho GTPases family and was found to bind to CUL3.[32] De novo nonsense variants in *CUL3* were identified in two separate next-generation sequencing studies using ASD (autism spectrum disorder) probands,[50 51] *EIF4G1* encodes a translation initiation factor involved in the recruitment of mRNA to the 40S ribosomal subunit.[33] Variants in *EIF4G1* have been associated with autosomal dominant forms of Lewy body dementia[52] and Parkinson's disease (with and without dementia) however the real impact of some of these variants is still unclear.[53–56] Interestingly patient 11 (homozygous) and her mother (heterozygous) present resting tremor, a typical sign of Parkinson's disease. Considering the

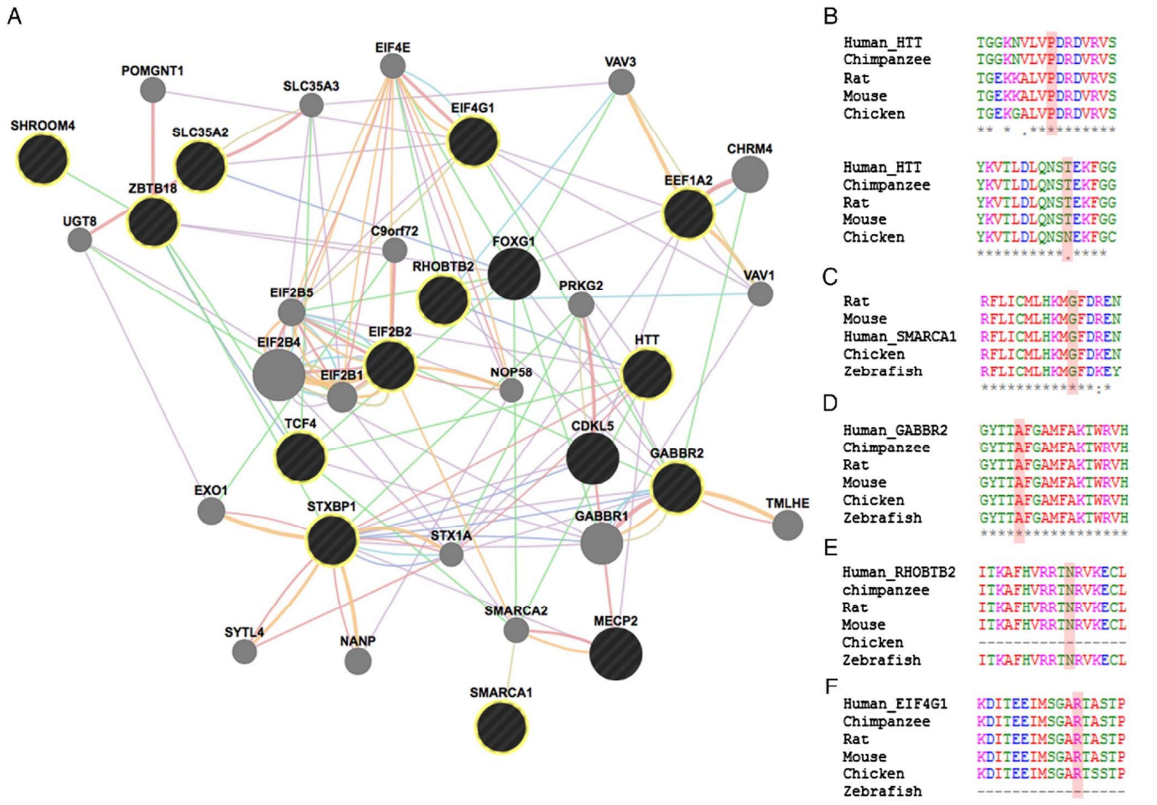


Figure 2 (A) Functional network showing the interaction among genes with variants in our cohort and genes previously associated with RTT. GeneMANIA retrieved interactions between the query genes (black circles—highlighted with yellow circumference if identified in our cohort) and added extra genes (grey circles) that are strongly connected to the query genes. Analysis was based on physical interactions (red edges), predicted interactions (orange edges), coexpression (purple edges), genetic interactions (green edges), pathway (light blue edges), colocalisation (dark blue edges) and shared protein domains (brown edges). (B) Species alignment for the changed animo acid for Htt p.(P703L) (above) and Htt p.(T1260M) (bellow), (C) Smarca1 p.(G966V), (D) Gabbr2 p.(A567T), (E) Rhobtb2 p.(N510D) and (F) Eif4g1 p.(R201H).

biological processes in which *RHOBTB2* and *EIF4G1* are involved, it is likely that either one of the variants or the combination of both in the patient might have contributed towards the development of the disease.

When analysing the phenotypes associated with genes previously implicated in NDDs, they overlap with those observed in our sample of RTT-like patients. In fact, there seems to be a spectrum of clinical presentation that allows for the delineation of a core phenotype as well as distinctive clinical features, that could help guide/interpret genetic testing in future patients, as summarised in figure 1. Furthermore, the gene-gene interaction analysis revealed that our candidate genes interact with each other and with the already known RTT genes mainly by means of protein interactions (71%), predicted functional relationships (15%) and coexpression (7%) (figure 2).[16] Functional enrichment analysis revealed that the top GO biological process terms were under the parent terms Translation (GO:0006412) and Glial cell differentiation (GO:0010001). Careful analysis of the network also allows for identification of possible additional candidates such as *GABBR1*, which has genetic interactions with *FOXG1*, physical interactions with *GABBR2* and coexpression with *STXBP1* and *TCF4*.

In this work, the identification of variants in genes that had already been associated with overlapping but still distinctive NDDs brings new insight into the differential diagnosis of RTT and might allow for the aetiological diagnosis of RTT-like patients. We point out seven novel candidate genes, which may be implicated in RTT-like clinical presentations. It is important to highlight that replication of these results in more patients is required for a proper genotype-phenotype correlation and the establishment of differences and similarities with RTT. Functional studies would also be of great value. In conclusion, we expanded the phenotypical spectrum of previously known NDD genes to encompass RTT-like clinical features, and suggest novel genes that might be associated with those. Although this group of disorders is genetically heterogeneous, the novel and previously identified genes converge in common pathways and only a better understanding of the pathophysiology of NDDs will allow for development of efficient targeted therapies in the future.

Author affiliations
[1]Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, Portugal
[2]ICVS/3B's—PT Government Associate Laboratory, Braga/Guimarães, Portugal
[3]Department of Genetics and Genomic Sciences, The Mindich Child Health & Development Institute, The Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, USA
[4]Instituto Gulbenkian de Ciência, Oeiras, Portugal
[5]Science for Life Laboratory Uppsala, Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden

[6]Center for Medical Genetics Dr. Jacinto Magalhães, Centro Hospitalar do Porto, Porto, Portugal
[7]Serviço de Genética Médica, Hospital Pediátrico, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal
[8]Serviço de Neurologia Pediátrica, Hospital D. Estefânia, Centro Hospitalar Lisboa Central, Lisbon, Portugal
[9]Unidade de Neurodesenvolvimento e Autismo, Centro de Desenvolvimento da Criança e Centro de Investigação e Formação Clínica, Hospital Pediátrico, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal
[10]Faculty of Medicine, University Clinic of Pediatrics and Institute for Biomedical Imaging and Life Science, University of Coimbra, Coimbra, Portugal
[11]Department of Neurology, Egas Moniz Hospital, Lisbon, Portugal
[12]Department of Neuropediatrics, Centro Hospitalar do Porto, Porto, Portugal
[13]Centro de Desenvolvimento Luís Borges, Hospital Pediátrico, Centro Hospitalar Universitário de Coimbra, Coimbra, Portugal
[14]Otto-Warburg-Laboratory, Max-Planck-Institute for Molecular Genetics, Berlin, Germany
[15]CU Systems Medicine, University of Wuerzburg, Wuerzburg, Germany
[16]Molecular Medicine and Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Uppsala, Sweden

Acknowledgements The authors thank the patients and families for participation in the genetic studies and for allowing this publication. SOLiD sequencing was performed at the Uppsala node of the National Genomics Infrastructure (NGI) Sweden and computational resources from Uppnex were used for the data analysis.

Contributors FL and MB: laboratory work, acquisition of data, analysis, interpretation of results and manuscript preparation; AA: bioinformatics and statistical analysis; GS: recruitment of patients, acquisition and analysis of clinical data and manuscript preparation; JdS, AID, GO, PC, TT, EC, IFC, JPV, RO: recruitment of patients and acquisition of clinical data; SE: in silico tools analysis; IJ, A-CS, UG: study supervision; DP: aCGH analysis; PM: study concept and design, data interpretation, critical revision of the manuscript for important intellectual content, study supervision.

Funding This work was supported by the Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 262055. This work was also supported by the FEDER through the Programa Operacional Factores de Competitividade—COMPETE and by Portuguese national funds through the FCT—Fundação para a Ciência e Tecnologia, grants number PIC/IC/83026/2007 and PIC/IC/83013/2007, PhD scholarship grant to MB number SFRH/BDINT/ 51549/2011 and PhD scholarship grant to FL number SFRH/BD/84650/2010.

Competing interests None declared.

Patient consent Obtained.

Ethics approval Hospital Geral de St. António, Portugal.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Neul JL, Kaufmann WE, Glaze DG, Christodoulou J, Clarke AJ, Bahi-Buisson N, Leonard H, Bailey MES, Schanen NC, Zappella M, Renieri A, Huppke P, Percy AK. Rett syndrome: revised diagnostic criteria and nomenclature. *Ann Neurol* 2010;68:944–50.
2 Hagberg B. Clinical manifestations and stages of Rett syndrome. *Ment Retard Dev Disabil Res Rev* 2002;8:61–5.
3 Depienne C, Gourfinkel-An I, Baulac S, LeGuern E. Genes in infantile epileptic encephalopathies. In: Noebels JL, Avoli M, Rogawski MA, Olsen RW, Delgado-Escueta AV, eds. *Jasper's Basic Mechanisms of the Epilepsies*. Bethesda, MD: National Center for Biotechnology Information (US), 2012. http://www.ncbi.nlm.nih.gov/books/NBK98182/ (accessed 3 Mar 2014).
4 Guerrini R, Parrini E. Epilepsy in Rett syndrome, and CDKL5- and FOXG1-gene-related encephalopathies. *Epilepsia* 2012;53:2067–78.
5 Marangi G, Ricciardi S, Orteschi D, Lattante S, Murdolo M, Dallapiccola B, Biscione C, Lecce R, Chiurazzi P, Romano C, Greco D, Pettinato R, Sorge G, Pantaleoni C, Alfei E, Toldo I, Magnani C, Bonanni P, Martinez F, Serra G, Battaglia D, Lettori D, Vasco G, Baroncini A, Daolio C, Zollino M. The Pitt-Hopkins syndrome: report of 16 new patients and clinical diagnostic criteria. *Am J Med Genet A* 2011;155:1536–45.
6 de Ligt J, Willemsen MH, van Bon BWM, Kleefstra T, Yntema HG, Kroes T, Vulto-Van Silfhout AT, Koolen DA, De Vries P, Gilissen C, Del Rosario M, Hoischen A, Scheffer H, De Vries BBA, Brunner HG, Veltman JA, Vissers LELM. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* 2012;367:1921–9.
7 Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W, Hosseini M, Behjati F, Haas S, Jamali P, Zecha A, Mohseni M, Püttmann L, Vahid LN, Jensen C, Moheb LA, Bienek M, Larti F, Mueller I, Weissmann R, Darvish H, Wrogemann K, Hadavi V, Lipkowitz B, Esmaeeli-Nieh S, Wieczorek D, Kariminejad R, Firouzabadi SG, Cohen M, Fattahi Z, Rost I, Mojahedi F, Hertzberg C, Dehghan A, Rajab A, Banavandi MJS, Hoffer J, Falah M, Musante L, Kalscheuer V, Ullmann R, Kuss AW, Tzschach A, Kahrizi K, Ropers HH. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* 2011;478:57–63.
8 Vissers LELM, de Ligt J, Gilissen C, Janssen I, Steehouwer M, De Vries P, Van Lier B, Arts P, Wieskamp N, Del Rosario M, Van Bon BWM, Hoischen A, De Vries BBA, Brunner HG, Veltman JA. A de novo paradigm for mental retardation. *Nat Genet* 2010;42:1109–12.
9 Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res* 2001;11:863–74.
10 Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–9.
11 Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* 2011;39:e118.
12 Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010;7:575–6.
13 Ferrer-Costa C, Gelpí JL, Zamakola L, Parraga I, de la Cruz X, Orozco M. PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinforma Oxf Engl* 2005;21:3176–8.
14 González-Pérez A, López-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet* 2011;88:440–9.
15 Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol* 2008;9(Suppl. 1):S4.
16 Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* 2010;38:W214–220.
17 Ballif BC, Rosenfeld JA, Traylor R, Theisen A, Bader PI, Ladda RL, Sell SL, Steinraths M, Surti U, Mcguire M, Williams S, Farrell SA, Filiano J, Schnur RE, Coffey LB, Tervo RC, Stroud T, Marble M, Netzloff M, Hanson K, Aylsworth AS, Bamforth JS, Babu D, Niyazov DM, Ravnan JB, Schultz RA, Lamb AN, Torchia BS, Bejjani BA, Shaffer LG. High-resolution array CGH defines critical regions and candidate genes for microcephaly, abnormalities of the corpus callosum, and seizure phenotypes in patients with microdeletions of 1q43q44. *Hum Genet* 2012;131:145–56.
18 de Munnik SA, García-Miñaúr S, Hoischen A, Van Bon BW, Boycott KM, Schoots J, Hoefsloot LH, Knoers NVAM, Bongers EMHF, Brunner HG. A de novo non-sense mutation in ZBTB18 in a patient with features of the 1q43q44 microdeletion syndrome. *Eur J Hum Genet* 2014;22:844–6.
19 Forrest M, Chapman RM, Doyle AM, Tinsley CL, Waite A, Blake DJ. Functional analysis of TCF4 missense mutations that cause Pitt-Hopkins syndrome. *Hum Mutat* 2012;33:1676–86.
20 Kousoulidou L, Tanteles G, Moutafi M, Sismani C, Patsalis PC, Anastasiadou V. 263.4 kb deletion within the TCF4 gene consistent with Pitt-Hopkins syndrome, inherited from a mosaic parent with normal phenotype. *Eur J Med Genet* 2013;56:314–18.
21 Scali O, Di Perri C, Federico A. The spectrum of mutations for the diagnosis of vanishing white matter disease. *Neurol Sci* 2006;27:271–7.
22 Hamdan FF, Gauthier J, Dobrzeniecka S, Lortie A, Mottron L, Vanasse M, D'Anjou G, Lacaille JC, Rouleau GA, Michaud JL. Intellectual disability without epilepsy associated with STXBP1 disruption. *Eur J Hum Genet* 2011;19:607–9.
23 Kodera H, Nakamura K, Osaka H, Maegaki Y, Haginoya K, Mizumoto S, Kato M, Okamoto N, Iai M, Kondo Y, Nishiyama K, Tsurusaki Y, Nakashima M, Miyake N, Hayasaka K, Sugahara K, Yuasa I, Wada Y, Matsumoto N, Saitsu H. De novo mutations in SLC35A2 encoding a UDP-galactose transporter cause early-onset epileptic encephalopathy. *Hum Mutat* 2013;34:1708–14.
24 Ng BG, Buckingham KJ, Raymond K, Kircher M, Turner EH, He M, Smith JD, Eroshkin A, Szybowska M, Losfeld ME, Chong JX, Kozenko M, Li C, Patterson MC, Gilbert RD, Nickerson DA, Shendure J, Bamshad MJ, Freeze HH. Mosaicism of the UDP-galactose transporter SLC35A2 causes a congenital disorder of glycosylation. *Am J Hum Genet* 2013;92:632–6.
25 Veeramah KR, Johnstone L, Karafet TM, Wolf D, Sprissler R, Salogiannis J, Barth-Maron A, Greenberg ME, Stuhlmann T, Weinert S, Jentsch TJ, Pazzi M, Restifo LL, Talwar D, Erickson RP, Hammer MF. Exome sequencing reveals new causal mutations in children with epileptic encephalopathies. *Epilepsia* 2013;54:1270–81.
26 Hagens O, Dubos A, Abidi F, Barbi G, Van Zutven L, Hoeltzenbein M, Tommerup N, Moraine C, Fryns JP, Chelly J, Van Bokhoven H, Gécz J, Dollfus H, Ropers HH, Schwartz CE, de Cassia Stocco Dos Santos R, Kalscheuer V, Hanauer A. Disruptions of the novel KIAA1202 gene are associated with X-linked mental retardation. *Hum Genet* 2006;118:578–90.
27 McFarland KN, Huizenga MN, Darnell SB, Sangrey GR, Berezovska O, Cha J-HJ, Outeiro TF, Sadri-Vakili G. MeCP2: a novel Huntingtin interactor. *Hum Mol Genet* 2014;23:1036–44.
28 Roux J-C, Zala D, Panayotis N, Borges-Correia A, Saudou F, Villard L. Modification of Mecp2 dosage alters axonal transport through the Huntingtin/Hap1 pathway. *Neurobiol Dis* 2012;45:786–95.

29 Yip DJ, Corcoran CP, Alvarez-Saavedra M, Demaria A, Rennick S, Mears AJ, Rudnicki MA, Messier C, Picketts DJ. Snf2l regulates Foxg1-dependent progenitor cell expansion in the developing brain. *Dev Cell* 2012;22:871–8.
30 Eckey M, Kuphal S, Straub T, Rummele P, Kremmer E, Bosserhoff AK, Becker PB. Nucleosome remodeler SNF2L suppresses cell proliferation and migration and attenuates Wnt signaling. *Mol Cell Biol* 2012;32:2359–71.
31 Blein S, Hawrot E, Barlow P. The metabotropic GABA receptor: molecular insights and their functional consequences. *Cell Mol Life Sci* 2000;57:635–50.
32 Wilkins A, Ping Q, Carpenter CL. RhoBTB2 is a substrate of the mammalian Cul3 ubiquitin ligase complex. *Genes Dev* 2004;18:856–61.
33 Villa N, Do A, Hershey JWB, Fraser CS. Human eukaryotic initiation factor 4G (eIF4G) protein binds to eIF3c, -d, and -e to promote mRNA recruitment to the ribosome. *J Biol Chem* 2013;288:32932–40.
34 Whalen S, Héron D, Gaillon T, Moldovan O, Rossi M, Devillard F, Giuliano F, Soares G, Mathieu-Dramard M, Afenjar A, Charles P, Mignot C, Burglen L, Van Maldergem L, Piard J, Aftimos S, Mancini G, Dias P, Philip N, Goldenberg A, Le Merrer M, Rio M, Josifova D, Van Hagen JM, Lacombe D, Edery P, Dupuis-Girod S, Putoux A, Sanlaville D, Fischer R, Drévillon L, Briand-Suleau A, Metay C, Goossens M, Amiel J, Jacquette A, Giurgea I. Novel comprehensive diagnostic strategy in Pitt-Hopkins syndrome: clinical score and further delineation of the TCF4 mutational spectrum. *Hum Mutat* 2012;33:64–72.
35 Rodríguez-Santiago B, Malats N, Rothman N, Armengol L, Garcia-Closas M, Kogevinas M, Villa O, Hutchinson A, Earl J, Marenne G, Jacobs K, Rico D, Tardón A, Carrato A, Thomas G, Valencia A, Silverman D, Real FX, Chanock SJ, Pérez-Jurado LA. Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome. *Am J Hum Genet* 2010;87:129–38.
36 Luedi PP, Dietrich FS, Weidman JR, Bosko JM, Jirtle RL, Hartemink AJ. Computational and experimental identification of novel human imprinted genes. *Genome Res* 2007;17:1723–30.
37 Ellery PM, Ellis RJ, Holder SE. Interstitial 3p25 deletion in a patient with features of 3p deletion syndrome: further evidence for the role of SRGAP3 in mental retardation. *Clin Dysmorphol* 2014;23:29–31.
38 Endris V, Wogatzky B, Leimer U, Bartsch D, Zatyka M, Latif F, Maher ER, Tariverdian G, Kirsch S, Karch D, Rappold GA. The novel Rho-GTPase activating gene MEGAP/ srGAP3 has a putative role in severe mental retardation. *Proc Natl Acad Sci USA* 2002;99:11754–9.
39 Saitsu H, Kato M, Mizuguchi T, Hamada K, Osaka H, Tohyama J, Uruno K, Kumada S, Nishiyama K, Nishimura A, Okada I, Yoshimura Y, Hirai S, Kumada T, Hayasaka K, Fukuda A, Ogata K, Matsumoto N. De novo mutations in the gene encoding STXBP1 (MUNC18-1) cause early infantile epileptic encephalopathy. *Nat Genet* 2008;40:782–8.
40 Deprez L, Weckhuysen S, Holmgren P, Suls A, Van Dyck T, Goossens D, Del-Favero J, Jansen A, Verhaert K, Lagae L, Jordanova A, Van Coster R, Yendle S, Berkovic SF, Scheffer I, Ceulemans B, De Jonghe P. Clinical spectrum of early-onset epileptic encephalopathies associated with STXBP1 mutations. *Neurology* 2010;75:1159–65.
41 Mignot C, Moutard M-L, Trouillard O, Gourfinkel-An I, Jacquette A, Arveiler B, Morice-Picard F, Lacombe D, Chiron C, Ville D, Charles P, Leguern E, Depienne C, Héron D. STXBP1-related encephalopathy presenting as infantile spasms and generalized tremor in three patients. *Epilepsia* 2011;52:1820–7.
42 Rauch A, Wieczorek D, Graf E, Wieland T, Endele S, Schwarzmayr T, Albrecht B, Bartholdi D, Beygo J, Di Donato N, Dufke A, Cremer K, Hempel M, Horn D, Hoyer J, Joset P, Röpke A, Moog U, Riess A, Thiel CT, Tzschach A, Wiesener A, Wohlleber E, Zweier C, Ekici AB, Zink AM, Rump A, Meisinger C, Grallert H, Sticht H, Schenck A, Engels H, Rappold G, Schröck E, Wieacker P, Riess O, Meitinger T, Reis A, Strom TM. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* 2012;380:1674–82.
43 Kim YO, Korff CM, Villaluz MMG, Suls A, Weckhuysen S, De Jonghe P, Scheffer IE, Head stereotypies in STXBP1 encephalopathy. *Dev Med Child Neurol* 2013;55:769–72.
44 Romaniello R, Saettini F, Panzeri E, Arrigoni F, Bassi MT, Borgatti R, A de-novo STXBP1 gene mutation in a patient showing the Rett syndrome phenotype. *Neuroreport* 2015;26:254–7.
45 Olson HE, Tambunan D, Lacoursiere C, Goldenberg M, Pinsky R, Martin E, Ho E, Khwaja O, Kaufmann WE, Poduri A. Mutations in epilepsy and intellectual disability genes in patients with features of Rett syndrome. *Am J Med Genet A* 2015;167:2017–25.
46 Hara M, Ohba C, Yamashita Y, Saitsu H, Matsumoto N, Matsuishi T. De novo *SHANK3* mutation causes Rett syndrome-like phenotype in a female patient. *Am J Med Genet A* 2015;167:1593–6.
47 Reiss AL, Faruque F, Naidu S, Abrams M, Beaty T, Bryan RN, Moser H. Neuroanatomy of Rett syndrome: a volumetric imaging study. *Ann Neurol* 1993;34:227–34.
48 MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, Jostins L, Habegger L, Pickrell JK, Montgomery SB, Albers CA, Zhang ZD, Conrad DF, Lunter G, Zheng H, Ayub Q, Depristo MA, Banks E, Hu M, Handsaker RE, Rosenfeld JA, Fromer M, Jin M, Mu XJ, Khurana E, Ye K, Kay M, Saunders GI, Suner MM, Hunt T, Barnes IHA, Amid C, Carvalho-Silva DR, Bignell AH, Snow C, Yngvadottir B, Bumpstead S, Cooper DN, Xue Y, Romero IG, Wang J, Li Y, Gibbs RA, Mccarroll SA, Dermitzakis ET, Pritchard JK, Barrett JC, Harrow J, Hurles ME, Gerstein MB, Tyler-Smith C. A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 2012;335:823–8.
49 EuroEPINOMICS-RES Consortium, Epilepsy Phenome/Genome Project, Epi4K Consortium. De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. *Am J Hum Genet* 2014;95:360–70.
50 Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, Wong WSW, Sigurdsson G, Walters GB, Steinberg S, Helgason H, Thorleifsson G, Gudbjartsson DF, Helgason A, Magnusson OT, Thorsteinsdottir U, Stefansson K. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 2012;488:471–5.
51 O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, Levy R, Ko A, Lee C, Smith JD, Turner EH, Stanaway IB, Vernot B, Malig M, Baker C, Reilly B, Akey JM, Borenstein E, Rieder MJ, Nickerson DA, Bernier R, Shendure J, Eichler EE. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 2012;485:246–50.
52 Fujioka S, Sundal C, Strongosky AJ, Castanedes MC, Rademakers R, Ross OA, Vilariño-Güell C, Farrer MJ, Wszolek ZK, Dickson DW. Sequence variants in eukaryotic translation initiation factor 4-gamma (eIF4G1) are associated with Lewy body dementia. *Acta Neuropathol (Berl)* 2013;125:425–38.
53 Blanckenberg J, Ntsapi C, Carr JA, Bardien S. EIF4G1 R1205H and VPS35 D620N mutations are rare in Parkinson's disease from South Africa. *Neurobiol Aging* 2014;35:445.e1–3.
54 Li K, Tang B, Guo J, Lou MX, Lv ZY, Liu ZH, Tian Y, Song CY, Xia K, Yan XX. Analysis of EIF4G1 in ethnic Chinese. *BMC Neurol* 2013;13:38.
55 Puschmann A. Monogenic Parkinson's disease and parkinsonism: clinical phenotypes and frequencies of known mutations. *Parkinsonism Relat Disord* 2013;19:407–15.
56 Sudhaman S, Behari M, Govindappa ST, Muthane UB, Juyal RC, Thelma BK, VPS35 and EIF4G1 mutations are rare in Parkinson's disease among Indians. *Neurobiol Aging* 2013;34:2442.e1–3.

Part 1 - Detailed clinical and results description

Participants

Patient selection and recruitment

The Life and Health Sciences Research Institute of the School of Health Sciences at University of Minho offers the molecular diagnostic for Rett syndrome in Portugal since 2002 by *MECP2* genetic analysis (Sanger sequencing and dosage analysis by qPCR of all exons). The lab´s work on Rett syndrome includes the setup of a clinical database and detailed clinical analysis of mutation-positive and mutation-negative cases and genoptype-phenotype correlations (and also the behavioral, neuroanatomical and molecular study of a mouse model of the disease, the Mecp2 KO mouse). Due to previous research on this pathology and in intellectual disability as a wider category, a large bank of DNA samples of patients and parents (as well as extensive clinical information) has been obtained. Patients were selected to participate in this project based on clinical suspicion of RTT (typical or atypical) according to the recently revised diagnostic criteria(Neul *et al.*, 2010)and availability of both parents' DNA. The clinical information was gathered in an anonymous database. The enrollment of the patients and their families was done, after explanation about its purposes, potential pitfalls and eventual benefits by the referring doctor. Written informed consent was obtained for all the participants. This study was approved by the ethics committee of Hospital de Santo António, Centro Hospitalar do Porto.

Clinical description of the patients

Patient 1 – This girl is the middle child of a healthy non-consanguineous couple; her two male sibs are healthy. She was born at full term through eutocic delivery with a weight of 2630g (5-10th centile) and OFC of 33cm (10th centile). The neonatal period had no intercurrences. Head control occurred around 7 months, sitting without support and purposeful grasp at 12 months. Independent walking and language were never acquired. From an early age autistic traits (e.g. gaze avoidance) were noted. Partial complex seizures began when the child was around 3 years old; good control was attained with valproate and clonazepam. At 6 years, purposeful grasp was lost. Currently, at 15 years of age, the patient presents severe intellectual disability, swallowing difficulties, breathing disturbances and laughing spells. On physical exam the patient presents growth retardation and severe scoliosis; head circumference is on the 25th centile. On neurologic exam, hand washing and clapping stereotypies, as well as tremor of the upper limbs and dystonia of the lower limbs were noted. Brain MRI revealed frontal periventricular heterotopies and frontal and temporal atrophy. Metabolic screening, karyotype, microarrays (Agilent 180k), *MECP2* and *CDKL5* Sanger sequencing showed no mutations.

Patient 1 presents a 83,5 kb paternal microdeletion in chromosome 3 (chr3:113,826,308-113,909,889) encompassing the *DRD3* (DOPAMINE RECEPTOR D3) gene. *DRD3* encodes the D3 subtype of the five (D1-D5) dopamine receptors. The activity of the D3 subtype receptor is mediated by G proteins, which inhibit adenylyl cyclase. This receptor is localized to the limbic areas of the brain, which are associated with cognitive, emotional, and endocrine functions. Association studies have linked *DRD3* to risk for schizophrenia (Spurlock *et al.*, 1998) and essential tremor (Lucotte *et al.*, 2006). Deletions encompassing exons of *DRD3* are not observed in DGV, which suggests that preservation of this gene is very relevant for health. *DRD3* has also been pointed out as being one of the most relevant genes in the 3q13.31 deletion syndrome, in which patients invariably present ID and frequently also craniofacial dysmorphic features (high arched palate, short philtrum and protruding lips), skeletal malformations (scoliosis, lordosis, thoracic kyphosis, joint contractures), postnatal overgrowth, agenesis of the corpus callosum and hypoplastic male genitalia. Behavioral problems such as autism and attention deficit disorders have also been described in some affected individuals (Vuillaume *et al.*, 2013; Wiśniowiecka-Kowalnik *et al.*, 2013). No variants were found in the the *DRD3* gene using whole exome sequencing (WES). A list of the variants found in the patient is presented in table S1.

Patient 2 – This girl is the only child of a healthy non-consanguineous couple, whose family history is negative for developmental delay/intellectual disability. She was born at full term through eutocic delivery with a weight of 2780g (10th centile) and OFC of 33cm (10th centile). There were no intercurrences in the neonatal period. Head control occurred around 5 months, sitting with support at 6 months. Purposeful grasp was acquired at 3 months and lost around 6 months. Stagnation was noted at 8 months, coinciding with the beginning of complex partial seizures and lack of social interaction; virtually no acquisitions were achieved posteriorly, including gait or sphincter control. Currently, with 17 years of age, the patient presents severe intellectual disability, crying and laughing spells, respiratory dysfunction, frequent psychomotor agitation and sleep disturbances (waking up in the middle of the night). On neurologic exam, dystonia, bradykinesia, continuous stereotypies, tremor of the upper limbs and pyramidal signs of the lower limbs were noted, as well as swallowing difficulties. Brain MRI at the age of 4years and 10 months showed significant striatum atrophy (specially the caudate nuclei) as well as mild atrophy of the cortex and cerebellar vermis. Metabolic screen, *MECP2* and *CDKL5* Sanger sequencing were normal. This patient has two compound heterozygous variants in *HTT* gene. HTT encodes a protein that plays a role in embryogenesis and that, when absent, results in the impairment of specification of ectodermal and mesodermal lineages.(Nguyen *et al.*, 2013) Even though RTT is a neurodevelopmental disorder and Huntington Disease is a late onset neurodegenerative disease, they seem to share some features at the molecular and clinical level.(Reiss *et al.*, 1993; McFarland *et al.*, 2014) We believe that it is possible that the combination of two missense variants in *HTT* (though one is a rare polymorphism) in patient 2 may contribute for the RTT-like phenotype. Noteworthy, *HTT* homozygous loss of function variants have not been observed in any individual from the 1000 genomes project cohort suggesting that a *HTT* KO (and likely KD) has significant impact in health in humans, as seen in mice.(Duyao *et al.*, 1995; MacArthur *et al.*, 2012) A list of the variants found in the patient is presented in table S2.

Patient 3 – This girl is the youngest daughter of a healthy non-consanguineous couple; her brother and sister are healthy. She was born at full term through eutocic delivery with a weight of 2245g (<5th centile) and OFC of 32cm (<5th centile). The neonatal period had no intercurrences but at 5 months West syndrome was diagnosed. Head control was achieved at 4 months, sitting without support at 11 months, first words at 12 months. Purposeful grasp was noted at 17

months but partially lost at 3 years. At 2 years she started walking (dyspraxic gait) and at 4 years she acquired sphincter control and the capability to say a few sentences (but not always purposeful; jargon and echolalia were frequent). Currently, with 13 years of age, the patient presents severe intellectual disability, poor eye contact, laughing spells, breathing disturbances (hyperpneia) and periods of psychomotor agitation. In the neurologic exam, dystonia and hand washing stereotypies were observed until the age of 5 years. Metabolic screening, karyotype, microarrays (Agilent 180k), *MECP2* and *CDKL5* Sanger sequencing were normal. A list of the variants found in the patient is presented in table S3.

Patient 4 – This girl is the daughter of a healthy non-consanguineous couple. She has a healthy brother and the remaining family history is negative for developmental delay/intellectual disability. She was born at full term through eutocic delivery with a weight of 3600g (50-75th centile) and OFC of 36cm (75th centile). Neonatal period had no intercurrences. At 7 months she started having complex partial seizures. Developmental acquisitions occurred very slowly, with sitting without support occurring around 14 months; walking and language were never attained. Purposeful grasp was acquired when she was 5 years old but lost 2 years later. Currently, with 25 years of age, the patient presents severe intellectual disability, autism, laughing spells, eye pointing and sleep disturbances (waking up in the middle of the night). On physical exam kyphosis/scoliosis, peripheral vasomotor disturbances and small cold feet were noted. The neurologic exam revealed stereotypies, dystonia, pyramidal signs. Microarrays (Agilent 180k), *MECP2* and *CDKL5* Sanger sequencing revealed no significant abnormalities. This patient carries a *de novo* variant in *SMARCA1* (alias *SNF2L*) gene. This gene encodes a chromatin remodeling ATPase involved in the control of Wnt signaling by regulation of proliferation and cell migration(Eckey *et al.*, 2012) and was also found to function antagonistically with Foxg1 in the regulation of brain size in mice.(Yip *et al.*, 2012) A list of the variants found in the patient is presented in table S4.

Patient 5 – This girl is the daughter of a healthy non-consanguineous couple, whose remaining family history is negative for developmental delay/intellectual disability; her elder brother is healthy. She was born at full term through eutocic delivery with a weight of 3260g (25-50th centile) and OFC of 33cm (10-25th centile). Motor delay was present from the beginning, with head control achieved at 4 months, and seems to have been superimposed by regression at 8 months. Sitting with support was achieved at 12 months and tiptoe gait at 26 months. The child never learnt any words. Currently, at 15 years of age, the patient presents severe intellectual

disability, screaming spells, bruxism when awake and breathing disturbances (apnea followed by hyperpnoea). The neurologic exam revealed microcephaly, stereotypies, and dystonia. Brain MRI, metabolic screen, karyotype and *MECP2* Sanger sequencing revealed no significant abnormalities. A list of the variants found in the patient is presented in table S5.

Patient 6 – This girl is the only child of a healthy non-consanguineous couple; a maternal great aunt had ID and epilepsy reportedly after having had meningitis. She was born at full term through dystocic delivery (requiring forceps) with a weight of 3565g (50-75th centile), length of 49cm (25-50th centile) and OFC of 34,5cm (25-50th centile). Congenital hip dislocation was diagnosed and casts were used until 9 months of age. Head control was achieved around 3 months, sitting without support at 6 months, first words at 12 months and walking without support at 13 months. Regression occurred around 14-20 months of age, with loss of almost all previously acquired motor skills. Currently, at 13 years of age, the patient presents severe intellectual disability. On physical exam growth retardation was clear (weight and height < 5th centile), though the child was normocephalic. Atetosis, pyramidal signs and wallowing difficulties due to oropharyngeal dystonia were noted on the neurologic exam. Brain MRI revealed hyperintensities in the putamen and caudate nuclei; spectroscopy was normal. Metabolic screening, karyotype and *MECP2* Sanger sequencing revealed no significant abnormalities.

A *de novo* variant c.A1801G (N601D) in the *LARP4* (LA RIBONUCLEOPROTEIN DOMAIN FAMILY, MEMBER 4) gene was found in this patient, resulting in the substitution of an asparagine for an aspartic acid and predicted to be pathogenic by the SIFT prediction tool. *LARP4* encodes a La-related protein which can bind poly(A) RNA and interact with the poly(A) binding protein (PABP), contributing in this way for mRNA stability and homeostasis (Yang *et al.*, 2011; Hussain *et al.*, 2013; Merret *et al.*, 2013). So far, no variants were described in *LARP4*. However, due to its participation in such an important biological process, known to be perturbed in several ID syndromes, we cannot ignore its eventual contribution to ID pathogenesis. A list of the variants found in the patient is presented in table S6.

Patient 7 – This girl is the second child of a healthy non-consanguineous couple; her two sibs (one male and one female) are healthy. The couple has lost spontaneously two pregnancies (one in the 2nd and another in the 3rd trimester); gender and presence of other malformations in the fetuses is unknown. She was born full term through eutocic delivery with a weight of 3800g (75th centile) and OFC of 35,0cm (50th centile). Head control occurred around 6 months, sitting without support at 10 months, walking at 3,5 years, girst words with 5 years. Stereotypies were noted

when she was 2,5 years and for many years she didn't achieve purposeful grasp. She started having purposeful hand use around 6 years of age, but lost it between 9 and 10 years. When she was 7, she had the first hyperpnea episodes and tonico-clonic generalized seizures as well as nocturnal crying and laughing spells. Currently, as a 31 year old adult, the patient presents severe intellectual disability, bruxism when awake, eye pointing, episodes of psychomotor agitation. On physical exam kyphosis/scoliosis and peripheral vasomotor disturbance were registered. Brain MRI, metabolic screening, karyotype and *MECP2* Sanger sequencing showed no significant abnormalities.

Patient 8 – This girl is the first child of a healthy non-consanguineous couple. She has a healthy younger sister. A first grade maternal cousin died with spinal muscular atrophy type I. The remaining family history is negative for developmental delay/intellectual disability. She was born at full term through eutocic delivery with a weight of 2870g (10-25th centile) and OFC of 35cm (50th centile). Head control occurred around 6 months, sitting without support at 6 months. Deceleration of head growth was noted at 6 months. Walking without support and first words were achieved at 18 months and building sentences at 24 months. Regression occurred between 2 and 2,5 years with loss of language and purposeful grasp and coincided with the appearance of severe autism (screened with CARS and confirmed with ADI-R). As time went by, a cervical kyphosis became evident. Currently, at the age of 13 , the patient presents severe intellectual disability, epilepsy (well controlled with anti-epileptic drugs), psychomotor agitation, screaming spells and poor eye contact. On neurological exam the patient presents stereotypies, ataxia and dysmetria and peripheral vasomotor disturbances. Brain MRI, karyotype, metabolic screen, microarrays (Agilent 180k), *MECP2* and *CDKL5* Sanger sequencing revealed no significant abnormalities. A list of the variants found in the patient is presented in table S11.

Patient 9 – This girl is the only child of a healthy non-consanguineous couple; the remaining family history is negative for developmental delay/intellectual disability. She was born at full term through eutocic delivery with a weight of 3140g (25th centile) and OFC of 34,0cm (25-50th centile). Stagnation of psychomotor development was noted early on (7 months) followed by regression, as well as lack if interest in exploring/interacting with the environment. Also at 7 months of life head growth acceleration was documented with posterior stabilization at 2 years of age around the 98th centile. Head control was attained at around 5 months of age, sitting without support at 18 months and walking at 24 months. Purposeful grasp, language and sphincter control were never acquired. Currently, at 19 years of age, the patient presents severe intellectual

disability, episodes of psychomotor agitation with crying spells, bruxism when awake, sleep disturbances (waking up in the middle of the night) and breathing disturbances (hyperventilation). On neurological exam stereotypies, dystonia (lower limbs) and swallowing difficulties were noted. Brain MRI revealed frontoparietal enlargement of CSF spaces. Metabolic screen and *MECP2* Sanger sequencing revealed no significant abnormalities. Patient 9 has a *de novo* variant in *GABBR2* gene. *GABBR2* encodes a Gamma-aminobutyric acid (GABA) type B receptor involved in neuronal activity inhibition and that is co-expressed with FOXG1.(Blein *et al.*, 2000; Wang, 2006) Because GABA receptors play an important role in maintaining the excitatory-inhibitory balance in brain, the possibility that their deregulation may be associated with epilepsy is not to be dismissed. *GABBR2* expression was found to be reduced in lateral cerebella from subjects with schizophrenia, bipolar disorder and major depression (Fatemi *et al.*, 2011) as well as in the cerebella of autistic patients.(Fatemi *et al.*, 2009) Common insertion/deletion polymorphisms affecting *GABBR2* were described in patients with autism although without statistically significant enrichment when compared to control populations.(Hedges *et al.*, 2012) Additionally, *de novo* missense variants in *GABRR2* were identified in two different patients with infantile spasms.(EuroEPINOMICS-RES Consortium *et al.*, 2014) This information brings new insight into the association of *GABRR2* with neurodevelopmental disorders and highlights the importance of clinical and WES data sharing. A list of the variants found in the patient is presented in table S9.

Patient 10 – This female patient is the second child of a healthy non-consanguineous couple; her male sibling had absence seizures that later evolved to tonic-clonic seizures; he is otherwise healthy. She was born full term through eutocic delivery with a weight of 2730g (10th centile) and OFC of 30,0cm (<2th centile). In the neonatal period the following dysmorphisms were described: beaked nose, widened space between the nipples, cauliflower anus, low implantation of the 5th toes. Head control occurred around 12 months, sitting without support at 36 months. Walking, talking and fine motor coordination were never attained. Epilepsy (partial seizures) started at 24 months, and so did midline hand stereotypies. When the child was 6 years old, she started having episodes of hyperpena followed by apnea. Currently, with 24 years of age, the patient presents severe intellectual disability, has screaming spells and sleep disturbances (waking up in the middle of the night). On physical exam, growth retardation and microcephaly were apparent, as well as kyphosis/scoliosis and peripheral vasomotor disturbances. On neurological exam dystonia and swallowing difficulties were noted. Brain MRI, metabolic screening, karyotype and *MECP2* and *CDKL5* Sanger sequencing revealed no significant abnormalities.She was classified

as a RTT patient (see table 1 in main text) in light of her loss of acquired hand skills, spoken language, gait abnormalities and stereotypical hand movements. A paternally inherited variant c.G2971A (p.V991I) in the *MAGEL2* (MAGE-LIKE 2) gene (OMIM 605283) was found, resulting in the substitution of a valine for an isoleucine, predicted not to have a functional impact by the SIFT, PolyPhen2 and Mutation Assessor prediction tools. However, given the imperfection of these tools it is not possible to exclude this variant as a basis of disease. *MAGEL2* is located at the critical region for Prader-Willi syndrome and paternally expressed in the CNS (Lee *et al.*, 2000). Recently, *de novo* truncating variant in the paternal allele of *MAGEL2* were described in patients with Prader-Willi Syndrome features, resulting in the loss of expression of functional protein (Schaaf *et al.*, 2013). The contribution of this alteration to the disease in this patient is not clear since it was not possible to determine the parental origin of the variant in the father (and/or if it occurred *de novo* in the father, being then inherited). A list of the variants found in the patient is presented in table S8.

Patient 11 – This girl is the second child of a non consanguineous couple. The mother has resting tremor and is suspected of having psychiatric disease, possibly early-onset dementia. The maternal grandmother is bed-ridden and dementiated. The father is suspected of also having psychiatric disease. Formal neurological and psychiatric evaluation of both parents is pending. The older male sibling is healthy and so are the remainder more distant relatives. She was born full term through dystocic delivery (forceps) with a weight of 2880g (10-25th centile), length 46,3cm (5-10th centile) and OFC of 34,3cm (25-50th centile). She gained head control at 3 months and sitting with support at 6 months. Between 6 and 9 months, her development seemed to stagnate and her head growth decelerated. At 9 months generalized epilepsy began, good response to anti-epileptic drugs. Currently, with 6 years of age, the patient presents severe intellectual disability, growth retardation, microcephaly, restless nights with diurnal sleep, bruxism when awake, diminished response to pain and eye pointing. On neurological exam the patient presents stereotypies, resting tremor and peripheral vasomotor disturbances. Brain MRI, metabolic screening, karyotype and *MECP2, CDKL5, UBE3A* and *PCDH19* Sanger sequencing revealed no significant abnormalities.

Patient 11 has a combination of two variants: a *de novo* variant in *RHOBTB2* gene and a homozygous variant in *EIF4G1* gene. *RHOBTB2* belongs to the Rho GTPases family and was found to bind to CUL3 and to be a substrate of the Cul3-based ubiquitin ligase complex, which is necessary for mitotic cell division.(Wilkins *et al.*, 2004) Moreover, *de novo* nonsense variants in

CUL3 were identified in two separate next-generation sequencing reports using ASD probands,(Kong *et al.*, 2012; O'Roak, Vives, Girirajan, *et al.*, 2012) further strengthening the relationship between *RHOBTB2* and neurodevelopmental disorders. RhoBTB2 is also likely to interfere with several neuron related functions, given its participation in cytoskeleton and membrane trafficking networks.(Siripurapu *et al.*, 2005, p. 2) *EIF4G1* encodes a translation initiation factor involved in the recruitment of mRNA to the 40S ribosomal subunit, which is a rate-limiting step during the initiation phase of protein synthesis.(Villa *et al.*, 2013) Variants in *EIF4G1* have been associated with autosomal dominant forms of Lewy body dementia(Fujioka *et al.*, 2013) and Parkinson disease (with and without dementia). However, the true pathogenicity of some of these variants remains unknown as it has been difficult to replicate the findings in populations with different genetic backgrounds.(Li *et al.*, 2013; Puschmann, 2013; Sudhaman *et al.*, 2013; Blanckenberg *et al.*, 2014) In the Human Genome Mutation Database, the mutations that have been clearly associated with Parkinson disease are in heterozigosity, whereas our patient has a variant in homozigosity, which could justify an early and more severe presentation of neurological disease. A list of the variants found in the patient is presented in table S13.

Patient 12 – This girl is the first child of a healthy non-consanguineous couple. She has a healthy younger sister; the remaining family history is also negative for developmental delay/intellectual disability. During pregnancy, intra-uterine growth retardation and single umbilical artery were detected. She was born at full term through eutocic delivery with a weight of 2150g (<5th centile), length of 42cm (<5th centile) and OFC of 30,5cm (2nd centile). During the neonatal period, left ptosis and cleft palate were diagnosed. Head control was achieved at around 6 months, sitting without support at 2 years, dyspraxic gait with support was eventually attained, but never independent walking. Even though she has hand stereotypies, hand use is preserved. At 10 years of age, on physical exam the child presented normal growth but severe microcephaly (weight: 35Kg, height:145cm, OFC:43cm), scoliosis/kyphosis and peripheral vasomotor disturbances. Her significant swallowing difficulties and gastroesophageal reflux prompted the decision of performing a gastrostomy. Currently, at 19 years of age, the patient presents severe intellectual disability, eye pointing and breathing disturbances (hyperapnea). On neurological exam stereotypies and ataxia were noted. Brain MRI, metabolic screening, karyotype, microarrays (Agilent 180k) and *MECP2* Sanger sequencing revealed no significant abnormalities. A list of the variants found in the patient is presented in table S10.

Patient 13 – This girl is the daughter of a healthy but possibly remotely consanguineous couple. She was born at full term through eutocic delivery with a weight of 3610g (50-75th centile) and OFC of 35cm (50th centile).The beginning of epilepsy (well controlled with therapy) at 24 months coincided with regression (namely loss of language) at 24 months. Currently, at the age of 20 years , the patient presents severe intellectual disability, crying and screaming spells, and restless nights with diurnal sleep. The patient is also affected with scoliosis/kyphosis and gastro-esophageal reflux disease. On neurological exam she presents pyramidal signs. Brain MRI, metabolic screening, karyotype and *MECP2* Sanger sequencing revealed no significant abnormalities.

Patient 13 presents two compound heterozygous variant in the *TCTN2* (TECTONIC FAMILY, MEMBER 2) (OMIM 613846) gene: a maternal c.C668T (p.T223M), resulting in a substitution of a threonine for a methionine, a very rare variant in the population (rs145374149, MAF T=0.0009/2) and predicted not to be pathogenic by in silico softwares; a paternal c.T1538C (p.I513T), resulting in a substitution of a isoleucine for a threonine predicted not to be pathogenic by in silico softwares. The *TCTN2* gene is a paralogous of *TCTN1* (TECTONIC FAMILY, MEMBER 1) (OMIM 609863) that plays a role in sonic hedgehog signaling pathway modulation (Shaheen *et al.*, 2011). Homozygous and compound heterozygous variants in *TCTN2* gene were described in patients with Meckel-Gruber 8 syndrome and Joubert syndrome (Sang *et al.*, 2011; Shaheen *et al.*, 2011, p. 2). A list of the variants found in the patient is presented in table S12.

Patient 14 – This boy is the only child of a healthy non consanguineous couple. The older male sibling is healthy and so are the remainder more distant relatives. He was born at full term through C-section with a weight of 3444g (25-50th centile), length 51,5cm (50-75th centile) and OFC of 35,5cm (25-50th centile). At 1 month West syndrome was diagnosed; control of seizures has been difficult. Development was significantly delayed, with purposeful grasp and sitting being attained at 15 months. At 6 years old he started walking without support but few other psychomotor milestones were conquered. Currently, with 6 years of age, the patient presents severe intellectual disability, psychomotor agitation, restless nights with diurnal sleep and bruxism when awake. On physical exam it was confirmed that growth is within a normal interval. On neurological exam the patient presents poor eye contact, stereotypies, mioclonies, hyperventilation and decreased response to pain. Brain MRI showed moderate enlargement of pericerebral spaces – likely generalized brain atrophy. Metabolic screening, karyotype and

MECP2, CDKL5, ARX Sanger sequencing revealed no significant abnormalities. A list of the variants found in the patient is presented in table S14.

Patient 15 – This boy is the second child of a healthy non consanguineous couple. The older female sibling is healthy and so are the remainder more distant relatives. He was born at full term through eutocic delivery with a weight of 3250g (25-50th centile) and OFC of 33cm (10th centile). Food refusal and recurrent vomiting lead to growth decelaration: around 4 months of life the patient was crossing centiles both in weight (from 50th centile to 10th centile), height (from 25th centile to 10th centile) and OFC (from 10th centile to <3rd centile). Around 10 months the patient had his first apyretic seizure (difficulty breathing, involuntary movements, gaze deviation). Since then, few psychomotor acquisitions were achieved and autism spectrum features were detected. Currently, with 8 years of age, the patient presents severe intellectual disability, psychomotor agitation, crying spells and disrupted sleep (wakes up in the middle of the night). On physical exam it was verified that the patient had caught up in terms of weight and height but microcephaly was still present. Some minor dysmorphism were observed: sunken eyes, prominent forehead, shortening of the 4th and th fingers, brittle nails. On neurological exam the patient presents stereotypies of the head and hands. Brain MRI (performed at 19 months and repeated at 4 years of age) showed global brain atrophy, thin corpus callosum and bilateral fronto-parietal periventricular heterotopies. EEG at 7 years showed a disorganized pattern during sleep, with predominantly occipital continuos lentification and paroxystic activity. Metabolic screening, karyotype and *MECP2* Sanger sequencing revealed no significant abnormalities. A list of the variants found in the patient is presented in table S15.

Patient 16 – This girl is the conception of a healthy and possibly remotely consanguineous couple. A paternal second grade male cousin is affected with idiopathic intellectual disability. She was born at full term through eutocic delivery with a weight of 3800g (75th centile) and OFC of 34,5cm (25-50th centile). At 6 months, coinciding with inaugural West syndrome there was developmental arrest, lack of interest in interacting/exploring the environment and head growth deceleration (currently being microcephalic). Currently, with 9 years of age, the patient presents severe intellectual disability, psychomotor agitation episodes, screaming and crying spells. On neurological exam the patient presents poor eye contact and lack of interest in social interaction, repetitive behavior, stereotypies (hand wringing) and ataxia. Brain MRI, metabolic screening, karyotype and *MECP2* and *CDKL5* Sanger sequencing had no significant abnormalities. A list of the variants found in the patient is presented in table S16.

Patient 17 – This girl is the second child of a healthy non-consanguineous couple; her older brother is healthy and so are more distant relatives. She was born at full term through dystocic delivery (vacuum extraction) with a weight of 3700g (50-75th centile). Developmental delay was noticeable all along, with further stagnation at 10 months with the beginning of epilepsy. She sat at 16 months, walked with support at 30 months and without support at 48 months. Her first words were learnt at 36 months but she never constructed sentences. Currently, with 13 years of age, the patient presents severe intellectual disability, autistic behavior, crying spells and bruxism when awake. Disrupted sleep became so problematic that hipnotics were prescribed. On neurological exam the patient presents poor eye contact, stereotypies, respiratory dysfunction (hyperpnoea alternating with apneas) and diminished response to pain. Brain MRI showed global enlargement of the CSF space suggestive of generalized brain atrophy. Metabolic screening, karyotype, FISH for 17p11.2 (Smith-Magenis syndrome), DNA methylation analysis of 15q11.2-q13 (AS/PWS critical region), *MECP2* and *UBE3A* Sanger sequencing revealed no significant abnormalities. A list of the variants found in the patient is presented in table S17.

Patient 18 – This girl is the first child born to a non-consanguineous couple. Her mother has mild intellectual disability; the remaining family history is negative for developmental delay/intellectual disability. She was born at full term from dystocic delivery (vacuum extraction) with a weight of 2770g (5-10th centile), length 49,0cm (25-50th centile) and OFC of 33,3cm (10-25th centile). She sat without support at 11 months, said her first words at 12 months (but did not develop language), achieved purposeful grasp at 18 months and walked without support at 30 months. Also around 30 months, ASD was suspected on the basis of hand flapping, temper tantrums, autoagression. Currently, at 9 years old, the patient presents severe intellectual disability, bruxism when awake, crying spells and disrupted sleep (waking up in the middle of the night). On physical exam her growth is normal. On neurological exam the patient presents stereotypies (hand flapping), peripheral vasomotor disturbances, ataxia, diminished response to pain. Brain MRI, metabolic screening, karyotype, microarrays (Agilent 180k), *MECP2* Sanger sequencing revealed no significant abnormalities.

In this patient a maternally inherited variant c.G3226T (A1076S) in *CHD8* (CHROMODOMAIN HELICASE DNA-BINDING PROTEIN 8) (OMIM 610528) was found, resulting in the substitution of an alanine for a serine and predicted by SIFT and Mutation Taster to be damaging. *CHD8* encodes a chromodomain helicase DNA-binding protein that participates in the Wnt-β-catenin signaling pathway trough histone H1 recruitment during development (Nishiyama *et al.*, 2012). In

rat, duplin (corresponding to *CHD8*), binds to beta-catenin inhibiting the binding of Tcf to beta-catenin and consequently inhibiting the Wnt signaling pathway (Kobayashi *et al.*, 2002). Noteworthy, *CHD8* seems to be co-expressed with *MECP2*, the main cause of RTT (Smirnov *et al.*, 2009; Kang *et al.*, 2010). Rare *de novo* variants in *CHD8* have been described as associated with autism spectrum disorder in recent years (Neale *et al.*, 2012; O'Roak, Vives, Fu, *et al.*, 2012; O'Roak, Vives, Girirajan, *et al.*, 2012; Talkowski *et al.*, 2012; Krumm *et al.*, 2014). A list of the variants found in the patient is presented in table S18.

Patient 19 – This boy is the child of a healthy non-consanguineous couple; his younger sister is healthy and so are the remainder distant relatives. He was born at full term through dystocic delivery (vacuum extraction) with a weight of 3300g (25-50th centile), length 49,0cm (25-50th centile) and OFC of 34,0cm (25-50th centile). He sat without support at 6-9 months, acquired purposeful grasp at 2 years and controlled sphincters at 10 years. Currently, with 14 years of age, the patient presents severe intellectual disability, dyspraxic gate, eye pointing, bruxism when awake and scoliosis/kyphosis. He is also affected with gastroesophageal reflux. On neurologic exam he presents microcephaly, peripheral vasomotor disturbances, diminished response to pain and autistic traits (including lack of interest in social interaction). The patient does not present short stature, hip dislocation or epilepsy. Brain MRI, metabolic screening, karyotype, microarrays (Agilent 180k) and *MECP2* Sanger sequencing revealed no significant abnormalities. A list of the variants found in the patient is presented in table S19.

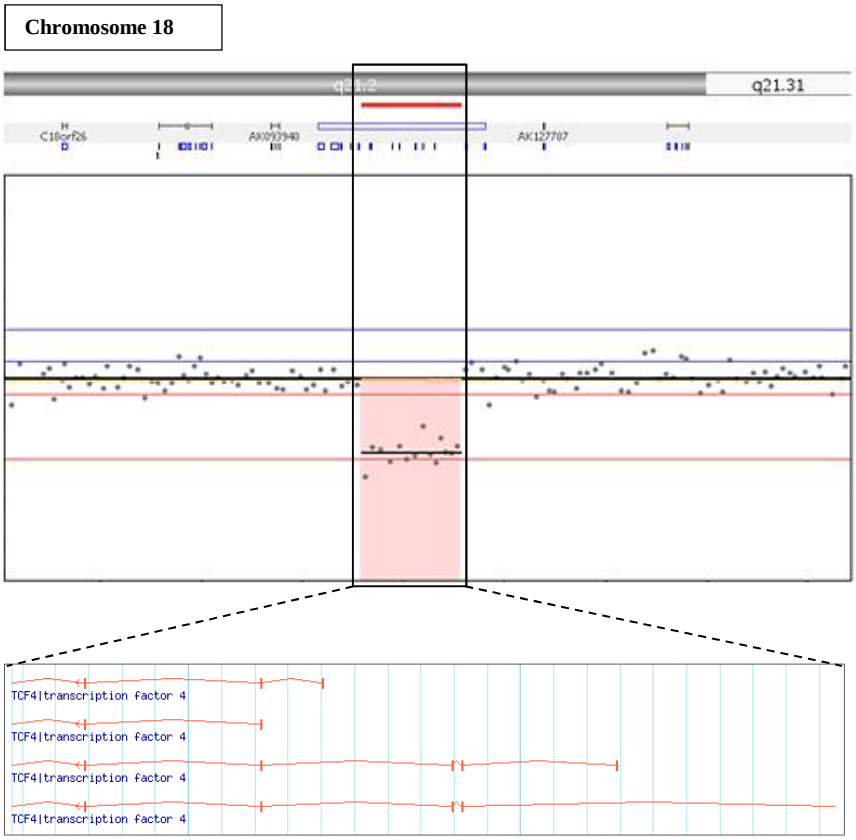


Figure S1.1 – Schematic representation of the 18q21.1 deletion in patient 7. It is possible to observe the decrease in the LRR of the probes contained in that region when compared to neighboring probes (LRR=0). The deletion is 250Kb long and encompasses two genes: *TCF4* and *MIR4529*.

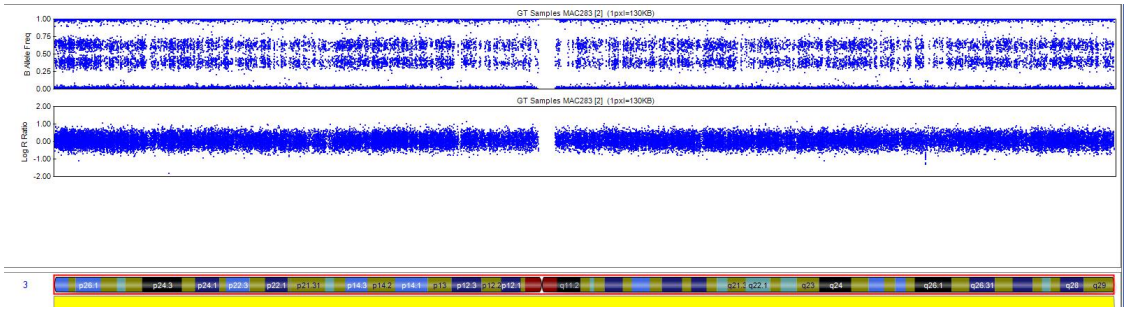


Figure S1.2 – Schematic representation of the UPD mosaic of chromosome 3 in patient 16. It is possible to observe that the entire chromosome 3 has a LRR=0 and a BAF split (0,3 and 0,6), compatible with mosaic UPD.

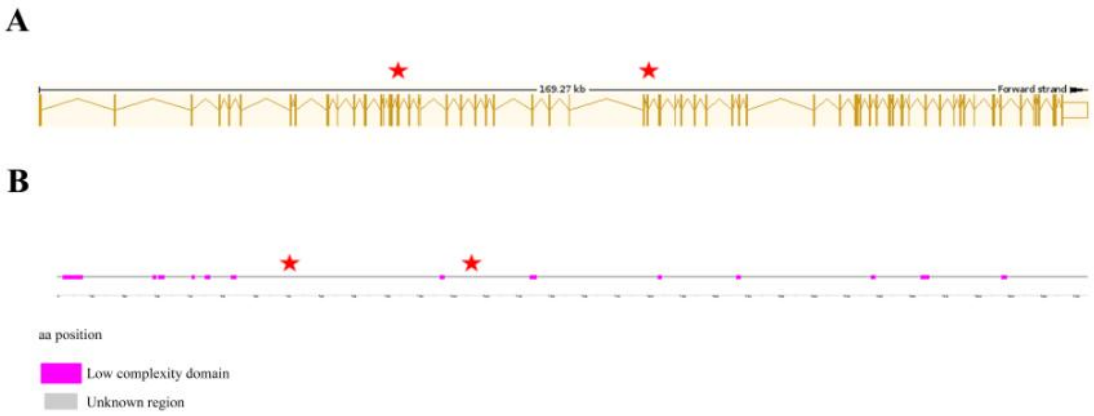


Figure S1.3 – Schematic representation of the *HTT* gene and HTT protein domains. The gene structure was retrieved from Ensembl and the domain structure was determined based on SMART webtool. The variants identified in this gene are indicated by red stars with their respective location at exons and protein levels (aa). Both variants affect and unknown region/domain.

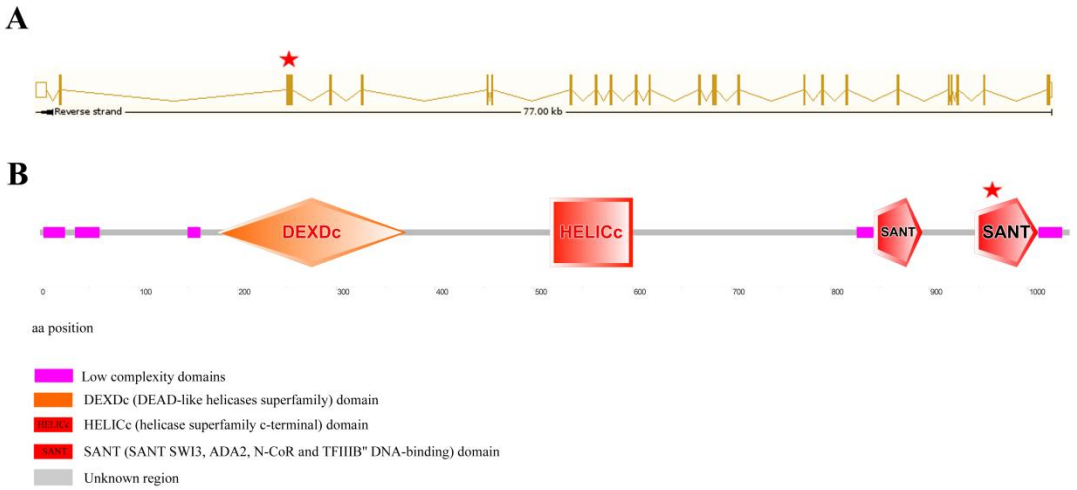


Figure S1.4 – Schematic representation of the *SMARCA1* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structure was determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant is located id the SANT domain, which is present in the subunits of many chromatin-remodelling complexes (Aasland *et al.*, 1996).

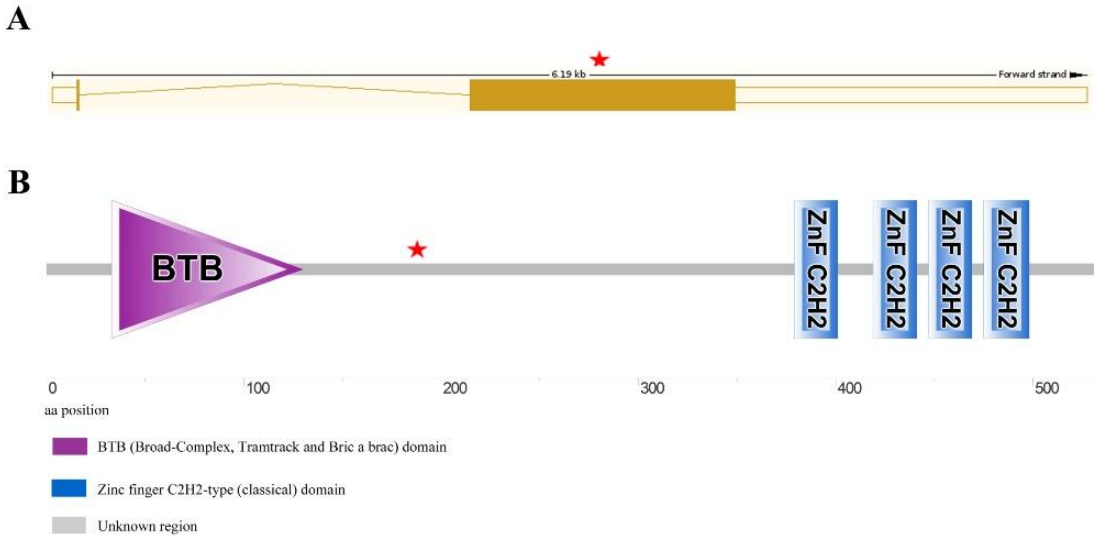


Figure S1.5 – Schematic representation of the *ZNF238* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects and unknown region/domain.

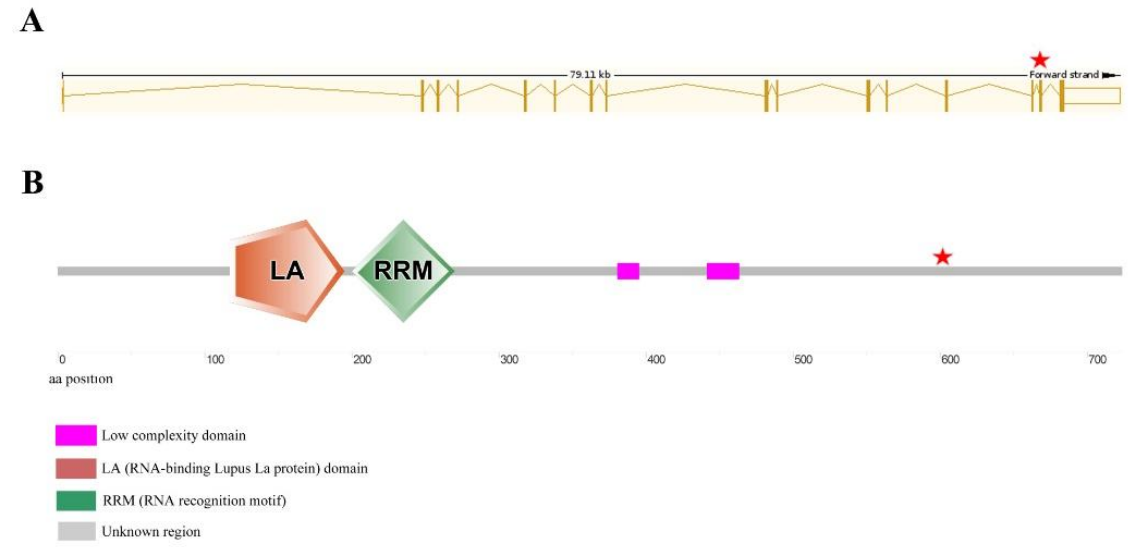


Figure S1.6 – Schematic representation of the *LARP4* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects and unknown region/domain.

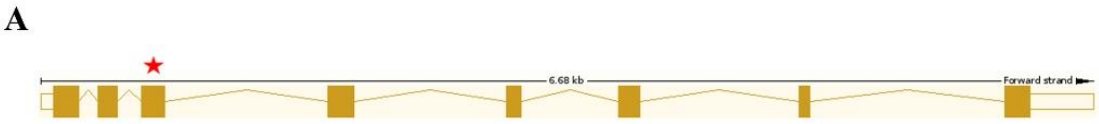


Figure S1.7 – Schematic representation of the *EIF2B2* gene structure. The gene structure was retrieved from Ensembl. The variant identified in this gene is indicated by a red star with its respective location at exon level.

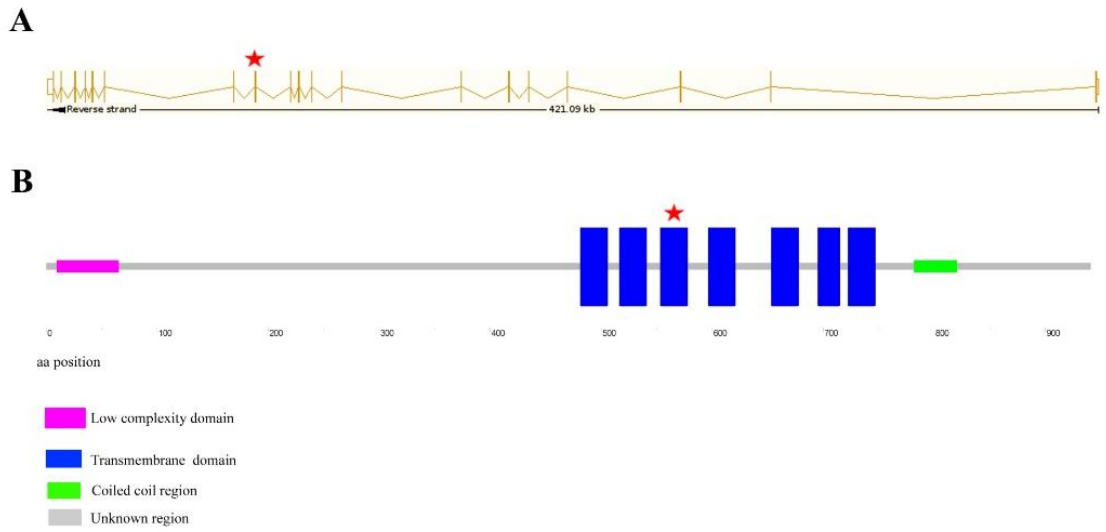


Figure S1.8 – Schematic representation of the *GABBR2* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant is located in one of the transmembrane helix regions fo the protein.

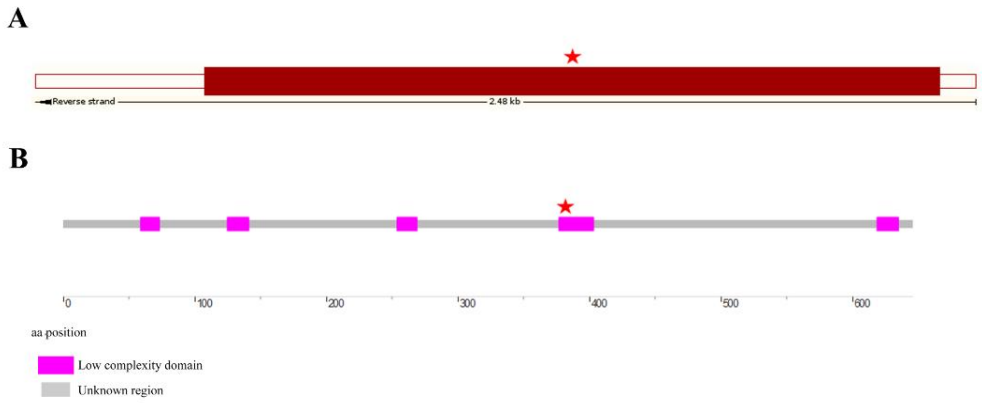


Figure S1.9 – Schematic representation of the *MAGEL2* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant is located in one of the low compositional complexity regions fo the protein.

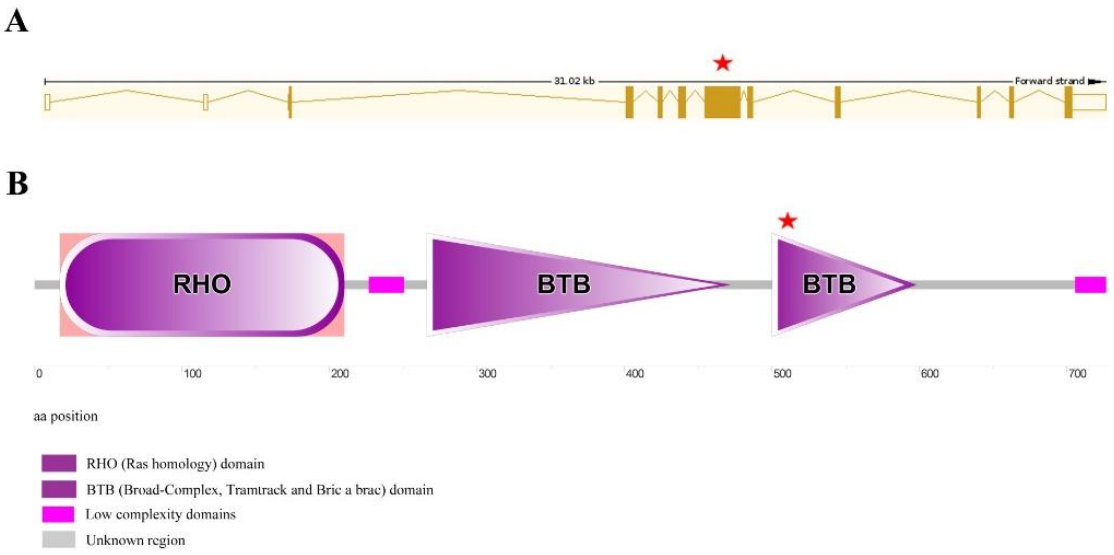


Figure S1.10 – Schematic representation of the *RHOBTB2* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant is located in the BTB domain of the protein, which is known to mediate homomeric dimerisation and, in some circunstances, heteromeric dimerization (Bardwell and Treisman, 1994).

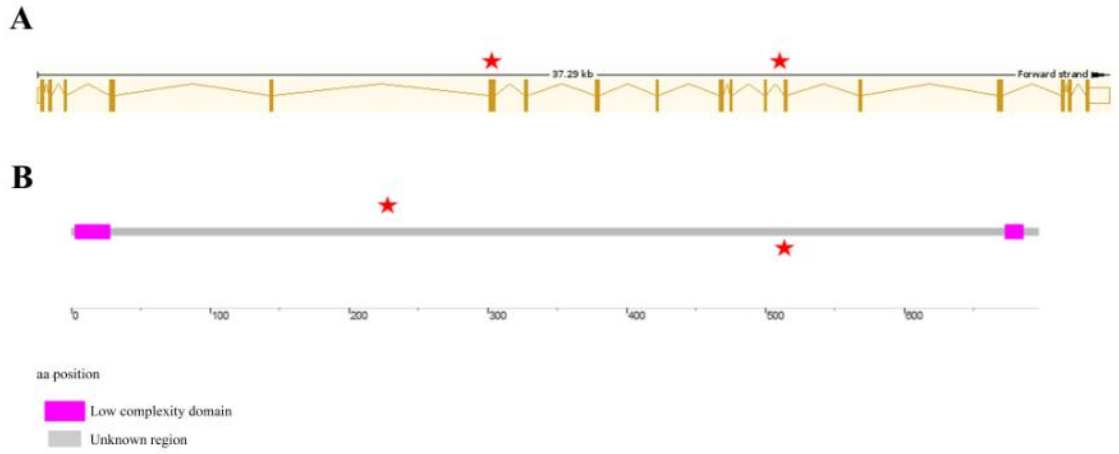


Figure S1.11 – Schematic representation of the *EIF4G1* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects and unknown region/domain.

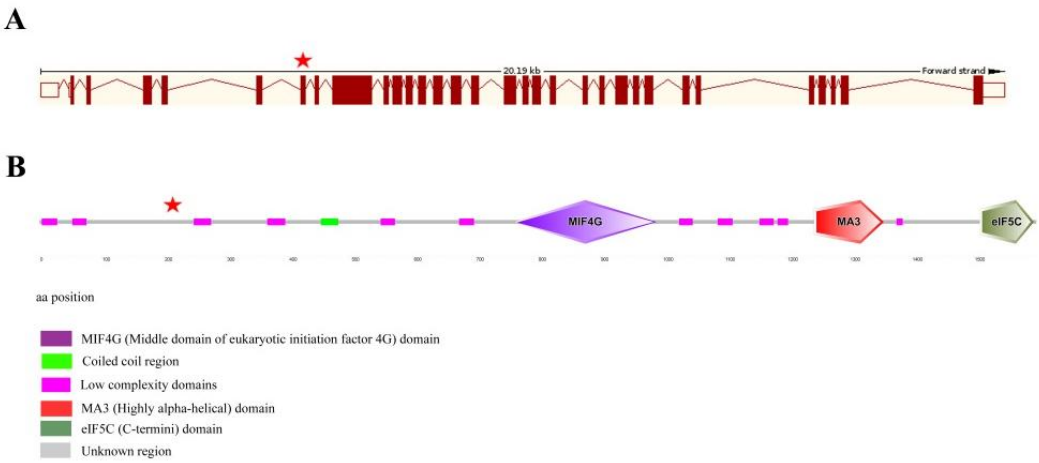


Figure S1.12 – **Schematic representation of the *TCTN2* gene and protein domains.** The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects and unknown region/domain.

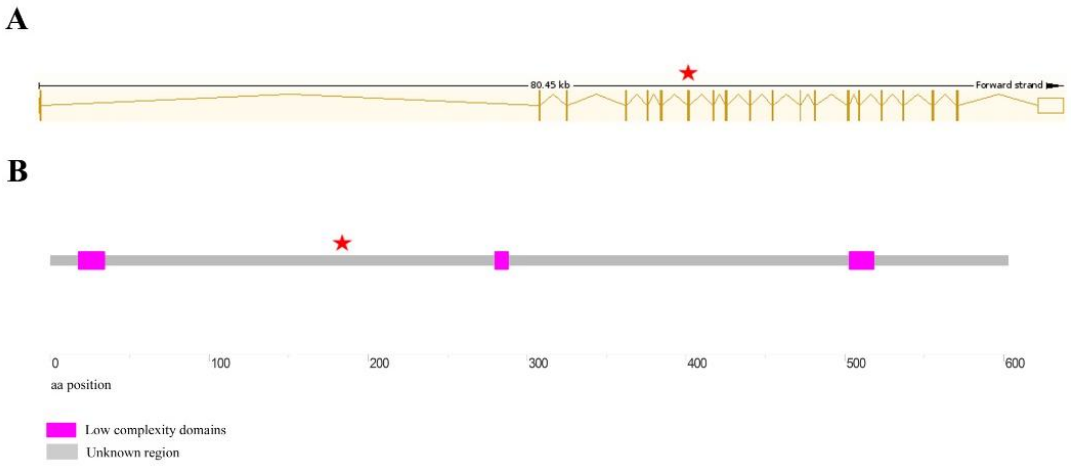


Figure S1.13 – Schematic representation of the *STXBP1* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects and unknown region/domain.

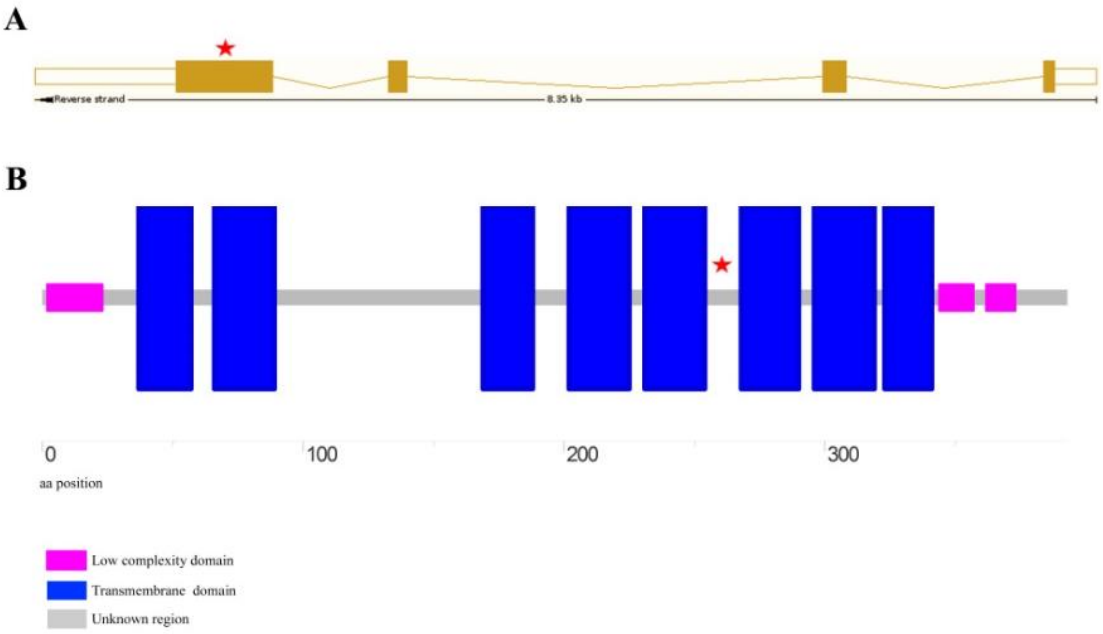


Figure S1.14 – Schematic representation of the *SLC35A2* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structure was determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects and unknown region/domain.

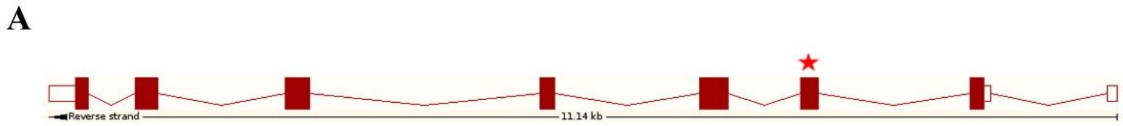


Figure S1.15 – Schematic representation of the *EEF1A2* gene structure. The gene structure was retrieved from Ensembl. The variant identified in this gene is indicated by a red star with its respective location at exon level.

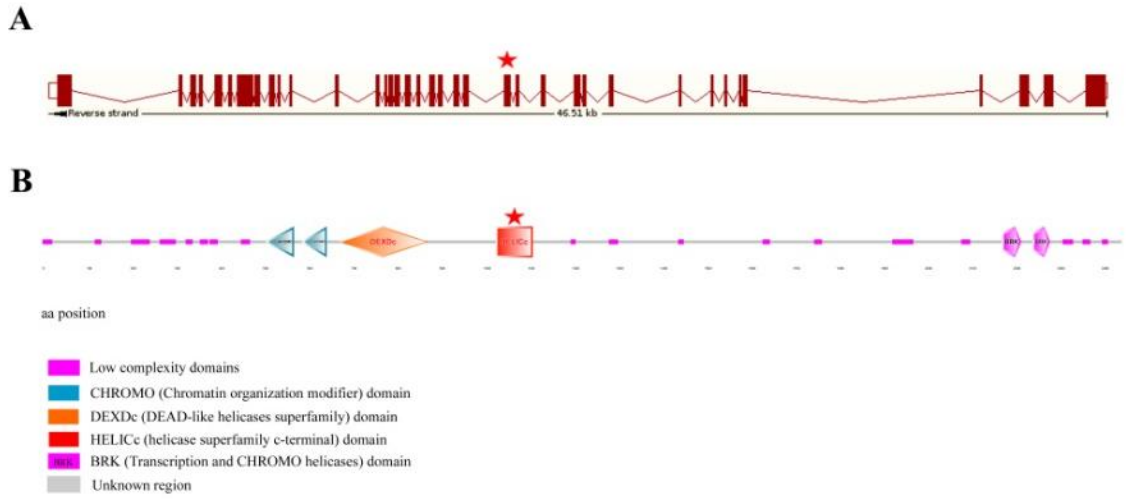


Figure S1.16 – Schematic representation of the C*HD8* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structure was determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects the HELICc domain, which corresponds to the C-terminal domain found in proteins belonging to the helicase superfamilies 1 and 2 (Gorbalenya *et al.*, 1989).

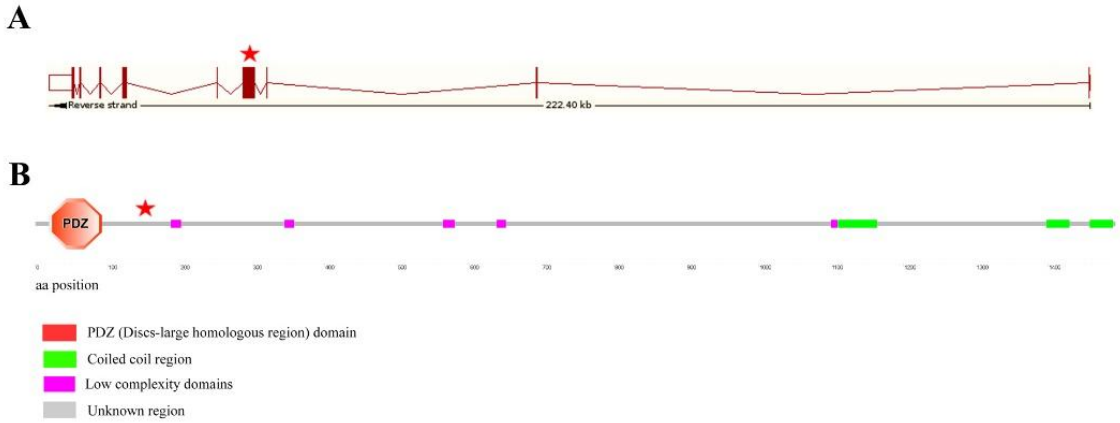


Figure S1.17 – Schematic representation of the *SHROOM4* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structure was determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects and unknown region/domain.

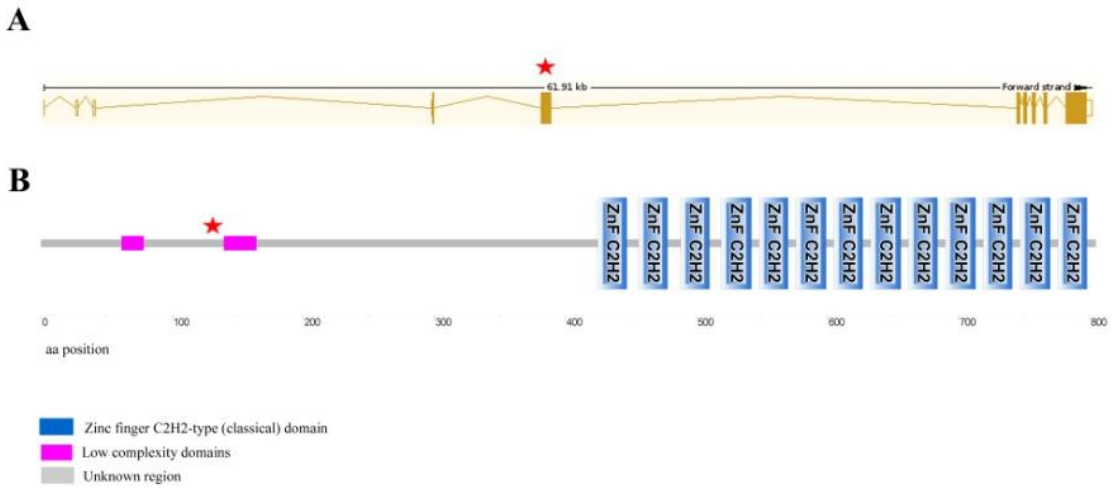


Figure S1.18 – Schematic representation of the *ZFX* gene and protein domains. The gene structure was retrieved from Ensembl and the domain structurewas determined based on SMART webtool. The variant identified in this gene is indicated by a red star with its respective location at exon and protein levels (aa). The variant affects and unknown region/domain.

Table S1.1 – Summary of the aCGH findings.

ID	Coordinates (hg19)	Size (bp)	#Probes	Algorithm*	%Rare	Type	Median LRR	Gene symbols	Comments
	chr2:194920864-194985510	64,646	12	GN,PN,QT	100%	1	-0.46	-	CNV does not encompass any genes
1	chr2:42494273-42537995	43,722	16	PN,QT	100%	3	0.42	*EML4*	Plotted on GenomeStudio:waves
	chr3:113826308-113909889	83,581	27	GN,PN,QT	100%	1	-0.37	*DRD3*	inherited from healthy father
2									Not performed
3	-	-	-	-	-	-	-	-	No relevant CNVs to report
4	chr11:38998451-39234089	235,638	23	GN,PN,QT	100%	1	-0.40	-	CNV does not encompass any genes
5									Not performed
6	chr8:22188330-22288995	100,665	31	PN,QT	100%	3	0.2	*PIWIL2; SLC39A14*	*PIWIL2* belongs to the Argonaute family (role in development and maintenance of germline stem cells); *SLC39A14* shows structural characteristics of zinc transporter. DECIPHER patient269297 has a 150kb duplication (*SLC39A14; PPP3CC*).
7	chr18:52996207-53243605	247,399	14	FASST2	100%	1	-0.8	*MIR4529, TCF4*	Loss of functions mutations and exonic or whole-gene deletions of *TCF4* cause Pitt-Hopkins syndrome
8	chr11:37476792-37532517	55,725	18	GN,PN,QT	100%	1	-0.57	-	CNV does not encompass any genes
	chr18:65848177-65897747	49,570	24	GN,PN,QT	100%	1	-0.45	-	CNV does not encompass any genes
9	-	-	-	-	-	-	-	-	No relevant CNVs to report
10	chr6:65669109-65781547	112,438	26	GN,PN,QT	100%	1	-0.41	*EYS*	*EYS* is expressed in the photoreceptor layer of the retina. Mutated in autosomal recessive retinitis pigmentosa. Likely the patietn is a carrier for retinitis pigmentosa.
	chr14:46325257-46435899	110,642	24	GN,PN,QT	100%	1	-0.44	-	CNV does not encompass any genes
11	chr21:44823479-44837555	14,076	12	PN,QT	100%	3	0.22	*SIK1*	*SIK1*: salt-inducible kinase 1. Function not well understood
12	chr4:61339224-61395456	56,232	15	PN,QT	100%	3	0.40	-	CNV does not encompass any genes
	chr13:92874384-92968290	93,906	38	GN,PN,QT	100%	1	-0.49	-	CNV does not encompass any genes
13	chr4:13102351-13221390	119,039	32	GN,PN,QT	100%	1	-0.39	-	CNV does not encompass any genes
14	-	-	-	-	-	-	-	-	No relevant CNVs to report

15	-	-	-	-	-	-	-	-	No relevant CNVs to report
16	chr3:234726-192095111	191,860,385	3032	GN,PN,QT	100%	2,3	0.11	≈1900 genes	Mosaic UPD chr3
17	chr9:5377177-5423103	45,926	22	GN,PN,QT	100%	1	-0.47	-	CNV does not encompass any genes
18	chr9:5373441-5418334	44,893	22	GN,PN,QT	100%	1	-0.51	-	CNV does not encompass any genes
18	chr16:78045800-78092197	46,397	17	GN,PN,QT	100%	1	-0.45	*CLEC3A*	*CLEC3A*: C-type lectin domain family 3, member A. Function not well understood
19	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	No relevant CNVs to report

*Algorithms: GN – GNOSIS; PN – PennCNV; QT – QuantiSNP; FASST2 - FASST2 Segmentation

Table S1.2 – Variants found by WES in patient 1.

								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	FNDC7			chr1|109265010|G|A	nonsynonymous	chr1	82	P	NP	NP	NP	NP	P	No info	Not performed
Mother	FNDC7			chr1|109265010|G|A	nonsynonymous	chr1	53	P	NP	NP	NP	NP	P	No info	Not performed
Patient	FNDC7			chr1|109273387|T|A	nonsynonymous	chr1	195	P	NP	NP	NP	NP	NP	No info	Not performed
Father	FNDC7			chr1|109273387|T|A	nonsynonymous	chr1	141	P	NP	NP	NP	NP	NP	No info	Not performed

indels_de novo

sample_id	gene	snp137	freq	indel_id	class	chr	start	end	Nervous system function	type	coverage	Sanger confirmation
Patient	NXF1			chr11|62569105|62569105|-|AGCTACAG	splicing	chr11	62569105	62569105	Yes	ins	88	Confirmed. Not *de novo* (maternal)

Table S1.3 - Variants found by WES in patient 2.

denovo

								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	RIN2	rs183028833	T=0.0009/2	chr20|19867384|C|T	nonsynonymous	chr20	16	P	NP	NP	NP	NP	NP	Yes/Likely	Confirmed. Not *de novo* (maternal)
Patient	CCRL2			chr3|46449931|A|G	nonsynonymous	chr3	75	P	P	P	P	NP	NP	Unlikely	Not performed

recessive_compound

sample_id	gene	snp135	freq	SNP_id	class	chr	coverage								Sanger confirmation
Patient	HTT			chr4|3133374|C|T	nonsynonymous	chr4	20	NP	P	P	P	NP	P	Yes	Confirmed. Present
Father	HTT			chr4|3133374|C|T	nonsynonymous	chr4	21	NP	P	P	P	NP	P	Yes	Confirmed. Present
Patient	HTT	rs34315806	T=0.0234/51	chr4|3162034|C|T	nonsynonymous	chr4	114	NP	P	Np	NP	NP	P	Yes	Confirmed. Present
Mother	HTT	rs34315806	T=0.0234/51	chr4|3162034|C|T	nonsynonymous	chr4	94	NP	P	Np	NP	NP	P	Yes	Confirmed. Present
Patient	TG	rs114944116	T=0.0023/5	chr8|133919047|G|T	nonsynonymous	chr8	6	P	P	P	P	P	NP	Unlikely	Not performed
Father	TG	rs114944116	T=0.0023/5	chr8|133919047|G|T	nonsynonymous	chr8	28	P	P	P	P	P	NP	Unlikely	Not performed
Patient	TG	rs150728539	NA	chr8|133920582|C|G	nonsynonymous	chr8	21	P	NP	P	P	NP	NP	Unlikely	Not performed
Mother	TG	rs150728539	NA	chr8|133920582|C|G	nonsynonymous	chr8	9	P	NP	P	P	NP	NP	Unlikely	Not performed

Table S1.4* - Variants found by WES in patient 3.

denovo

								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	CCDC73			chr11|32636452|A|G	nonsynonymous	chr11	23	P	NP	NP	NP	NP	NP	No info	Confirmed. Not *de novo* (maternal)
Patient	PPP1R32			chr11|61254529|A|G	nonsynonymous	chr11	21	NP	NP	NP	NP	NP	NP	Yes	Confirmed. Not *de novo* (maternal)
Patient	SYNE1	rs150179494	A=0.0014/3	chr6|152665271|G|A	nonsynonymous	chr6	27	P	P	NP	NP	NP	NP	Yes	Confirmed. Not *de novo* (paternal)
Patient	HSDL1			chr16|84163608|C|T	nonsynonymous	chr16	745	NP	NP	NP	NP	NP	NP	Unlikely	Not performed
Patient	ATG10			chr5|81354416|A|G	nonsynonymous	chr5	23	NP	NP	NP	NP	NP	NP	Yes	Not performed

recessive_compound

sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	HMCN1	rs140061598	C=0.0014/2	chr1|185956606|T|C	nonsynonymous	chr1	53								Not performed
Mother	HMCN1	rs140061598	C=0.0014/2	chr1|185956606|T|C	nonsynonymous	chr1	52								Not performed
Patient	HMCN1	rs76748242	A=0.0087/19	chr1|186031730|C|A	nonsynonymous	chr1	60								Not performed
Father	HMCN1	rs76748242	A=0.0087/19	chr1|186031730|C|A	nonsynonymous	chr1	60								Not performed
Patient	CKAP5	rs145146116	C=0.0005/1	chr11|46817323|T|C	nonsynonymous	chr11	31								Not performed
Father	CKAP5	rs145146116	C=0.0005/1	chr11|46817323|T|C	nonsynonymous	chr11	29								Not performed
Patient	CKAP5	rs138443179	T=0.0064/14	chr11|46837930|C|T	nonsynonymous	chr11	66								Not performed
Mother	CKAP5	rs138443179	T=0.0064/14	chr11|46837930|C|T	nonsynonymous	chr11	74								Not performed
Patient	AHNAK2	rs116553680	T=0.0087/19	chr14|105406201|C|T	nonsynonymous	chr14	75								Not performed
Mother	AHNAK2	rs116553680	T=0.0087/19	chr14|105406201|C|T	nonsynonymous	chr14	96								Not performed
Patient	AHNAK2	rs144504264	C=0.0114/25	chr14|105419755|G|C	nonsynonymous	chr14	15								Not performed
Father	AHNAK2	rs144504264	C=0.0114/25	chr14|105419755|G|C	nonsynonymous	chr14	15								Not performed
Patient	TRMT61B	rs144501479	NA	chr2|29073096|T|G	nonsynonymous	chr2	26								Not performed
Mother	TRMT61B	rs144501479	NA	chr2|29073096|T|G	nonsynonymous	chr2	28								Not performed
Patient	TRMT61B	rs146318990	A=0.0055/12	chr2|29092933|G|A	nonsynonymous	chr2	46								Not performed
Father	TRMT61B	rs146318990	A=0.0055/12	chr2|29092933|G|A	nonsynonymous	chr2	57								Not performed
Patient	DNAH1			chr3|52366284|C|A	nonsynonymous	chr3	39								Not performed
Father	DNAH1			chr3|52366284|C|A	nonsynonymous	chr3	55								Not performed
Patient	DNAH1	rs147123898	C=0.0046/10	chr3|52400812|A|C	nonsynonymous	chr3	23								Not performed
Mother	DNAH1	rs147123898	C=0.0046/10	chr3|52400812|A|C	nonsynonymous	chr3	24								Not performed
Patient	ARHGAP31	rs751793	T=0.0005/1	chr3|119102053|C|T	nonsynonymous	chr3	138								Not performed
Father	ARHGAP31	rs751793	T=0.0005/1	chr3|119102053|C|T	nonsynonymous	chr3	155								Not performed
Patient	ARHGAP31	rs183837502	C=0.0014/3	chr3|119133106|T|C	nonsynonymous	chr3	36								Not performed
Mother	ARHGAP31	rs183837502	C=0.0014/3	chr3|119133106|T|C	nonsynonymous	chr3	32								Not performed
Patient	SOX30	rs889057	T=0.0005/1	chr5|157053365|C|T	nonsynonymous	chr5	71								Not performed
Mother	SOX30	rs889057	T=0.0005/1	chr5|157053365|C|T	nonsynonymous	chr5	50								Not performed
Patient	SOX30	rs75818287	A=0.0055/12	chr5|157065439|G|A	nonsynonymous	chr5	25								Not performed
Father	SOX30	rs75818287	A=0.0055/12	chr5|157065439|G|A	nonsynonymous	chr5	23								Not performed
Patient	GIF	rs35867471	C=0.0174/38	chr11|59604754|T|C	nonsynonymous	chr11	88								Not performed
Father	GIF	rs35867471	C=0.0174/38	chr11|59604754|T|C	nonsynonymous	chr11	69								Not performed
Patient	GIF	rs11825834	T=0.0041/8	chr11|59611415|C|T	nonsynonymous	chr11	135								Not performed

Mother	GIF	rs11825834	T=0.0041/8	chr11|59611415|C|T	nonsynonymous	chr11	145								Not performed
Patient	TTLL5	rs141614130	A=0.0005/1	chr14|76165549|G|A	nonsynonymous	chr14	17								Not performed
Father	TTLL5	rs141614130	A=0.0005/1	chr14|76165549|G|A	nonsynonymous	chr14	25								Not performed
Patient	TTLL5	rs11848004	A=0.0101/21	chr14|76232470|G|A	nonsynonymous	chr14	22								Not performed
Mother	TTLL5	rs11848004	A=0.0101/21	chr14|76232470|G|A	nonsynonymous	chr14	29								Not performed
Patient	TNFAIP2	rs137885091	NA	chr14|103599106|C|T	nonsynonymous	chr14	45								Not performed
Mother	TNFAIP2	rs137885091	NA	chr14|103599106|C|T	nonsynonymous	chr14	52								Not performed
Patient	TNFAIP2	rs2229727	T=0.0266/57	chr14|103599847|C|T	nonsynonymous	chr14	15								Not performed
Father	TNFAIP2	rs2229727	T=0.0266/57	chr14|103599847|C|T	nonsynonymous	chr14	11								Not performed
Patient	CCDC135	rs116668060	A=0.0087/19	chr16|57738918|G|A	nonsynonymous	chr16	72								Not performed
Father	CCDC135	rs116668060	A=0.0087/19	chr16|57738918|G|A	nonsynonymous	chr16	81								Not performed
Patient	CCDC135	rs147667972	NA	chr16|57761251|C|T	nonsynonymous	chr16	24								Not performed
Mother	CCDC135	rs147667972	NA	chr16|57761251|C|T	nonsynonymous	chr16	26								Not performed
Patient	UIMC1	rs10475633	T=0.0289/62	chr5|176335585|C|T	nonsynonymous	chr5	25								Not performed
Father	UIMC1	rs10475633	T=0.0289/62	chr5|176335585|C|T	nonsynonymous	chr5	34								Not performed
Patient	UIMC1	rs199580368	C=0.0005/1	chr5|176396261|A|C	nonsynonymous	chr5	172								Not performed
Mother	UIMC1	rs199580368	C=0.0005/1	chr5|176396261|A|C	nonsynonymous	chr5	127								Not performed
Patient	ETV7			chr6|36322447|A|T	nonsynonymous	chr6	13								Not performed
Father	ETV7			chr6|36322447|A|T	nonsynonymous	chr6	23								Not performed
Patient	ETV7	rs34306145	T=0.0087/19	chr6|36339176|C|T	nonsynonymous	chr6	21								Not performed
Mother	ETV7	rs34306145	T=0.0087/19	chr6|36339176|C|T	nonsynonymous	chr6	32								Not performed
Patient	VWDE	rs116125922	A=0.0092/20	chr7|12381687|G|A	nonsynonymous	chr7	27								Not performed
Father	VWDE	rs116125922	A=0.0092/20	chr7|12381687|G|A	nonsynonymous	chr7	24								Not performed
Patient	VWDE	rs17165872	A=0.0344/75	chr7|12384089|C|A	nonsynonymous	chr7	29								Not performed
Mother	VWDE	rs17165872	A=0.0344/75	chr7|12384089|C|A	nonsynonymous	chr7	39								Not performed
Patient	CSMD1	rs199914130	NA	chr8|2966208|G|A	nonsynonymous	chr8	72								Not performed
Father	CSMD1	rs199914130	NA	chr8|2966208|G|A	nonsynonymous	chr8	75								Not performed
Patient	CSMD1	rs115021133	G=0.0106/22	chr8|3226864|C|G	nonsynonymous	chr8	31								Not performed
Mother	CSMD1	rs115021133	G=0.0106/22	chr8|3226864|C|G	nonsynonymous	chr8	27								Not performed

*Not a Caucasian patient (both parents are Angolan)

Table S1.5 – Variants found by WES in patient 4.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	SMARCA1			chrX|128599594|C|A	nonsynonymous	chrX	42	P	P	P	P	P	P	Yes	Confirmed. *De novo*
Patient	OR5V1			chr6|29323941|T|A	nonsynonymous	chr6	17	P	P	P	P	P	NP	No Info	Not performed
recessive_homoz															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	OR11H4	rs140555973	C=0.0005/1	chr14|20711688|A|C	nonsynonymous	chr14	67	P	P	P	P	NP	NP	Yes	Not performed
Mother	OR11H4	rs140555973	C=0.0005/1	chr14|20711688|A|C	nonsynonymous	chr14	75	P	P	P	P	NP	NP	Yes	Not performed
Father	OR11H4	rs140555973	C=0.0005/1	chr14|20711688|A|C	nonsynonymous	chr14	39	P	P	P	P	NP	NP	Yes	Not performed
Patient	UNKL	rs61741579	T=0.0027/6	chr16|1442911|C|T	nonsynonymous	chr16	154	P	P	NP	P	NP	NP	No Info	Not performed
Mother	UNKL	rs61741579	T=0.0027/6	chr16|1442911|C|T	nonsynonymous	chr16	166	P	P	NP	P	NP	NP	No Info	Not performed
Father	UNKL	rs61741579	T=0.0027/6	chr16|1442911|C|T	nonsynonymous	chr16	75	P	P	NP	P	NP	NP	No Info	Not performed

Table S1.6 – Variants found by WES in patient 5.

denovo								Functional impact prediction							
sample_id	**Gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	ZNF238			chr1|244217659|C|T	stopgain	chr1	12	P	P	-	-	P	P	Yes	Confirmed. *De novo*
Patient	S100PBP			chr1|33295564|G|A	splicing	chr1	19	-	-	-	-	-	-	No Info	Not performed

Table S1.7 – Variants found by WES in patient 6.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	PLEKHG6			chr12|6436851|C|T	nonsynonymous	chr12	14	NP	NP	NP	NP	NP	NP	Unknown	Confirmed. Not *de novo* (paternal)
Patient	LARP4			chr12|50867915|A|G	nonsynonymous	chr12	76	P	NP	NP	NP	NP	NP	Unknown	Confirmed. *De novo*
Patient	NEK9			chr14|75570699|A|C	nonsynonymous	chr14	34	P	NP	NP	P	P	P	No	Not performed
Patient	AP3B1			chr5|77385269|A|C	nonsynonymous	chr5	23	P	NP	NP	P	P	P	Yes	Not performed

Table S1.8 – Variants found by WES in patient 8.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	MOCOS			chr18|33775244|G|A	nonsynonymous	chr18	18	P	P	P	P	P	P	No	Not performed
recessive_homoz															
sample_id	**gene**	**snp135**		**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	EIF2B2	rs150617429	T=0.0023/5	chr14|75470349|C|T	nonsynonymous	chr14	63	NP	NP	NP	NP	NP	P	Yes	Confirmed. In homozygosity
Mother	EIF2B2	rs150617429	T=0.0023/5	chr14|75470349|C|T	nonsynonymous	chr14	69	NP	NP	NP	NP	NP	P	Yes	Confirmed. In heterozygosity
Father	EIF2B2	rs150617429	T=0.0023/5	chr14|75470349|C|T	nonsynonymous	chr14	67	NP	NP	NP	NP	NP	P	Yes	Confirmed. In heterozygosity

Table S1.9 – Variants found by WES in patient 9.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	RCC2			chr1|17736540|C|T	nonsynonymous	chr1	8	P	P	P	P	P	P	Yes	Confirmed. Not *de novo* (maternal)
Patient	GABBR2			chr9|101133817|C|T	nonsynonymous	chr9	7	P	NP	P	NP	NP	P	Yes	Confirmed. *De novo*
Patient	H2AFY2	rs149708840	T=0.0037/8	chr10|71871347|G|T	nonsynonymous	chr10	25	NP	NP	NP	NP	NP	NP	Yes	Not performed
Patient	C10orf96			chr10|118084833|G|C	nonsynonymous	chr10	15	P	P	P	NP	P	NP	No Info	Not performed
Patient	NCF4			chr22|37271932|A|C	nonsynonymous	chr22	29	P	NP	-	P	NP	NP	No Info	Not performed
Patient	FYCO1	rs114145679	T=0.0064/14	chr3|46009810|C|T	nonsynonymous	chr3	18	NP	NP	-	-	NP	NP	Yes	Not performed
recessive_compound															
sample_id	**gene**	**snp135**		**SNP_id**	**class**	**chr**	**coverage**								**Confirmação por Sanger**
Patient	TTN			chr2|179414062|G|T	nonsynonymous	chr2	108	-	-	-	-	-	NP	No Info	Not performed
Father	TTN			chr2|179414062|G|T	nonsynonymous	chr2	145	-	-	-	-	-	NP	No Info	Not performed
Patient	TTN	rs34819099	T=0.005/11	chr2|179628918|C|T	nonsynonymous	chr2	27	NP	NP	-	-	-	P	No Info	Not performed
Mother	TTN	rs34819099	T=0.005/11	chr2|179628918|C|T	nonsynonymous	chr2	25	NP	NP	-	-	-	P	No Info	Not performed

Table S1.10 – Variants found by WES in patient 10.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	MAGEL2			chr15|23889919|C|T	nonsynonymous	chr15	27	-	-	-	-	-	-	Yes	Performed. Not *de novo* (paternal)
Patient	RPGRIP1L			chr16|53636037|T|C	nonsynonymous	chr16	4	NP	NP	NP	NP	P	NP	Yes	Performed. Not *de novo* (paternal)
Patient	TMEM18			chr2|669832|C|G	nonsynonymous	chr2	18	P	P	P	P	NP	P	Yes	Performed. Not *de novo* (maternal)
Patient	MAST4	rs115056755	T=0.0018/4	chr5|66461954|C|T	nonsynonymous	chr5	6	NP	NP	NP	NP	NP	NP	Yes	Performed. Not *de novo* (maternal)
Patient	TMCC3			chr12|94976229|G|A	nonsynonymous	chr12	29	NP	NP	NP	NP	NP	P	No Info	Not performed

Table S1.11 – Variants found by WES in patient 11.

								Functional impact prediction							
denovo															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	LAMB2			chr3|49160344|C|T	nonsynonymous	chr3	7	P	P	NP	P	NP	P	Yes	Confirmed. Not *de novo* (maternal)
Patient	RHOBTB2			chr8|22865220|A|G	nonsynonymous	chr8	70	P	P	NP	P	NP	NP	Yes	Confirmed. *De novo*
Patient	MARCH8			chr10|46028610|A|G	nonsynonymous	chr10	35	NP	NP	NP	NP	NP	NP	Unknown	Not performed
Patient	TTC40			chr10|134736171|G|A	nonsynonymous	chr10	18	P	-	-	-	-	-	No Info	Not performed
Patient	TAS2R30			chr12|11286282|G|A	nonsynonymous	chr12	40	-	-	NP	NP	NP	-	No Info	Not performed
Patient	CIITA			chr16|11001853|T|C	nonsynonymous	chr16	21	P	P	NP	P	P	P	Unknown	Not performed
Patient	ALPPL2			chr2|233273068|T|G	nonsynonymous	chr2	46	NP	NP	NP	NP	NP	NP	Unknown	Not performed
Patient	BCR			chr22|23615846|G|A	nonsynonymous	chr22	14	NP	NP	NP	NP	NP	P	Unlnown	Not performed
Patient	SLC36A2			chr5|150696620|C|T	nonsynonymous	chr5	19	P	P	NP	NP	NP	P	Uknown	Not performed
Patient	TINAG			chr6|54186172|A|G	nonsynonymous	chr6	38	NP	NP	NP	NP	NP	NP	Unknown	Not performed
recessive_homoz															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	EIF4G1	rs34838305	T=0.0005/1	chr3|184038482|G|A	nonsynonymous	chr3	55	P	P	NP	P	NP	P	Yes	Confirmed. In homozygosity
Mother	EIF4G1	rs34838305	T=0.0005/1	chr3|184038482|G|A	nonsynonymous	chr3	60	P	P	NP	P	NP	P	Yes	Confirmed. In heterozygosity
Father	EIF4G1	rs34838305	T=0.0005/1	chr3|184038482|G|A	nonsynonymous	chr3	61	P	P	NP	P	NP	P	Yes	Confirmed. In heterozygosity

Table S1.12 – Variants found by WES in patient 12.

								Functional impact prediction							
denovo															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	EEF1A2			chr20|62120361|T|G	nonsynonymous	chr20	21	P	P	P	P	P	P	Yes	Confirmed. Not *de novo* (maternal)
Patient	DNAH1	rs61734653	G=0.005/11	chr3|52381865|A|G	nonsynonymous	chr3	11	NP	NP	NP	NP	NP	NP	Yes	Confirmed. Not *de novo* (maternal)
Patient	PRDM5			chr4|121739655|G|A	nonsynonymous	chr4	57	NP	NP	NP	NP	NP	NP	Yes	Not performed

Tables S1.13 – Variants found by WES in patient 13.

denovo

								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	RRAS2			chr11|14317388|G|A	nonsynonymous	chr11	20	P	NP	NP	NP	NP	P	Yes	Confirmed. Not *de novo* (maternal)
Patient	BNIP2	rs142043306	NA	chr15|59961478|A|G	nonsynonymous	chr15	11	P	P	-	P	NP	P	Yes	Confirmed. Not *de novo* (paternal)
Patient	MTERFD3			chr12|107371889|C|T	nonsynonymous	chr12	17	P	NP	NP	NP	NP	P	Unknown	Not performed
Patient	ENAM			chr4|71508671|C|A	nonsynonymous	chr4	31	P	P	P	NP	NP	NP	Unknown	Not performed

recessive_homoz

sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	PSMD1			chr2|231948329|A|G	nonsynonymous	chr2	45	NP	NP	NP	NP	NP	NP	No Info	Confirmed. Present in homozygosity
Mother	PSMD1			chr2|231948329|A|G	nonsynonymous	chr2	67	NP	NP	NP	NP	NP	NP	No Info	Confirmed. Present in heterozygosity
Father	PSMD1			chr2|231948329|A|G	nonsynonymous	chr2	87	NP	NP	NP	NP	NP	NP	No Info	Confirmed. Present in heterozygosity
Patient	CBR1	rs146758729	A=0.0005/1	chr21|37444929|G|A	nonsynonymous	chr21	14	NP	NP	NP	NP	P	P	Yes	Not performed
Mother	CBR1	rs146758729	A=0.0005/1	chr21|37444929|G|A	nonsynonymous	chr21	31	NP	NP	NP	NP	P	P	Yes	Not performed
Father	CBR1	rs146758729	A=0.0005/1	chr21|37444929|G|A	nonsynonymous	chr21	42	NP	NP	NP	NP	P	P	Yes	Not performed

recessive_compound

sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	TCTN2	rs145374149	T=0.0009/2	chr12|124171486|C|T	nonsynonymous	chr12	56	NP	NP	NP	NP	NP	NP	Yes	Not performed
Mother	TCTN2	rs145374149	T=0.0009/2	chr12|124171486|C|T	nonsynonymous	chr12	99	NP	NP	NP	NP	NP	NP	Yes	Not performed
Patient	TCTN2			chr12|124184283|T|C	nonsynonymous	chr12	51	NP	NP	NP	NP	P	NP	Yes	Not performed
Father	TCTN2			chr12|124184283|T|C	nonsynonymous	chr12	114	NP	NP	NP	NP	P	NP	Yes	Not performed
Patient	PLCH1	rs150381264	T=0.0018/4	chr3|155199598|C|T	nonsynonymous	chr3	41	NP	NP	NP	NP	NP	NP	Yes	Not performed
Mother	PLCH1	rs150381264	T=0.0018/4	chr3|155199598|C|T	nonsynonymous	chr3	65	NP	NP	NP	NP	NP	NP	Yes	Not performed
Patient	PLCH1	rs150143990	A=0.0046/10	chr3|155314114|G|A	nonsynonymous	chr3	17	NP	NP	NP	P	P	NP	Yes	Not performed
Father	PLCH1	rs150143990	A=0.0046/10	chr3|155314114|G|A	nonsynonymous	chr3	50	NP	NP	NP	P	P	NP	Yes	Not performed

Table S1.14 – Variants found by WES in patient 14.

								Functional impact prediction							
denovo															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	STXBP1			chr9|130425592|T|C	nonsynonymous	chr9	83	P	P	P	P	P	P	Yes	Confirmed. *De novo*
recessive_compound															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	HPX	rs150488733	NA	chr11|6452695|G|A	nonsynonymous	chr11	56	P	P	NP	P	P	P	Yes	Not performed
Father	HPX	rs150488733	NA	chr11|6452695|G|A	nonsynonymous	chr11	77	P	P	NP	P	P	P	Yes	Not performed
Patient	HPX			chr11|6459640|C|A	nonsynonymous	chr11	68	P	NP	P	P	NP	P	Yes	Not performed
Mother	HPX			chr11|6459640|C|A	nonsynonymous	chr11	63	P	NP	P	P	NP	P	Yes	Not performed
Patient	HSD17B8			chr6|33172705|G|C	nonsynonymous	chr6	12	P	NP	NP	NP	NP	P	Yes	Not performed
Mother	HSD17B8			chr6|33172705|G|C	nonsynonymous	chr6	5	P	NP	NP	NP	NP	P	Yes	Not performed
Patient	HSD17B8	rs116381506	T=0.0046/10	chr6|33173457|C|T	nonsynonymous	chr6	338	P	P	P	P	P	P	Yes	Not performed
Father	HSD17B8	rs116381506	T=0.0046/10	chr6|33173457|C|T	nonsynonymous	chr6	319	P	P	P	P	P	P	Yes	Not performed
xlinked															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	GPR64			chrX|19024128|C|A	nonsynonymous	chrX	8	P	NP	-	-	NP	NP	Yes	Not performed
Mother	GPR64			chrX|19024128|C|A	nonsynonymous	chrX	9	P	NP	-	-	NP	NP	Yes	Not performed
Patient	FAM48B2			chrX|24329737|C|A	nonsynonymous	chrX	5	-	-	-	-	P	-	No info	Not performed
Mother	FAM48B2			chrX|24329737|C|A	nonsynonymous	chrX	16	-	-	-	-	P	-	No info	Not performed

Table S1.15 – Variants found by WES in patient 15.

								Functional impact prediction							
denovo															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	CHPF			chr2|220404347|C|G	nonsynonymous	chr2	16	NP	NP	NP	NP	NP	NP	Yes	Confirmed. Not *de novo* (paternal)
Patient	ZDBF2			chr2|207172196|A|G	nonsynonymous	chr2	53	P	P	NP	P	P	NP	Unknown	Not performed
recessive_compound															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Confirmação por Sanger**
Patient	CAGE1	rs45437691	A=0.0005/1	chr6|7378978|G|A	nonsynonymous	chr6	74	NP	-	-	-	NP	NP	Unknown	Not performed
Father	CAGE1	rs45437691	A=0.0005/1	chr6|7378978|G|A	nonsynonymous	chr6	81	NP	-	-	-	NP	NP	Unknown	Not performed
Patient	CAGE1	rs183206380	A=0.0009/2	chr6|7379149|C|A	nonsynonymous	chr6	71	NP	-	-	NP	NP	NP	Unknown	Not performed
Mother	CAGE1	rs183206380	A=0.0009/2	chr6|7379149|C|A	nonsynonymous	chr6	47	NP	-	-	NP	NP	NP	Unknown	Not performed
xlinked															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Confirmação por Sanger**
Patient	SLC35A2			chrX|48762414|C|T	nonsynonymous	chrX	12	P	P	NP	P	NP	P	Yes	Confirmed. Maternal (X-linked)
Mother	SLC35A2			chrX|48762414|C|T	nonsynonymous	chrX	19	P	P	NP	P	NP	P	Yes	Confirmed. In heterozigosity

Table S1.16 – Variants found by WES in patient 17.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	DDX23			chr12|49230558|G|A	nonsynonymous	chr12	120	P	P	P	P	P	P	Unknown	Condirmed. *De novo*
Patient	EEF1A2			chr20|62127259|C|T	nonsynonymous	chr20	37	P	P	NP	P	NP	P	Yes	Condirmed. *De novo*
Patient	LIN9			chr1|226475489|C|T	nonsynonymous	chr1	20	P	NP	NP	NP	NP	NP	Unknwon	Not performed
Patient	L3MBTL3			chr6|130460830|G|C	nonsynonymous	chr6	29	P	P	P	P	P	P	Unknwon	Not performed
recessive_compound															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	AHNAK			chr11|62290815|A|G	nonsynonymous	chr11	115	P	P	NP	P	P	NP	Yes	Not performed
Father	AHNAK			chr11|62290815|A|G	nonsynonymous	chr11	117	P	P	NP	P	P	NP	Yes	Not performed
Patient	AHNAK			chr11|62298843|G|A	nonsynonymous	chr11	124	P	NP	NP	NP	NP	NP	Yes	Not performed
Mother	AHNAK			chr11|62298843|G|A	nonsynonymous	chr11	120	P	NP	NP	NP	NP	NP	Yes	Not performed

Table S1.17 – Variants found by WES in patient 18.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	ASPM			chr1|197097663|C|T	nonsynonymous	chr1	13	NP	NP	NP	NP	NP	P	Yes	Confirmed. Not *de novo* (maternal)
Patient	CHD8			chr14|21873449|C|A	nonsynonymous	chr14	35	P	NP	NP	NP	NP	P	Yes	Confirmed. Not *de novo* (maternal)
Patient	LMOD1			chr1|201915453|T|C	nonsynonymous	chr1	85	P	P	NP	NP	NP	NP	Unknown	Not performed
Patient	LAMP1			chr13|113975974|C|T	nonsynonymous	chr13	60	NP	NP	NP	NP	NP	NP	Yes	Not performed

Table S1.18 – Variants found by WES in patient 19.

denovo								Functional impact prediction							
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**	**SIFT**	**PolyPhen2**	**MutAsse.**	**Condel**	**Pmut**	**MutTast.**	**Nervous system function**	**Sanger confirmation**
Patient	SPRED2			chr2|65659108|G|T	nonsynonymous	chr2	17	NP	NP	NP	NP	P	P	Yes	Not performed
Patient	SLC38A9	rs35800744	T=0.005/11	chr5|54945091|C|T	nonsynonymous	chr5	38	NP	NP	NP	NP	NP	NP	No Info	Not performed
Patient	TTC37			chr5|94856512|T|A	nonsynonymous	chr5	15	P	P	P	P	NP	P	Unknown	Not performed
Patient	DTNBP1			chr6|15651554|A|T	nonsynonymous	chr6	15	P	NP	NP	NP	NP	P	Yes	Not performed
xlinked															
sample_id	**gene**	**snp135**	**freq**	**SNP_id**	**class**	**chr**	**coverage**								**Sanger confirmation**
Patient	ZFX	rs150375972	A=0.0006/1	chrX|24197650|G|A	nonsynonymous	chrX	41	P	NP	NP	NP	-	NP	Yes	Confirmed. Maternal (X-linked)
Mother	ZFX	rs150375972	A=0.0006/1	chrX|24197650|G|A	nonsynonymous	chrX	93	P	NP	NP	NP	-	NP	Yes	Confirmed. In heterozigosity
Patient	SHROOM4	rs189694750	A=0.003/5	chrX|50378637|G|A	nonsynonymous	chrX	12	P	P	NP	P	P	P	Yes	Confirmed. Maternal (X-linked)
Mother	SHROOM4	rs189694750	A=0.003/5	chrX|50378637|G|A	nonsynonymous	chrX	27	P	P	NP	P	P	P	Yes	Confirmed. In heterozigosity
Patient	PHKA2			chrX|18943832|G|A	nonsynonymous	chrX	11	P	P	P	P	P	P	No	Not performed
Mother	PHKA2			chrX|18943832|G|A	nonsynonymous	chrX	39	P	P	P	P	P	P	No	Not performed
Patient	ACOT9			chrX|23722015|G|A	nonsynonymous	chrX	17	P	NP	NP	P	P	-	Unknown	Not performed
Mother	ACOT9			chrX|23722015|G|A	nonsynonymous	chrX	70	P	NP	NP	P	P	-	Unknown	Not performed
Patient	HDX			chrX|83695581|G|C	nonsynonymous	chrX	30	P	NP	NP	P	NP	P	No Info	Not performed
Mother	HDX			chrX|83695581|G|C	nonsynonymous	chrX	47	P	NP	NP	P	NP	P	No Info	Not performed

Supplementary table S1.19 – Genes harboring variants submitted confirmation by to Sanger sequencing

Trio	#Genes with variants selected for Sanger confirmation	Genes with variants selected for Sanger confirmation	#Genes with variants confirmed by Sanger	Gene with variants confirmed by Sanger
1	5	*PSIP1, TRAPPC6A, WASF2, PICALM, NXF1*	5	*PSIP, TRAPPC6A, WASF2, PICALM, NXF1*
2	4	*FXR2, HTT, HEXDC, RIN2*	3	*HTT, HEXDC, RIN2*
3	8	*CCDC73, PPP1R32, GAL, EXOC3L4, CACNA1H, KATNB1, CELSR3, SYNE1*	8	*CCDC73, PPP1R32, GAL, EXOC3L4, CACNA1H, KATNB1, CELSR3, SYNE1*
4	1	*SMARCA1*	1	*SMARCA1*
5	1	*ZNF238*	1	*ZNF238*
6	2	*PLEKHG6, LARP4*	2	*PLEKHG6, LARP4*
8	2	*SLC26A4, EIF2B2*	1	*EIF2B2*
9	3	*GABBR2, RCC2, ITGA7*	2	*GABBR2, RCC2*
10	6	*MAGEL2, DROSHA, TMEM18, DKK3, RPGRIP1L, MAST4*	4	*MAGEL2, TMEM18, RPGRIP1L, MAST4*
11	4	*JMJD1C, RHOBTB2, LAMB2, EIF4G1*	3	*RHOBTB2, LAMB2, EIF4G1*
12	5	*DDX23, EEF1A2, LEPROTL1, DNAH1, TCOF1*	2	*EEF1A2, DNAH1*
13	3	*BNIP2, PSMD1, RRAS2*	3	*BNIP2, PSMD1, RRAS2*
14	1	*STXBP1*	1	*STXBP1*
15	3	*SLC35A2, KCNT2, CHPF*	2	*SLC35A2, CHPF*
17	2	*EEF1A2, DDX23*	2	*EEF1A2, DDX23*
18	3	*CHD8, ASPM, CADPS2*	2	*CHD8, ASPM*
19	3	*SHROOM4, CACNA1G, ZFX*	2	*SHROOM4, ZFX*

Part 2 - Detailed methodology

Array comparative genomic hybridization analysis

Agilent 180K

The aCGH analysis was performed on a human genome CGH Agilent 180K custom array designed by the Low Lands Consortium (LLC, Professor Klass Kok), specifically tailored for the study of children with ID/DD (AMADID:023363; Agilent, Santa Clara, CA). A total of 500 ng of DNA was labeled with Cy3 (test DNA) and Cy5 (reference DNA) using Klenow fragment at 37°C during 4 hours (http://www.enzolifesciences.com/ENZ-42671/cgh-labeling-kit-for-oligo-arrays/). The removal of uncoupled nucleotides was carried out in a MinElute PCR purification kit (Qiagen) according to the manufacturer´s instructions (http://www.qiagen.com/Products/DnaCleanup/GelPcrSiCleanupSystems/MinEluteReactionCleanupKit.aspx?r=1644). As reference DNA we used Kreatech´s MegaPoll Reference DNA (Kreatech Diagnostics). Arrays were then hybridized using the Agilent SurePint G3 Human CGH Microarray Kit. Briefly, samples and controls were hybridized in the presence of Cot-1 DNA and blocking agents for 24 hours, 65°C and 20rpm. After the hybridization period, the slides were washed and scanned with Agilent Microrray Scanner. Data was extracted with the Agilent Feature Extraction (FE) Software v10.5 using default settings for CGH hybridizations. Image analysis was performed using the across-array methodology described previously (Buffart *et al.*, 2008). CGH data was analyzed using Nexus Copy Number 6.0 software with FASST2 Segmentation algorithm and a minimum of three probes in a region required to be considered a copy number alteration. The stipulated minimal thresholds for calling a copy number gain were 0.2 (Copy number Gain) and 0.6 (High Copy Gain) and for calling a copy number loss were -0.2 (Copy number Loss) and -1 (Homozygous Copy Loss).

Illumina HumanOmniExpress

The microarray analysis was performed on a Illumina HumanOmniExpress beadchip array (WG312-1120; Illumina, San Diego, CA). The array contained about 730.525 markers probes with a mean and median resolution of one probe every 4.0kb and 2.1kb, respectively. Samples were processed using the manufacturer's recommended assay: Infinium® HD assay. Using this protocol, around 750 ng each of DNA samples were denatured and neutralized to prepare them for amplification. The denatured DNA was isothermally amplified in an overnight incubation at 37°C. The amplified product was enzymatically fragmented to 300-600 base pairs, purified by isopropanol precipitation and later resuspended. Hybridization of sheared DNA to the bead chip

was carried out in a capillary flow through chamber. Samples were pipetted into the BeadChip and incubated overnight at 48°C in the Illumina Hybridization Oven. Only the labeled probes are left on the beads. Green fluorescent Streptavidin and red fluorescent Anti-DNP Antibody are used to bind specifically to the labeled probes. The beads arrays are scanned using Illumina HiScan which has a laser that excites the fluorophore of the single-base extension product on the beads. The fluorescence intensity data generated is used to score genotype calls for each SNP. The scanned data are analyzed on Genome Analyzer software provided by Illumina, Inc.

CNVision (Sanders *et al.*, 2011), a pipeline developed for Illumina microarray data, was used to analyze CNVs using three algorithms with two different approaches: PennCNV (Wang *et al.*, 2007), QuantiSNP (Colella *et al.*, 2007) and GNOSIS (Sanders *et al.*, 2011). GNOSIS is a distribution function algorithm that uses sliding windows to detect CNVs whereas PennCNV and QuantiSNP both implement a hidden Markov model. All algorithms use logRratio (LRR; a measure of total signal intensity of probes) and B allele frequency (BAF; a measure of relative intensity ratio of allelic probes) information from the genotyping arrays to detect CNVs. Priority was given to CNVs with > 40kb (i.e., > 10 probes), rare (frequency<1% in DGV) and *de novo,* called by both QuantiSNP and PennCNV. Accuracy of CNVs detected *in silico* was assessed by visual inspection of LRR and BAF plots in GenomeStudio software, with its built-in algorithm - cnvPartition (Illumina, 2013). CNV interpretation took into consideration the size and number of genes, as well as presence of similar CNVs in databases such as ISCA ('Welcome to the ISCA Consortium web site', n.d.) https://www.iscaconsortium.org/) and DECIPHER (Firth *et al.*, 2009) and information in OMIM and PubMed.

CNV interpretation

Interpretation of the CNVs found was carried out based on the workflow proposed by (Edelmann and Hirschhorn, 2009). For each patient the total number of CNVs was listed according to the position in the chromosome and classified according to the workflow represented in figure S2.1.

Quantitative PCR analysis

Primers for quantitative PCR analysis (qPCR) were designed using Primer3Plus software ('Primer3Plus', n.d.) (listed in table 22) and taking into account standard recommendations for qPCR primer development. A set of primers were designed for the *TFC4* gene (ENSG00000196628). The reference genes used were *SDC4* (ENSG00000124145) and *ZNF80* (ENSG00000174255) localized in the 20q12-q13 and 3p12 regions, respectively (primers table S2.1). qPCR reactions were carried out in a 7500-FAST Real Time PCR machine (Applied

Biosystems, Foster City, CA, USA) using Power SYBR Green® (Applied Biosystems) for a final volume of 20 µl, at the following amplifications conditions: denaturation for 10 minutes at 95°C, followed by 40 amplification cycles for 15 seconds at 60°C. The specificity of each of the reactions was verified by the generation of a melting curve for each of the amplified fragments. The primer efficiency was calculated by the generation of a standard curve fitting the accepted normal efficiency percentage. Quantification was performed as described elsewhere (Hoebeeck *et al.*, 2005). Ct values obtained for each test were analyzed in DataAssist ™ software from Applied Biosystems.

Exome sequencing and data analysis

Library preparation and SOLiD sequencing

Exome enrichment was performed using 3 µg genomic DNA. DNA samples were sheared by sonication with the Covaris S2 instrument (Covaris, Inc.). Fragment libraries were constructed from the sheared samples using the AB Library Builder System (Life Technologies) and target enrichment was performed according to the manufacturer's protocols (Agilent SureSelect Human All Exon v4 kit). The Agilent SureSelect Human All Exon v4 kit was used for enrichment, containing exonic sequences of 20.965 genes (corresponding to a total of 334.378 exons), covering a total of 51Mb of genomic sequence (as specified by the company). Exome capture was conducted by hybridizing the DNA libraries with biotinylated RNA baits for 24 h followed by extraction using streptavidin coated magnetic beads. Captured DNA was then amplified followed by emulsion PCR using the EZ Bead System (Life Technologies) and sequenced on the SOLiD5500xl system, generating over 40 million reads of length 75bp for each of the samples. Individual libraries were labeled by a post-hybridization barcoding procedure (Agilent, Santa Clara, CA, USA; barcodes compatible with SOLiD sequencing technology).

Mapping and variant calling

Alignment of color space reads to the human reference genome (hg19) was performed using v2.1 of the Lifescope Software (Life Technologies). Around 90% of the reads from each sample were uniquely mapped to the target regions, generating an average coverage of 35X and a median of 30x over the targeted exons across samples. Similar coverage has been described in other studies using a Agilent SureSelect human exome enrichment kit and SOLiD sequencing platform (Life Technologies) (Hoischen *et al.*, 2010, p. 1). Single nucleotide variants (SNVs) and small insertions and deletions (indels) were subsequently called by the diBayes algorithm available within the Lifescope software. At least 10 reads are required to obtain a 99% probability

that at least two reads contain the variant allele (assuming a binomial distribution with probability 0.5 of sequencing the variant allele at a heterozygous position. In our cohort, 80% of the targets were covered at least 10 fold (figure S2.2). Single Nucleotide Variants (SNVs) and indels were filtered as follows: exclusion of synonymous and common variants (MAF>1% dbSNP35, Exome Sequencing Project and 1000 genomes Project), variants present in the CanvasDB in-house exome database(Ameur *et al.*, 2014) in Uppsala (excluding individuals with neurodevelopmental phenotypes), presence of at least 3 unique variant reads (i.e. different start sites), as well as the variant being present in at least 15% of all reads. All called SNVs and indels were imported into a local installation of the CanvasDB database system (Ameur *et al.*, 2014) for annotation and further analysis of the variants.

Analysis pipeline

To avoid false positives (i.e., variant calling in duplicate reads due to PCR artifacts), only variants present in reads with at least 3 different starts were used in downstream analysis. Then, we excluded intronic or synonymous variants. Next we filtered out variants present in dbSNPv135 or in our in-house variant database. Our database contains variants from 500 in-house performed 'exomes' analyzed by the same pipeline. The inhouse database is composed of aproximately 300 healthy individuals and 200 patients with non-neurodevelopmental disorders. Variants with frequency >1% in the 6500 individuals of the Exome Variant Server/National Heart, Lung and Blood Institute ('Exome Variant Server', n.d.) or in the 1000 Genomes Project (1000 Genomes Project Consortium *et al.*, 2012) were also excluded This step further reduced the average number of variants (including stopgain, stoploss, missense, splicing, frameshift, in frame) to 26 per patient (which we called "private" variants).

Prioritization of candidate variants

The exome data from both parents was used to provide the parent-of-origin of each candidate variant. Variants that were not identified in either parent with >15% variation reads were considered to be candidate *de novo* variants.

For the male patients the variants located at the X-chromosome and of maternal inheritance were classified as such, allowing an additional filter step for X-linked maternally-inherited variants. For recessive analysis, non-synonymous homozygous, non-synonymous compound heterozygous variants and/or canonical splice site variants in the same gene were selected for evaluation, with each parent being carrier of one of the variants.

Variants were selected for Sanger confirmation if (I) Integrative Genomics Viewer(Robinson *et al.*, 2011) visualization confirmed that the position where the variant occurred was sufficiently covered and matched the pre-defined inheritance model, (II) the variant occurred in a known or candidate ID gene or a gene known to be functionally relevant for the nervous system, (III) the effect of the variant on the gene was predicted to be deleterious. Gene prioritization took into consideration information on biological function available in the literature (PubMed search) and in the OMIM, Gene Entrez, and GeneCards databases, genetic/protein interactions,(Warde-Farley, Sylva L Donaldson, *et al.*, 2010) brain expression(Michael J Hawrylycz *et al.*, 2012; 'Microarray Data :: Allen Brain Atlas: Human Brain', n.d.) and KO mice phenotype.(Judith A Blake *et al.*, 2014)

We also took into consideration the type of variant and gave preference to loss of function variants (i.e., nonsense, splice site, frameshift) followed by missense and in frame variants. Guidance was also obtained from six prediction softwares, which was particularly helpsulf on evauating missense variants. The prediction softwares and cutoffs used were: SIFT (damaging if SIFT score<=0.05) (Ng and Henikoff, 2001), PolyPhen2 (damaging if PolyPhen2 score >0.05) (Adzhubei *et al.*, 2010), Mutation Assessor (damaging if Mutation assessor score >1.938) (Reva *et al.*, 2011), Mutation Taster (damaging if Mutation taster classifies as disease causing) (Schwarz *et al.*, 2010), PMut (damaging if Pmut score >0.5) (Ferrer-Costa *et al.*, 2005), Condel (damaging if Condel classifies as deleterious) (González-Pérez and López-Bigas, 2011). Conservation scores were also taken into account, namely: PhyloP (conserved if score >0.95 in a range between 0 and 1) (Pollard *et al.*, 2010), GERP (constrained if score ranges from 0 to 6.18) (Davydov *et al.*, 2010). Finally, we compared the variants identified in our patients to variants in curated catalogs of pathogenic variants: Human Gene Mutation Database (HGMD) (Stenson *et al.*, 2003) and ClinVar (Landrum *et al.*, 2014). A workflow for this analysis is represented in figure S2.3.

Prioritization of candidate genes

Gene prioritization took into consideration (I) the biological function according to information available in the literature (PubMed search) ('Home - PubMed - NCBI', n.d.) and in the Online Mendelian Inheritance in Man (OMIM) ('Home - OMIM - NCBI', n.d.), Gene Entrez ('Home - Gene - NCBI', n.d.), and GeneCards ('GeneCards - Human Genes | Gene Database | Gene Search', n.d.) databases; (II) the genetic/protein interactions (Warde-Farley, Sylva L. Donaldson, *et al.*, 2010) (preference was given to genes that were on the same pathway or had co-expression, co-

localization, physical or genetic interactions with *MECP2*, *CDKL5*, *FOXG1*); (III) brain expression ('Microarray Data :: Allen Brain Atlas: Human Brain', n.d.; Michael J. Hawrylycz *et al.*, 2012) (preference was given to genes expressed in the brain); and (IV) KO mice phenotype (Judith A. Blake *et al.*, 2014) (preference was given to genes for which KO mice had a neurological phenotype). A workflow for this analysis is represented in figure S2.4.

Sanger validation of selected candidate genes

Validation of the selected candidate variants was performed using standard Sanger sequencing. Primers were designed to surround the candidate variant using Primer3Plus software ('Primer3Plus', n.d.) and PCR reactions were performed using MasterMix ™ PCR reaction mix (Finnzymmes, ©Thermo Fisher Scientific In) for a final volume of 20 µl, at the following amplifications conditions: denaturation for 5 minutes at 95°C, followed by 30 amplification cycles (95ªC for 1 minute, amplicom specific annealing temperature for 30 seconds, 72°C for 1 minute) and final extension for 10 minutes at 72°C. The products were sequenced in ABI 3730 XL equipment (Applied Biosystems, Life technologies. Carlsbad, CA). Of 56 candidate variants, 44 variants were confirmed by Sanger to be present (see supplementary table S20). Primer sequences are available in supplementary table S2.2.

Network analysis

We performed gene network analysis to: I) verify if our list of candidate genes interacted amongst themselves and with the known RTT genes (*MECP2, CDKL5, FOXG1*), II) study the topology of these interactions, III) predict additional genes that may be involved in RTT if they are shown to interact with a large number of genes in the query set, IV) identify common biological themes by exploring functional enrichment analysis of Gene Ontology (GO) terms.

Network analysis was performed with GeneMANIA (version 3.1.2.7, http://www.genemania.org/).(Mostafavi *et al.*, 2008; Warde-Farley, Sylva L. Donaldson, *et al.*, 2010) Given a set of input genes, GeneMANIA finds related genes using a very large set of functional association data, including protein interactions (two gene products are linked if they were found to interact in a protein-protein interaction study), genetic interactions (two genes are functionally associated if the effects of perturbing one gene perturb a second gene), pathway (two gene products are linked if they participate in the same reaction within a pathway), co-expression (two genes are linked if their expression levels are similar across conditions in a gene expression study), co-localization (two genes are linked if they are both expressed in the same tissue or if their gene products are both identified in the same cellular location), shared protein domain (two

gene products are linked if they have the same protein domain) and predicted functional relationship (a major source of predicted data being known functional relationships from another organism via orthology). GeneMANIA also allows for functional enrichment analysis. Enrichment is calculated using as the foreground set all the genes in either the query list or the related genes discovered by GeneMANIA that have at least one GO annotation and as the background set all genes with GO annotations and at least one interaction in our GeneMANIA's database. For our analysis, we used GeneMANIA's default datasets for *Homo sapiens* (last dataset update June 1st 2014) and equal weighting by network. The genes used as input were the already known RTT genes (*MECP, CDKL5, FOXG1*) as well as the genes selected as likely causing RTT-like phenotype in our cohort. In the results generated by GeneMANIA we allowed for up to 20 related genes and at most 10 related attributes to be displayed.

Supplementary figures

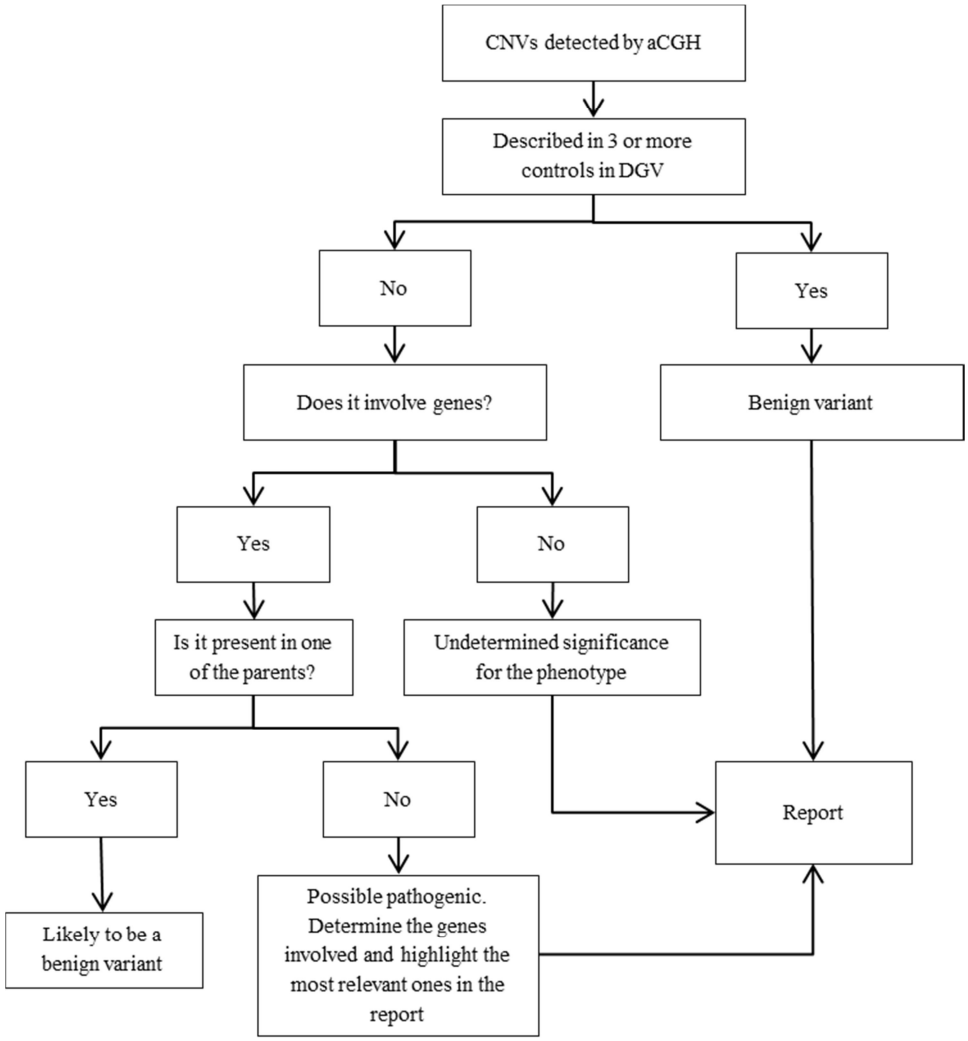


Figure S2.1 – Schematic representation of the workflow followed for the interpretation of the CNVs found in each patient. (DGV: Database of Genomic Variants http://dgv.tcag.ca/dgv/app/home)

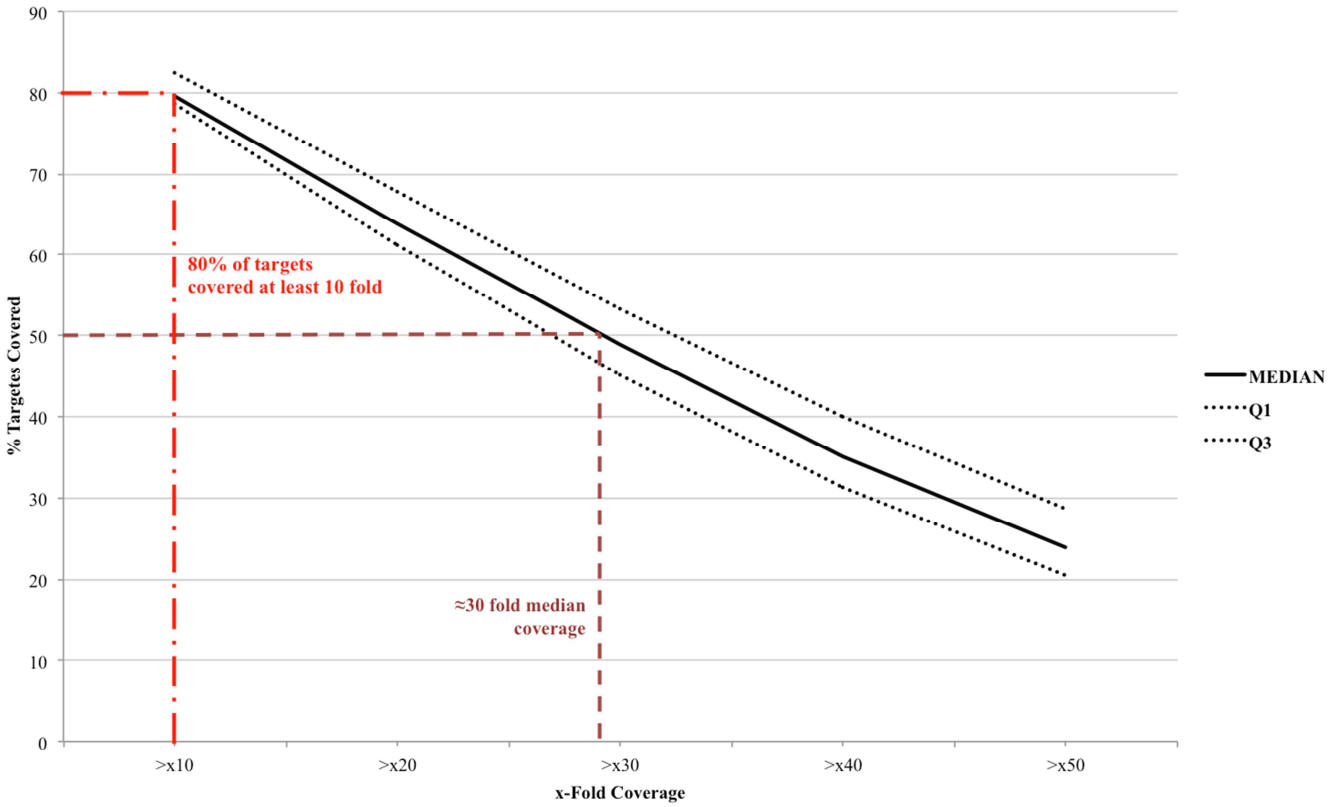
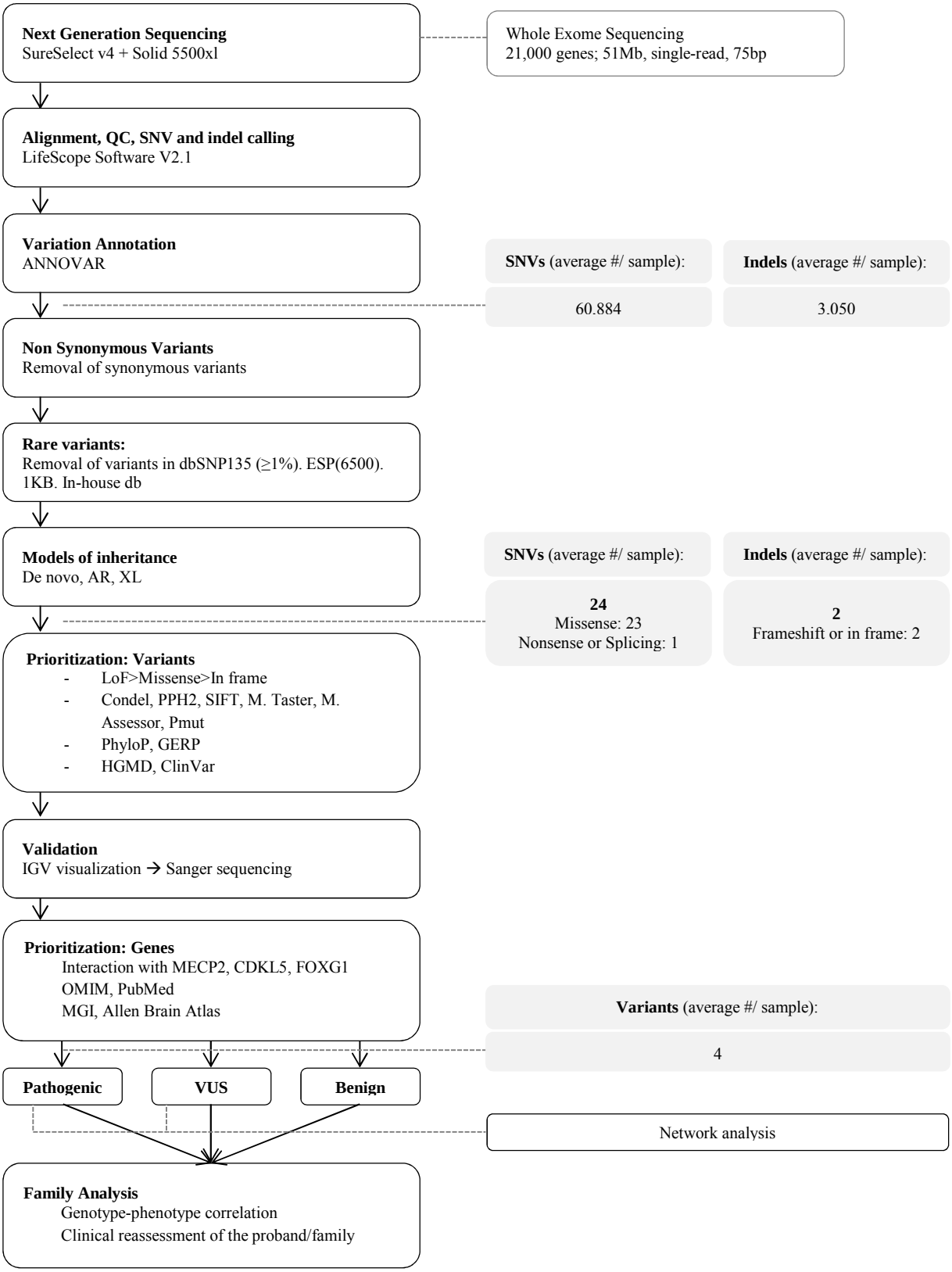


Figure S2.2 – WES coverage distribution. Coverage for all exons targeted by enrichment was evaluated. The median coverage across individuals was approximately 30 fold, with 80% of the targets covered at least 10 fold.


Next Generation Sequencing
SureSelect v4 + Solid 5500xl
Whole Exome Sequencing
21,000 genes; 51Mb, single-read, 75bp
Alignment, QC, SNV and indel calling
LifeScope Software V2.1
Variation Annotation
ANNOVAR
SNVs (average #/ sample):
Indels (average #/ sample):
60.884
3.050
Non Synonymous Variants
Removal of synonymous variants
Rare variants:
Removal of variants in dbSNP135 (≥1%). ESP(6500). 1KB. In-house db
Models of inheritance
De novo, AR, XL
SNVs (average #/ sample):
Indels (average #/ sample):
24
Missense: 23
Nonsense or Splicing: 1
2
Frameshift or in frame: 2
Prioritization: Variants
- LoF>Missense>In frame
- Condel, PPH2, SIFT, M. Taster, M. Assessor, Pmut
- PhyloP, GERP
- HGMD, ClinVar
Validation
IGV visualization → Sanger sequencing
Prioritization: Genes
Interaction with MECP2, CDKL5, FOXG1
OMIM, PubMed
MGI, Allen Brain Atlas
Variants (average #/ sample):
4
Pathogenic
VUS
Benign
Network analysis
Family Analysis
Genotype-phenotype correlation
Clinical reassessment of the proband/family

Exome sequencing and data analysis

Figure S2.3 - Overview of WES bioinformatic analysis pipeline. Abbreviatons: 1KG: 1000 genomes project; AR: autosomal recessive; ESP: Exome Sequencing Project; GERP: Genomic Evolutionary Rate Profiling; HGMD: Human Gene Mutation Database; IGV: Integrative Genomics Viewer; LoF: loss of function; M.ass: Mutation Assessor; M Taster: Mutation Taster; MGI: Mouse Genome Informatics; OMIM: Online Mendelian Inheritance in Man; PPH2: PolyPhen2; PhyloP: phylogenetic p-values; SIFT: Sorting Intolerant From Tolerant; VUS: Variants of Unkown Significance; XL: X-linked

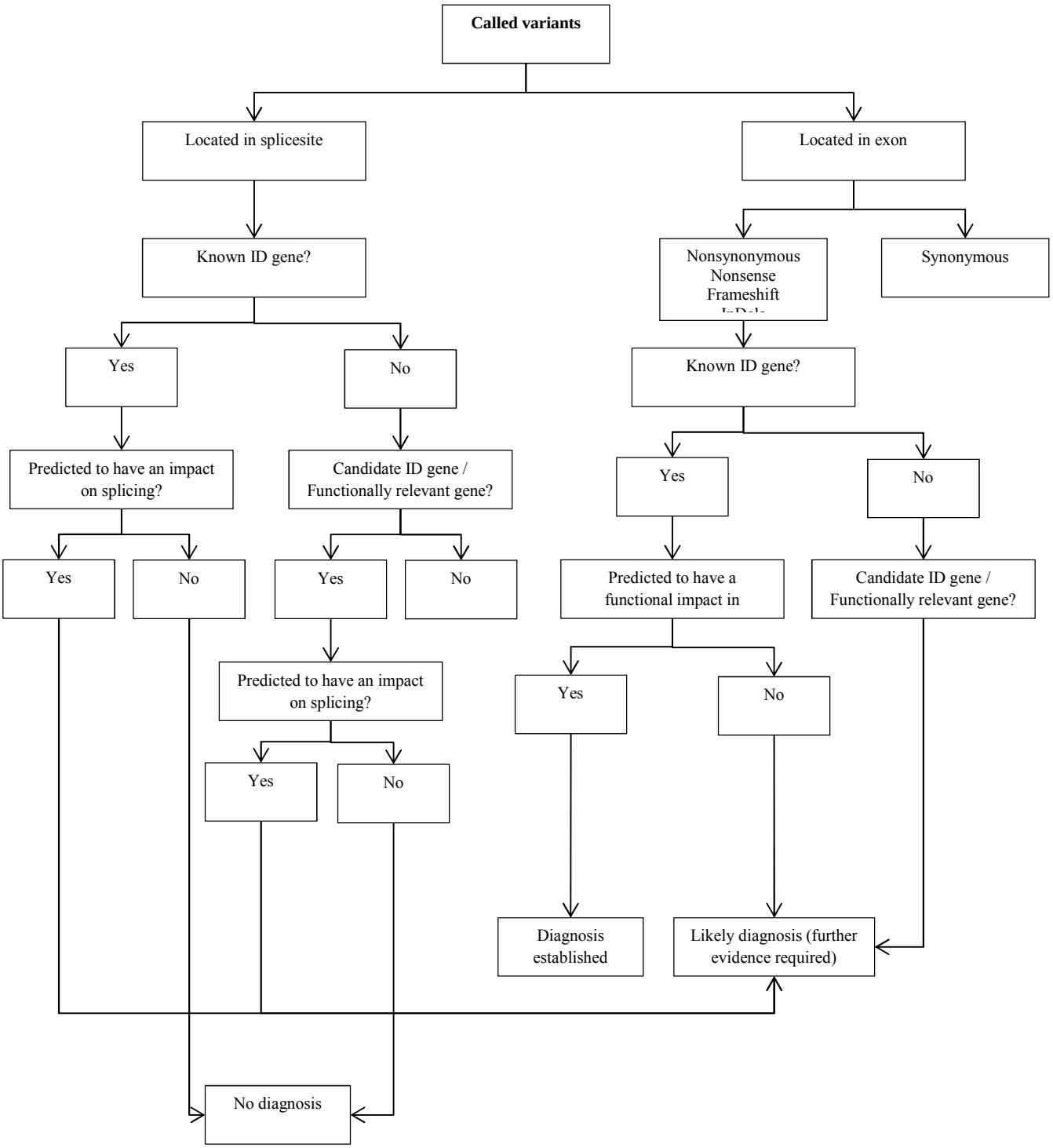


Figure S2.4 – Classification of the variants detected by WES. The workflow used to classify the variants called by WES is represented.

Supplementary tables

Supplementary table S2.1 – Primers used for quantitative PCR confirmation.

Gene	Primer Forward 5'→3'	Size (bp)	%GC	Tm (ºC)	Primer Reverse 5'→3'	Size (bp)	%GC	Tm (ºC)	Amplicon size (bp)
DRD3	CCATCTCACCATGCCTATCC	20	55	60	AAACAGAAGAGGGCAGGACA	20	50	60	168
ZNF80	GCTACCGCCAGATTCACACT	20	55	60	AATCTTCATGTGCCGGGTTA	20	45	60	182
SDC4	ACCGAACCCAAGAAACTAGA	20	45	56	GTGCTGGACATTGACACCT	19	53	57	101

Supplementary table S2.2 – Primers used for Sanger confirmation.

Gene	Primer Forward 5'→3'	Size (bp)	%GC	Tm (ºC)	Primer Reverse 5'→3'	Size (bp)	%GC	Tm (ºC)	Amplicon size (bp)
SLC35A2	AGATCCTCAAAGGCAGCTCA	20	50	60	GACAACCACAGCCACCAGTA	20	55	60	203
KCNT2	TGAACTTCTAGGATCATTTCAAGC	24	37	59	AATGCTGAATTCTCTTCTTTGG	22	36	58	250
CHPF	GACACTGGCCGCTTTGAT	18	55	60	TCCTGTTCAAAGAGTAGCATGG	22	45	59	317
RHOBTB2	CTGTACACGGGGGAGCTAGA	20	60	60	GCTGAGACAGACAGGCAATG	20	55	59	274
LAMB2	ATGCACCCTGTGCTACAAGC	20	55	61	TGGAAGCATTAGCCTTGTCC	20	50	60	261
JMJD1C	GAAATGGTGAGTTTCAGAAGGTT	23	39	59	CGTCGTGATGTAATGCCAAT	20	45	59	300
EIF4G1	TCCTGCTTCCCACTCATCTT	20	50	59	ATCCAGGAAAGCAGGGAAAT	20	45	60	245
PLEKHG6	CCAACCTGGTCTGAGGAAGA	20	55	60	CTCCCGGATCTCTCTGAGC	19	63	60	240
LARP4	TTCCTCTGTTCAGAAGGATGG	21	48	59	CCAACTTGGGCTACTTCCAA	20	50	60	248
DDX23 (exon10)	GCCCAGAGGCTCTACAGAAA	20	55	60	TGTAGCCACACTTATCAATGACC	23	44	59	291
EEF1A2 (exon3)	TTCAAGTATGCCTGGGTGCT	20	50	61	GAGCCAGACTGGGTGAGG	18	67	59	320
ASPM	CCTTTTGGCTTTTTCACGAG	20	45	60	CATTGATGTACCACTTCCCTGA	22	45	60	329
CHD8	AAGCCAATGATGCTGAGAAGA	21	43	60	AAGCATTCCCTCACCATTGA	20	45	60	246
CADPS2	TTTAATTGGGAAAAGAGTATTTTGA	25	24	58	CATCGAGAGGGGTTTAAATTG	21	43	59	300
DDX23 (exon15)	AGTGTTCATAGATGAGGGTTGG	22	45	58	GAAAGGGTGAGACTTGAAAGGA	22	45	60	286

EEF1A2 (exon7)	CAGGTCATCATCCTGAACCA	20	50	59	CGGGTACTGGGAGAAGCTC	19	63	60	231
LEPROTL1	CGTTTTAGGGATTTTTGCTTC	21	38	58	GGTAAGTGAACATTTCTCTGCATT	24	38	59	300
DNAH1	GTCATAGCGCTGGCAGGA	18	61	61	GCAGCACGTACCTTCTCCA	19	58	60	251
TCOF1	AGCTGCTGGAACAGGAAAGA	20	50	60	AGGAATGAGACCAGGTGCTG	20	55	60	473
SLC26A4	AGAATTGATTGTGTGTGTGTGC	22	41	59	TGTTTGTCAACCAAATAATGCTG	23	35	60	287
EIF2B2	TCCTCCCCACCTCTCTCTTT	20	55	60	CCTGTGACCAGTCCCATTTT	20	50	60	230
CCDC73	TTGATACAATAAAGGGAACCCAGT	24	38	60	TCCGTAACATTCGAGGTTTG	20	45	59	325
PPP1R32	CTCCCTTCTGGTGGTTCTGT	20	55	59	CTCCTGCACAAGAGGGGTAG	20	60	60	230
EXOC3L4	AACAAGCAGCAGGAAAGAGC	20	50	60	CTCGGGTTTCAGTTCCTCAG	20	55	60	298
CACNA1H	CATGAACTACCCCACGATCC	20	55	60	GAATGACACACCGGAGACC	19	58	59	249
KATNB1	AGGTGCAGTGCTGAATACCC	20	55	60	CCCCTCTGCAGCCACTTAC	19	63	61	223
SYNE1	AGAGGTAGAAAGCAGTTCTCATAATC	26	38	58	GAATGATTCACAAGAATAACACAAA	25	28	57	239
GAL	CCTGTAGCATGTGTCGTGGT	20	55	60	TGCATAAATTGGCCGAAGAT	20	40	60	235
CELSR3	TCAGGGAGCCTATCTTCGTG	20	55	60	TGCCGAATCAAAAAGTCTGA	20	40	59	298
PSIP1	TTGGTTGGTTTCAGTCTTTTCA	22	36	60	CCCTCAAAACAAGTTTTCAACA	22	36	59	299
TRAPPC6A	GAGGGTTGCATTGCTCCAT	19	53	61	TGGCTTCCCTTAGAGGATCA	20	50	60	373
WASF2	AGTGTGGTCAGCCCAAGC	18	61	60	GGTAGTCAGCTGCTGGTGGT	20	60	60	250
PICALM	TGCACAGATCCTTTCTCTGCT	21	48	60	TGCAGTGCTTTTAAAATAATTAATGG	26	27	60	247
NXF1	GAATGAACTGAAGCCAGAACAA	22	41	59	CTGAGAACTACTGCTGTCAACCA	23	48	60	240
STXBP1	TTTGTGGAAGAAACGGGAAC	20	45	60	AACCTTAAGGCCGGAGGAG	19	58	60	338
SHROOM4	TCTCTTTTGTGGCATGGGTA	20	45	59	TGGTTAGGGTACATGTTCTGGTC	23	48	60	250
CACNA1G	TGGAGCTCTTTGGAGACCTG	20	55	60	CTGTGGAGACTCGGAAGAGG	20	60	60	212
ZFX	GTGCCCAGATATCATGGAAGA	21	48	60	CGTCGGTAGTCAGAGGATCAG	21	57	60	293
TMEM18	TCCACATCAGCTCTCTCCAA	20	50	60	CACATGAAGAAAGACAACTGAAAG	24	38	58	232
DROSHA	AAATCACATTGAGATGAGATAATTTTT	27	22	58	CCACTGAATTGTATGCTTTAAATGAG	26	35	60	248
MAGEL2	CCCTAGCACCTCCAGGAT	18	61	58	CCTCTCATCCAAGGGAGACA	20	55	60	226
DKK3	GACCTCGGCTCCAGTCAA	18	61	60	CGCTCTTCCATGCCTTCC	18	61	62	212
RPGRIP1L	TTTCCTGCTCTTTTCTCCCTA	21	43	58	CCAGCTTCCCATGAATTATAGA	22	41	58	250
MAST4	CACACTGACAGGGCTCCTCTA	21	57	61	CGGGCTTTTGTCTGTCTGTC	20	55	61	249

SMARCA1	TGCACATTCAATTAAAAGCACT	22	32	58	GCAGTCCTAGACTTGATAAACCA	23	43	58	250
FXR2	GCAGCGACAAGGCTGGATA	19	58	62	CCCAGCCAATCAATCACTTT	20	45	60	245
HEXDC	AAAAACGGACCCTGTTAGGC	20	50	60	AGCAGCGCATCCACAGAG	18	61	61	197
HTT	GGAAATGATGGGAGCAGGTA	20	50	60	CCGTGGTGTCAAGAGGAACT	20	55	60	228
RIN2	GGGCTGCCTTCTTCTTCTTC	20	55	60	CTTCTCTTCACCTGCGCTTC	20	55	60	300
BNIP2	TTTTCTTTGTGTGTGTGTTTTT	24	29	59	TCAAAATATAGCTCTGTATCCCATAA	26	30	58	231
PSMD1	TGCTACTTTCAGTTGGATGTTTT	23	35	58	GAGCATCAGCCTCTTCCATC	20	55	60	246
TCTN2 (exon 6)	CAGCTCCTGCCTTTATGTTTG	21	48	60	GCGGCTGCTCAGACACTTAC	20	60	61	249
TCTN2 (exon 14)	GTGAAACCCGGACATTCACT	20	50	60	TCACAGCAACCAAGTTACAGG	21	48	59	236
RRAS2	GCCTCCCAGAGCATAGGATT	20	55	61	CGGGCTGCTCTGTCATCTAT	20	55	60	162
RGS4	GCAGAGCGGTCGTCTGAT	18	61	60	GCAATCTTACCTCCTCAAGCA	21	48	60	249
LIG4	CTGCACCTTGCGTTTTCC	18	57	60	TGGCTATCTGTTCCACTCATAA	22	41	57	300
GABBR2	GGGGAGTATTTGTCCCCATC	20	55	60	CCCTGAAACAGAAGGAGAGTG	21	52	59	244
RCC2	AGGAAATGTTGCCGTGATTC	20	45	60	CCCAAACAGACTGCGATGAT	20	50	61	249
ITGA7	GGGCATTGACATTTCCAAAC	20	45	60	CACAGCTCAGCCTCTCCTCT	20	60	60	266
ZNF238	CACATAGCAGGCGATTTGC	19	53	60	CTCACGGAGGTGACAGACCT	20	60	60	245

References

1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012; 491: 56–65.

Aasland R, Stewart AF, Gibson T. The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB. Trends Biochem. Sci. 1996; 21: 87–88.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat. Methods 2010; 7: 248–249.

Ameur A, Bunikis I, Enroth S, Gyllensten U. CanvasDB: a local database infrastructure for analysis of targeted- and whole genome re-sequencing projects. Database J. Biol. Databases Curation 2014; 2014

Bardwell VJ, Treisman R. The POZ domain: a conserved protein-protein interaction motif. Genes Dev. 1994; 8: 1664–1677.

Blake JA, Bult CJ, Eppig JT, Kadin JA, Richardson JE, Mouse Genome Database Group. The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. Nucleic Acids Res. 2014; 42: D810-817.

Blake JA, Bult CJ, Eppig JT, Kadin JA, Richardson JE, Mouse Genome Database Group. The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. Nucleic Acids Res. 2014; 42: D810-817.

Blanckenberg J, Ntsapi C, Carr JA, Bardien S. EIF4G1 R1205H and VPS35 D620N mutations are rare in Parkinson's disease from South Africa. Neurobiol. Aging 2014; 35: 445.e1-3.

Blein S, Hawrot E, Barlow P. The metabotropic GABA receptor: molecular insights and their functional consequences. Cell. Mol. Life Sci. CMLS 2000; 57: 635–650.

Buffart TE, Israeli D, Tijssen M, Vosse SJ, Mrsić A, Meijer GA, et al. Across array comparative genomic hybridization: a strategy to reduce reference channel hybridizations. Genes. Chromosomes Cancer 2008; 47: 994–1004.

Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, et al. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res. 2007; 35: 2013–2025.

Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput. Biol. 2010; 6: e1001025.

Duyao MP, Auerbach AB, Ryan A, Persichetti F, Barnes GT, McNeil SM, et al. Inactivation of the mouse Huntington's disease gene homolog Hdh. Science 1995; 269: 407–410.

Eckey M, Kuphal S, Straub T, Rummele P, Kremmer E, Bosserhoff AK, et al. Nucleosome Remodeler SNF2L Suppresses Cell Proliferation and Migration and Attenuates Wnt Signaling. Mol. Cell. Biol. 2012; 32: 2359–2371.

Edelmann L, Hirschhorn K. Clinical utility of array CGH for the detection of chromosomal imbalances associated with mental retardation and multiple congenital anomalies. Ann. N. Y. Acad. Sci. 2009; 1151: 157–166.

EuroEPINOMICS-RES Consortium, Epilepsy Phenome/Genome Project, Epi4K Consortium. De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. Am. J. Hum. Genet. 2014; 95: 360–370.

Fatemi SH, Folsom TD, Reutiman TJ, Thuras PD. Expression of GABA(B) receptors is altered in brains of subjects with autism. Cerebellum Lond. Engl. 2009; 8: 64–69.

Fatemi SH, Folsom TD, Thuras PD. Deficits in GABA(B) receptor system in schizophrenia and mood disorders: a postmortem study. Schizophr. Res. 2011; 128: 37–43.

Ferrer-Costa C, Gelpí JL, Zamakola L, Parraga I, de la Cruz X, Orozco M. PMUT: a web-based tool for the annotation of pathological mutations on proteins. Bioinforma. Oxf. Engl. 2005; 21: 3176–3178.

Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am. J. Hum. Genet. 2009; 84: 524–533.

Fujioka S, Sundal C, Strongosky AJ, Castanedes MC, Rademakers R, Ross OA, et al. Sequence variants in eukaryotic translation initiation factor 4-gamma (eIF4G1) are associated with Lewy body dementia. Acta Neuropathol. (Berl.) 2013; 125: 425–438.

González-Pérez A, López-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am. J. Hum. Genet. 2011; 88: 440–449.

Gorbalenya AE, Koonin EV, Donchenko AP, Blinov VM. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res. 1989; 17: 4713–4730.

Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature 2012; 489: 391–399.
Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature 2012; 489: 391–399.
Hedges DJ, Hamilton-Nelson KL, Sacharow SJ, Nations L, Beecham GW, Kozhekbaeva ZM, et al. Evidence of novel fine-scale structural variation at autism spectrum disorder candidate loci. Mol. Autism 2012; 3: 2.

Hoebeeck J, van der Luijt R, Poppe B, De Smet E, Yigit N, Claes K, et al. Rapid detection of VHL exon deletions using real-time quantitative PCR. Lab. Investig. J. Tech. Methods Pathol. 2005; 85: 24–33.

Hoischen A, van Bon BWM, Gilissen C, Arts P, van Lier B, Steehouwer M, et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nat. Genet. 2010; 42: 483–485.

Hussain RH, Zawawi M, Bayfield MA. Conservation of RNA chaperone activity of the human La-related proteins 4, 6 and 7. Nucleic Acids Res. 2013; 41: 8715–8725.

Illumina. DNA Copy Number and Loss of Heterozygosity Analysis Algorithms [Internet]. 2013[cited 2014 Apr 22] Available from: http://www.google.pt/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CDEQFjAA&url=http%3A%2F%2Fres.illumina.com%2Fdocuments%2Fproducts%2Ftechnotes%2Ftechnote_cnv_algorithms.pdf&ei=XYFWU62wIuKk0QX-yIDgAw&usg=AFQjCNGPmHgIQLQ0YFdD7hya3mv_5Ip8TQ&bvm=bv.65177938,d.bGQ&cad=rja

Kang H, Chen I-M, Wilson CS, Bedrick EJ, Harvey RC, Atlas SR, et al. Gene expression classifiers for relapse-free survival and minimal residual disease improve risk classification and outcome prediction in pediatric B-precursor acute lymphoblastic leukemia. Blood 2010; 115: 1394–1405.

Kobayashi M, Kishida S, Fukui A, Michiue T, Miyamoto Y, Okamoto T, et al. Nuclear localization of Duplin, a beta-catenin-binding protein, is essential for its inhibitory activity on the Wnt signaling pathway. J. Biol. Chem. 2002; 277: 5816–5822.

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, et al. Rate of de novo mutations and the importance of father's age to disease risk. Nature 2012; 488: 471–475.

Krumm N, O'Roak BJ, Shendure J, Eichler EE. A de novo convergence of autism genetics and molecular neuroscience. Trends Neurosci. 2014; 37: 95–105.

Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014; 42: D980-985.

Lee S, Kozlov S, Hernandez L, Chamberlain SJ, Brannan CI, Stewart CL, et al. Expression and imprinting of MAGEL2 suggest a role in Prader-willi syndrome and the homologous murine imprinting phenotype. Hum. Mol. Genet. 2000; 9: 1813–1819.

Li K, Tang B, Guo J, Lou M, Lv Z, Liu Z, et al. Analysis of EIF4G1 in ethnic Chinese. BMC Neurol. 2013; 13: 38.

Lucotte G, Lagarde JP, Funalot B, Sokoloff P. Linkage with the Ser9Gly DRD3 polymorphism in essential tremor families. Clin. Genet. 2006; 69: 437–440.

MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. A Systematic Survey of Loss-of-Function Variants in Human Protein-Coding Genes. Science 2012; 335: 823–828.

McFarland KN, Huizenga MN, Darnell SB, Sangrey GR, Berezovska O, Cha J-HJ, et al. MeCP2: a novel Huntingtin interactor. Hum. Mol. Genet. 2014; 23: 1036–1044.

Merret R, Martino L, Bousquet-Antonelli C, Fneich S, Descombin J, Billey E, et al. The association of a La module with the PABP-interacting motif PAM2 is a recurrent evolutionary process that led to the neofunctionalization of La-related proteins. RNA N. Y. N 2013; 19: 36–50.

Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. Genome Biol. 2008; 9 Suppl 1: S4.

Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 2012; 485: 242–245.

Neul JL, Kaufmann WE, Glaze DG, Christodoulou J, Clarke AJ, Bahi-Buisson N, et al. Rett syndrome: revised diagnostic criteria and nomenclature. Ann. Neurol. 2010; 68: 944–950.

Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res. 2001; 11: 863–874.

Nguyen GD, Molero AE, Gokhan S, Mehler MF. Functions of Huntingtin in Germ Layer Specification and Organogenesis. PLoS ONE 2013; 8: e72698.

Nishiyama M, Skoultchi AI, Nakayama KI. Histone H1 recruitment by CHD8 is essential for suppression of the Wnt-β-catenin signaling pathway. Mol. Cell. Biol. 2012; 32: 501–512.

O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science 2012; 338: 1619–1622.

O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 2012; 485: 246–250.

Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 2010; 20: 110–121.

Puschmann A. Monogenic Parkinson's disease and parkinsonism: clinical phenotypes and frequencies of known mutations. Parkinsonism Relat. Disord. 2013; 19: 407–415.

Reiss AL, Faruque F, Naidu S, Abrams M, Beaty T, Bryan RN, et al. Neuroanatomy of Rett syndrome: A volumetric imaging study. Ann. Neurol. 1993; 34: 227–234.

Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 2011; 39: e118.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat. Biotechnol. 2011; 29: 24–26.

Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 2011; 70: 863–885.

Sang L, Miller JJ, Corbit KC, Giles RH, Brauer MJ, Otto EA, et al. Mapping the NPHP-JBTS-MKS protein network reveals ciliopathy disease genes and pathways. Cell 2011; 145: 513–528.

Schaaf CP, Gonzalez-Garay ML, Xia F, Potocki L, Gripp KW, Zhang B, et al. Truncating mutations of MAGEL2 cause Prader-Willi phenotypes and autism. Nat. Genet. 2013; 45: 1405–1408.

Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat. Methods 2010; 7: 575–576.

Shaheen R, Faqeih E, Seidahmed MZ, Sunker A, Alali FE, AlQahtani K, et al. A TCTN2 mutation defines a novel Meckel Gruber syndrome locus. Hum. Mutat. 2011; 32: 573–578.

Siripurapu V, Meth J, Kobayashi N, Hamaguchi M. DBC2 significantly influences cell-cycle, apoptosis, cytoskeleton and membrane-trafficking pathways. J. Mol. Biol. 2005; 346: 83–89.

Smirnov DA, Morley M, Shin E, Spielman RS, Cheung VG. Genetic analysis of radiation-induced changes in human gene expression. Nature 2009; 459: 587–591.

Spurlock G, Williams J, McGuffin P, Aschauer HN, Lenzinger E, Fuchs K, et al. European Multicentre Association Study of Schizophrenia: a study of the DRD2 Ser311Cys and DRD3 Ser9Gly polymorphisms. Am. J. Med. Genet. 1998; 81: 24–28.

Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NST, et al. Human Gene Mutation Database (HGMD): 2003 update. Hum. Mutat. 2003; 21: 577–581.

Sudhaman S, Behari M, Govindappa ST, Muthane UB, Juyal RC, Thelma BK. VPS35 and EIF4G1 mutations are rare in Parkinson's disease among Indians. Neurobiol. Aging 2013; 34: 2442.e1-3.

Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A, et al. Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. Cell 2012; 149: 525–537.

Villa N, Do A, Hershey JWB, Fraser CS. Human eukaryotic initiation factor 4G (eIF4G) protein binds to eIF3c, -d, and -e to promote mRNA recruitment to the ribosome. J. Biol. Chem. 2013; 288: 32932–32940.

Vuillaume M-L, Delrue M-A, Naudion S, Toutain J, Fergelot P, Arveiler B, et al. Expanding the clinical phenotype at the 3q13.31 locus with a new case of microdeletion and first characterization of the reciprocal duplication. Mol. Genet. Metab. 2013; 110: 90–97.

Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007; 17: 1665–1674.

Wang Q. Integrative Genomics Identifies Distinct Molecular Classes of Neuroblastoma and Shows That Multiple Genes Are Targeted by Regional Alterations in DNA Copy Number. Cancer Res. 2006; 66: 6050–6062.

Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010; 38: W214-220.

Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010; 38: W214-220.

Wilkins A, Ping Q, Carpenter CL. RhoBTB2 is a substrate of the mammalian Cul3 ubiquitin ligase complex. Genes Dev. 2004; 18: 856–861.

Wiśniowiecka-Kowalnik B, Kastory-Bronowska M, Bartnik M, Derwińska K, Dymczak-Domini W, Szumbarska D, et al. Application of custom-designed oligonucleotide array CGH in 145 patients with autistic spectrum disorders. Eur. J. Hum. Genet. EJHG 2013; 21: 620–625.

Yang R, Gaidamakov SA, Xie J, Lee J, Martino L, Kozlov G, et al. La-related protein 4 binds poly(A), interacts with the poly(A)-binding protein MLLE domain via a variant PAM2w motif, and can promote mRNA stability. Mol. Cell. Biol. 2011; 31: 542–556.

Yip DJ, Corcoran CP, Alvarez-Saavedra M, DeMaria A, Rennick S, Mears AJ, et al. Snf2l Regulates Foxg1-Dependent Progenitor Cell Expansion in the Developing Brain. Dev. Cell 2012; 22: 871–878.

Welcome to the ISCA Consortium web site [Internet]. [cited 2014 Jul 24] Available from: https://www.iscaconsortium.org/

Primer3Plus [Internet]. [cited 2014 Aug 5] Available from: http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi

Exome Variant Server [Internet]. [cited 2014 Jul 24] Available from: http://evs.gs.washington.edu/EVS/

Microarray Data:: Allen Brain Atlas: Human Brain [Internet]. [cited 2014a Jun 16] Available from: http://human.brain-map.org/

Home - PubMed - NCBI [Internet]. [cited 2014 Aug 5] Available from: http://www.ncbi.nlm.nih.gov/pubmed

Home - OMIM - NCBI [Internet]. [cited 2014 Aug 5] Available from: http://www.ncbi.nlm.nih.gov/omim

Home - Gene - NCBI [Internet]. [cited 2014 Aug 5] Available from: http://www.ncbi.nlm.nih.gov/gene

GeneCards - Human Genes | Gene Database | Gene Search [Internet]. [cited 2014 Aug 5] Available from: http://www.genecards.org/

Microarray Data:: Allen Brain Atlas: Human Brain [Internet]. [cited 2014b Aug 5] Available from: http://human.brain-map.org/

Clinical exome sequencing: the importance of re-analyzing unsolved cases

Clinical exome sequencing: the importance of re-analyzing unsolved cases

Fatima Lopes, Julie Gauthier, Marguerite Miguet, Guylaine D'Amours, Philippe M. Campeau, Fadi H. Hamdan, Mark E. Samuels, Virginie Saillour, Emmanuelle Lemyre, Grant A. Mitchell, Jean-François Soucy, Jacques L. Michaud.

Abstract

Whole exome sequencing is currently a widely used technique for diagnosis of disease-causing variants in the diagnostic context. Although whole exome sequencing allows the genetic explanation of many cases in the clinical context, the number of unsolved cases is still elevated. This can be due to the almost daily discovery of new disease associated genes with the broader implementation clinical exome sequencing. These advents bring forward the need for re-analyzing exome results regularly, in order to update the reports in the light of more recent data.

The aims of this work were to identify the genetic cause of disease in a clinically heterogeneous group of patients and to evaluate the yield obtained in the re-analysis of older cases (studied up to 3 years before). Additionally, the efficacy of a software for variant prioritization (Exomiser) was also tested.

In this work we studied a group of 64 patients using two different approaches: (I) the application of a virtual gene panel selected based on the patient's phenotype and, if negative, (II) the entire exome analysis. A total of sixty-four patients previously studied between the year of 2013 and 2014 were again re-analysed in 2016.

Based on their specific clinical features, we applied seven different virtual gene panels to the 64 patients. Together, pathogenic, likely pathogenic and uncertain significance variants were found in 34% of the cases. With the use of virtual gene panels we were able to find a good candidate to explain the disease in 17% of the cases, while the other half (17%) of patients a disease causing variant was find in a gene outside of the panel used. We also found that the use of the Exomiser software for causative genes prioritization added value to the manual analysis. Such software could be useful for a faster and automated re-analysis of older unsolved cases since it is still able to detect human errors made along the analysis.

This work highlights the utility of using virtual panels for filtering variants in exome data, as well as the importance of re-analyzing results of exome analysis on a regular basis.

Introduction

Over the years the advances in next generation sequencing (NGS) techniques increased tremendously which consequently lead to the widespread use of those in research and clinical practice. Of these advents, sequencing all the protein coding regions of the genome (whole exome sequencing, WES) has been widely implemented in research and clinical context and proven to be a cost-effective method for disease-causing variants detection as well as new disease associated genes (Bao *et al.*, 2014). WES has proven to be a resourceful tool in the discovery of a genetic diagnosis reaching yield around 25% in patient collections of heterogeneous Mendelian disorders (Lee *et al.*, 2014; Yang *et al.*, 2014).

Although the added value of NGS is unquestionable, the strategy and pipeline in a diagnostic context can vary. The most common strategies currently used are the (I) disease related gene panel sequencing, (II) whole exome sequencing and (III) whole genome sequencing (the latter being one the less frequently used) (Saudi Mendeliome Group, 2015; Sun *et al.*, 2015; Meienberg *et al.*, 2016). The choice between one of these approaches is done at the individual labs level and most commonly based on factors such as the cost, diagnostic yield (and in this the false negative rate and the false positive rate should be taken into account) and overall cost-effectiveness (if a test retrieves such a high number of variants that it delays the analysis by many months with only a slight increase in yield it may not be a useful approach to adopt) (Sun *et al.*, 2015).

Another critical aspect in the clinical context is the interpretation and clarification of a variant's significance. A clear indication of this is the development of the ACMG recommended guidelines on the interpretation and reporting of sequence variations (Richards *et al.*, 2008). Over the last years these guidelines have dramatically broadened the interpretative categories, changing from six categories to a more sophisticated scoring system following a detailed algorithm, designed to combine multiple lines of evidence in order to assign a classification to a potentially causal variant (Richards *et al.*, 2015). More recently, these guidelines were put to the test among nine different laboratories, revealing that differences in variant classification can still be found (Amendola *et al.*, 2016). Currently, the Sequence Variant Inter-Laboratory Discrepancy Resolution task team from ACMG is working on an updated version that would tackle these gaps.

With this work we aim to (I) test the value of re-analysing unsolved cases as well as the yield obtained in a group of clinically heterogeneous patients and (II) test the value of the application of an automated software for variant prioritization.

Subjects and Methods

Patients

Patients were enrolled by medical geneticists at CHUSJ (Montreal, Canada), Centre Hospitalier Universitaire de l'Hôtel-Dieu (Quebec, Canada), Centre Universitaire de Santé McGill (Montreal, Canada) and Montreal Neurological Institute (Montreal, Canada). For the vast majority of the patients extensive clinical and molecular investigations were performed before WES enrollment. Only the cases with rare likely genetic conditions which remained unexplained despite extensive investigations (biochemical, genetics, etc.) and absence of explanatory copy number variants using a 135k-feature whole-genome microarray (SignatureChipOS®, Signature Genomic Laboratories, Spokane, WA, USA by Perkin Elmer, Waltham, MA USA; based on UCSC 2009 hg19 assembly) were considered for exome sequencing. This study was approved by the institutional ethics committee, and informed consent was obtained from the parents or legal guardians of all patients included in the study. Typically DNA was purified from peripheral blood samples, but other sources were occasionally used, based on medically-required biopsies: amniocytes, fibroblasts, liver, muscle, kidney, or via cordocenthesis. From a total of 97 patients previously studied, 64 were selected and classified into one or two (maximum) clinical categories corresponding to virtual gene panels (table I): Intellectual disability, Mitochondrial disorders, Neuropathies, Skeletal dysplasias, Vision disorders, Multiple congenital anomalies and Skin disorders.

Table XVI – Gene panels used in the analysis.

Panel	Number of genes included (2015)	Origin
Intellectual disability	641	RadboudUMC
Mitochondrial disorders	231	RadboudUMC
Multiple congenital anomalies	2746	RadboudUMC
Neuropathies	763	RadboudUMC
Vision disorders	282	RadboudUMC
Skeletal dysplasias	364	(Bonafe *et al.*, 2015)
Skin disorders	563	RadboudUMC

Exome sequencing data generation

Between 2013 and 2015 the Centre Hospitalier Universitaire Sainte-Justine (CHUSJ) performed exome sequencing approach to index cases in patients with clinically heterogeneous Mendelian disorders in order to better determine the molecular cause of the disease in the patients reaching a diagnostic yield of 34%. In 2016 the remaining unsolved cases were re-analysed using an updated pipeline that combines WES with the virtual filtering of gene panels.

WES was performed at Génome Québec/CHU Sainte-Justine Integrated Center for Pediatric Clinical Genomics (Montreal, Canada) between 2013 and 2015. Exon-containing fragments were captured using the Agilent SureSelect Human All Exon Capture V4 Kit. Sequencing was done using paired-end 2x100 bp read chemistry, with 3 exomes/lane format using the Illumina HiSeq2000 system. Data processing was done again in 2016 as follow: alignment was done using a Burrows-Wheeler algorithm, BWA-mem v0.7.5a, variant calling using the Genome Analysis Toolkit (GATK) UnifiedGenotyper v2.6-4 tool (https://www.broadinstitute.org/gatk/guide/tooldocs/org_broadinstitute_gatk_tools_walkers_genotyper_UnifiedGenotyper.php) (DePristo *et al.*, 2011) and variant annotation was done using Annovar v2014-11-12 (http://www.openbioinformatics.org/annovar/) (Wang *et al.*, 2010). A 137-fold mean coverage for targeted exons was obtained and 95% of the target bases were covered at least 20-fold. Single Nucleotide Variants (SNVs) and indels were filtered as follows: at least 3 independent variant-containing reads constituting at least 20% of all reads covering the variant's position. Common variants with minor allele frequency greater than 0.5% in either the 1000 Genomes Project or NHLBI exome database (EVS, 6500 exomes) were excluded. Variants seen more than once in an in-house dataset of 100 similarly sequenced exomes from patients with rare-monogenic diseases unrelated to the clinical phenotypes of our patients were also excluded. Intergenic, intronic and synonymous variants other than those the consensus splice sites were removed from analysis and preference was given to (i) *de novo* variants, (ii) homozygous or compound heterozygous variants compatible with an autosomal recessive mode of transmission and (iii) X-linked variants (for male probands).

Exome sequencing data analysis

Because the study was carried in a clinical setting of a pediatric hospital, increased relevance was given to variants previously associated with a disease gene in either the Online Mendelian Inheritance in Man (OMIM), the Human Gene Mutation Database (HGMD) and in ClinVar.

Manual analysis

For each sample all the variants were first analysed using a virtual panel (retrieved from RadboudUMC, all except the Skeletal Dysplasias panel (table I) (Bonafe *et al.*, 2015) and selected based on the clinical presentation of the patient. If appropriate in one patient could be applied more than one panel. If this initial step didn't retrieve any good candidate or if the selected candidates were dismissed after validation, whole exome analysis was undertaken first focusing on the OMIM genes and only after moving out to the non-OMIM genes. All the genes carrying homozygous and potential compound heterozygous variants were analysed. For the genes with simple heterozygous variants all the STOP, frameshift, non-frameshift, canonical splice site variants were analyzed while the nonsynonymous variants (not located in canonical splice sites) were filter for an ExAC frequency < 0.00001 and for a CADD ≥ and a GERP++ ≥ 2. Variants classified as pathogenic and likely pathogenic (according to ACMG guidelines) (Richards *et al.*, 2015) in known disease genes were discussed at a multidisciplinary meeting (including medical geneticists, molecular genetics and bioinformaticians) for genotype-phenotype evaluation. Only the variants selected as likely to explain the individuals' phenotype were confirmed by Sanger in the patients and, whenever possible, in in parents and other relevant family members (i.e. affected siblings).

Exomiser analysis

Exomiser 7.2.1 (http://www.sanger.ac.uk/science/tools/exomiser) was run in all the samples according with the protocol described by (Smedley *et al.*, 2015). The prioritization model used for all the samples was the integrated phenotypic and interactome analysis (hiPHIVE); an autosomal dominant and recessive inheritance model was applied to all the cases, while the X-linked recessive model was also applied to the male probands (Robinson *et al.*, 2014; Haendel *et al.*, 2015). A minor allele frequency (MAF) of 0.5% was used and all the variant frequency default databases were used. All the default pathogenicity sources were kept in the analysis (Polyphen, Mutation Taster, Sift, CADD and REMM). Exomiser results were carefully evaluated in only 33 of the solved cases (for optimization purposes) and in all the unsolved cases.

For each patient a maximum of 10 core phenotypic features (defined using HPO terms) were used. Only the first five hits for each inheritance model were considered for the remaining analysis.

Sanger sequencing confirmations

Validation was performed using standard Sanger sequencing. Primers were designed to surround the candidate variant using Primer3Plus software and PCR reactions were performed using Taq DNA Polymerase (Invitrogen, ©Thermo Fisher Scientific In). The PCR products were sequenced in both directions in Applied Biosystems 3130 Genetic Analyze (Applied Biosystems, Life technologies. Carlsbad, CA). A total of 87 first nation controls (24 Ojibwas, 32 Crees and 31 Algonquin) were also sequenced for the variant found in *DNAJC21*).

Results

Cohort's clinical profile

From an initial pool of 97 patients (40 females, 56 males and 1 foetus with undetermined gender) we studied 64 patients (24 females, 39 males and 1 fetus) with ages comprised between 3 and 56 years (mean age 11,5 years). Of those, a total of 13 patients were from consanguineous families, 45 from non-consanguineous families and for 6 it was not possible to determine the consanguinity status. These 64 previously unsolved cases had been initially analysed in 2013-2014. Although clinically heterogeneous, the patients included in this analysis were enriched for intellectual disability (n=33, 52%) and mitochondrial disorders (n=14, 22%) spectrum, while the other groups had a lower number of patients (neuropathies n=9 (14%); skeletal dysplasias n=2 (3%); vision disorders n=2 (3%), multiple congenital anomalies n=2 (3%) and skin disorders n=2 (3%) (table III).

Table XVII – Clinical description of the 64 patients re-analyzed.

Patient ID	Sex	Patient phenotype (HPO terms used to run Exomiser)	Panel	Detected in the panel?
P1	M	Seizures, Neutropenia, Lactic acidosis, Intellectual disability (moderate), Hypopigmented skin patches, Hypertrophic cardiomyopathy, Hyperalaninemia, Attention deficit hyperactivity disorder, Ascending aortic dilation, Aggressive behavior	ID, Mito	Yes (mito)
P2	M	Speech apraxia, Severe expressive language delay, Reduced tendon reflexes, Motor delay, Microcephaly, Intellectual disability (borderline), Dystonia, Dysphasia, Developmental regression, Decreased body weight	ID	No
P3	F	Nystagmus, Myopathy, Microcephaly, Hydronephrosis, Hearing impairment, Growth hormone deficiency, Central adrenal insufficiency, Bronchiectasis, Brain atrophy, Arthrogryposis multiplex congenita	MCA	Yes
P4	F	Scoliosis, Pancytopenia, Osteopenia, Muscular hypotonia, Mild neurosensory hearing impairment, Hepatosplenomegaly, Global developmental delay, Encephalopathy, Delayed myelination, Cerebral cortical atrophy	Skin	No
P5	F	Triangular face, Respiratory insufficiency due to muscle weakness, Proptosis, Myopathic facies, Mandibular prognathia, Generalized muscle weakness, Fatigable weakness of swallowing muscles, Downslanted palpebral fissures, Congenital, generalized hypertrichosis	ID	No
P6	M	Skeletal muscle atrophy, Scoliosis, Hydronephrosis, Global developmental delay, Generalized hypotonia, Dysphagia, Delayed CNS myelination, Cryptorchidism, Congenital bilateral ptosis, Abnormal renal corticomedullary differentiation	Neuro	No
P7	M	Vomiting, Mitochondrial depletion, Hypoglycemia, Hepatic failure, Chronic hepatitis	Mito	No
P8	M	Tachypnea, Strabismus, Lactic acidosis, Infantile spasms, Hypoglycemia, Hypermetropia, Global developmental delay	ID	Yes
P9	M	Spastic tetraplegia, Sensorineural hearing impairment, Seizures, Optic atrophy, Muscular hypotonia of the trunk, Motor delay, Epileptic encephalopathy, Delayed myelination, Cerebral atrophy, Blindness	Mito	No
P10	F	Small for gestational age, Plagiocephaly, Muscular hypotonia of the trunk, Microcephaly, Limb hypertonia, Intention tremor, Global developmental delay, Congenital microcephaly	ID	Yes
P11	F	Ventricular septal defect, Mitochondrial myopathy, Hyporeflexia, Hepatomegaly, Dilated cardiomyopathy, Developmental regression, Decreased body weight, Coarctation of aorta, Cerebral atrophy	Mito	No
P12	M	Scoliosis, Muscular hypotonia, Macrocephaly, Intellectual disability, Hypertelorism, Global developmental delay, Encephalopathy, Dysphagia, Choreoathetosis, Birth length greater than 97th percentile	ID	Yes
P13	M	Severe global developmental delay, Myopathy, Intermittent lactic acidemia, Severe intellectual disability, Episodic metabolic acidosis, Corpus callosum atrophy, Cerebral white matter agenesis, Cerebral atrophy, Cerebellar vermis atrophy, Aplasia/Hypoplasia of the cerebellum	Mito	No
P14	M	Spastic tetraparesis, Seizures, Microcephaly, Global developmental delay, Dyspnea, Dysphagia, Decreased body weight, Abnormal cortical gyration, Abnormal CNS myelination	ID	Yes
P15	F	Short stature, Seizures, Obesity, Intellectual disability (moderate), Global developmental delay	ID	No
P16	F	Seizures, Obesity, Muscular hypotonia, Intellectual disability (severe), Increased body weight, Hypoplasia of the pons, Hypoplasia of the corpus callosum, Global developmental delay, Aplasia/Hypoplasia of the cerebellar vermis, Abnormality of the periventricular white matter	ID	Yes
P17	F	Severe conductive hearing impairment, Nonketotic hyperglycinemia, Neonatal hypotonia, Motor delay, Microcephaly, Hypertelorism, Episodic tachypnea, Encephalopathy, Delayed speech and language development, Anotia	ID	No
P18	M	Widened subarachnoid space, Ventriculomegaly, Spastic tetraplegia, Single umbilical artery, Short stature, Protruding ear, Microcephaly, Mandibular prognathia, Brachycephaly	ID	No
P19	F	Spastic tetraparesis, Short stature, Obesity, Hyperactive deep tendon reflexes, Global developmental delay, Eczema, Dysphagia, Abnormality of the globus pallidus	ID	Yes
P20	F	Seizures, Scoliosis, Optic atrophy, Myopia, Muscular hypotonia, Microtia, Intellectual disability (mild), Hypothyroidism, Global developmental delay, Congenital sensorineural hearing impairment	ID	Yes
P21	M	Unilateral cryptorchidism, Muscular hypotonia, Maternal diabetes, Hypothyroidism, Global developmental delay, Febrile seizures, Chorea, Brain atrophy	ID	Yes

(Cont.)

P22	M	Microphthalmos, Lissencephaly, Hypoplasia of the corpus callosum, Hypoplasia of the brainstem, Hydrocephalus, Global developmental delay, Dilation of lateral ventricles, Congenital microcephaly, Cerebral white matter hypoplasia, Bilateral coxa valga	ID	Yes
P23	M	Resting tremor, Obesity, Motor delay, Intention tremor, Intellectual disability (moderate), Generalized myoclonic seizures, Developmental stagnation, Delayed speech and language development, Ataxia	ID	-
P24	M	Plagiocephaly, Limb hypertonia, Infantile axial hypotonia, Global developmental delay, Gastroesophageal reflux, Dysphagia, Delayed myelination, Cyanosis, Abnormality of the cerebral white matter	ID	-
P25	F (Fetus)	Ventriculomegaly, Type II lissencephaly, Talipes equinovarus, Retinal dysplasia, Periorbital edema, Hypotelorism, Hydromyelia, Hydrocephalus, Encephalocele, Abnormality of the diencephalon	ID	-
P26	M	Proximal muscle weakness, Muscular hypotonia, Joint laxity, Dyslexia, Disproportionate tall stature	Mito	-
P27	M	Ventriculomegaly, Retinal coloboma, Optic nerve hypoplasia, Moderate global developmental delay, Microtia, Growth delay, Encephalocele, Congenital microcephaly, Cleft palate, Choanal stenosis	ID	-
P28	M	Respiratory tract infection, Microcephaly, Global developmental delay, Generalized tonic-clonic seizures, Generalized myoclonic seizures, Gastroparesis, Gastroesophageal reflux, Developmental regression, Cerebellar atrophy, Absence seizures	ID	-
P29	F	Villous hypertrophy of choroid plexus, Toe clinodactyly, Seizures, Neonatal hypotonia, Hypertelorism, Hydrocephalus, Global developmental delay, Choroid plexus papilloma, Ascites, Apnea	ID	-
P30	F	Stereotypic behavior, Intellectual disability, Hypoplasia of the brainstem, Global developmental delay, Gastroesophageal reflux, Gait ataxia, Cyanotic episode, Caesarian section, Birth length greater than 97th percentile, Abnormality of the vertebrae	ID	-
P31	M	Intellectual disability, Ataxia	ID	-
P32	F	Seizures, Pain insensitivity, Myopathy, Muscular hypotonia, Moderate global developmental delay, Lactic acidosis, Increased muscle lipid content, Genu valgum, Gait ataxia, Easy fatigability	Mito	-
P33	M	Vocal cord paresis, Thoracolumbar kyphoscoliosis, Skeletal muscle atrophy, Peripheral axonal neuropathy, Hypoventilation, Hearing impairment, Distal muscle weakness, Central sleep apnea, Atrophy/Degeneration involving the spinal cord, Aortic regurgitation	Neuro	-
P34	Fetus	Short femur, Abnormality of the skull, Abnormality of the femur	Skeletal	-
P35	M	Short stature, Progressive macrocephaly, Pectus excavatum, Motor delay, Infantile axial hypotonia, Global developmental delay, Gastroesophageal reflux, Delayed myelination, Calcaneovalgus deformity	Mito	-
P36	M	Seizures, Proximal muscle weakness, Muscular hypotonia, Joint laxity, Global developmental delay, Gait disturbance, Delayed speech and language development	ID	-
P37	M	Peters anomaly, Optic nerve hypoplasia, Maternal diabetes, Macular hypoplasia, Delayed gross motor development, Decreased body weight, Cataract	Vision	-
P38	F	Scoliosis, Myopathic facies, Muscular hypotonia, Mixed demyelinating and axonal polyneuropathy, Macrocephaly, Leukoencephalopathy, Intellectual disability (mild), Global developmental delay Generalized amyotrophy	ID	-
P39	M	Sleep disturbance, Maternal hypertension, Leukodystrophy, Intrauterine growth retardation, Hearing impairment, Global developmental delay, Congenital nystagmus, Congenital microcephaly, Cerebral hypomyelination, Birth length less than 3rd percentile	Neuro	-
P40	M	Bone dysplasia, deafness, optic neuropathy	Skeletal	-
P41	F	Upper limb undergrowth, Micrognathia, Microcephaly, Lower limb undergrowth, Hypothyroidism, Femoral bowing, Cardiomyopathy, Aplasia/Hypoplasia of the thumb, Aplasia/Hypoplasia of the radius	MCA	-
P42	M	Mixed motor sensory neuropathy	Neuro	-
P43	M	Type 2 muscle fiber predominance, Severe global developmental delay, Sensorineural hearing impairment, Sensorimotor neuropathy, Osteopenia, Neurogenic bladder, Global developmental delay, Congenital cataract, Central core regions in muscle fibers, Abnormality of the cerebral white matter	ID	-

(Cont.)

P44	M	Ventricular hypertrophy, Stroke-like episodes, Seizures, Scoliosis, Recurrent infections, Limb-girdle muscle weakness, Lactic acidosis, Hemianopia, Global developmental delay, Developmental regression	Mito	-
P45	F	Thrombocytopenia, Palmar hyperkeratosis, Obesity, Hyperkeratosis, Depression	Skin	-
P46	M	Seizures, Progressive cerebellar ataxia, Obesity, Myoclonus, Muscular hypotonia, Motor polyneuropathy, Developmental regression, Congenital sensorineural hearing impairment, Chorea, Cerebellar atrophy	Neuro	-
P47	F	Ptosis, Proximal muscle weakness, Motor delay, Laryngomalacia, Hepatomegaly, Gastroesophageal reflux, Elevated hepatic transaminases, Congenital muscular torticollis	Neuro	-
P48	F	Scoliosis, Posterior embryotoxon, Muscle weakness, Hypopigmented skin patches, EMG: neuropathic changes, Dilated, ardiomyopathy, Corneal opacity, Cerebellar hypoplasia, Burkitt lymphoma, Abnormality of the retina	Neuro	-
P49	M	Neuronal loss in the cerebral cortex, Gliosis, Focal seizures, Encephalopathy, Delayed myelination, Cerebral atrophy	Mito	-
P50	M	Type 2 muscle fiber atrophy, Seizures, Mitochondrial encephalopathy, Lactic acidosis, Infantile axial hypotonia, Increased body weight, Hyperreflexia, Feeding difficulties, Easy fatigability, Developmental regression	Mito	-
P51	M	T-cell lymphoma, Seizures, Muscular hypotonia, Global developmental delay, Decreased body weight, Cortical gyral simplification, Congenital microcephaly, Colpocephaly, Cerebellar hypoplasia, Agenesis of corpus callosum	ID	-
P52	M	Limb pain, Hearing impairment, Dysarthria, Delayed fine motor development, Chorea, Attention deficit hyperactivity disorder, Abnormal pyramidal signs	Neuro	-
P53	M	Ventriculomegaly, Subarachnoid hemorrhage, Fetal pyelectasis, Cerebral arteriovenous malformation, Agenesis of corpus callosum, Abnormal cortical gyration	ID	-
P54	M	Scoliosis, Neonatal hypotonia, Limb-girdle muscle weakness, Horseshoe kidney, Glomerulopathy, Dilatation of the aortic arch, Defect in the atrial septum, Abnormal renal morphology, Abnormal mitochondria in muscle tissue	Mito	-
P55	F	Neuropathy	Neuro	-
P56	M	Retinal coloboma, Microphthalmos, Megalocornea	Vision	-
P57	F	Short stature, Renal tubular acidosis, Renal insufficiency, Pulmonary hypertension Myelodysplasia, Intrauterine growth retardation, Hypophosphatemic rickets, Elevated hepatic transaminases, Defect in the atrial septum, Congenital lactic acidosis	Mito	-
P58	M	Small for gestational age, Polymicrogyria, Patent foramen ovale, Nephrotic syndrome, Microcephaly, Hypertonia, Decreased body weight, Congenital microcephaly	ID	-
P59	F	Severe global developmental delay, Seizures, Optic nerve hypoplasia, Muscular hypotonia of the trunk, Microcephaly, Macrotia, Limb hypertonia, Dysphagia, Defect in the atrial septum	ID	-
P60	M	X-linked developmental delay	ID	-
P61	M	Vomiting, Stellate iris, Seizures, Hyperopic astigmatism, High palate, Global developmental delay, Gastroparesis, Gastroesophageal reflux, Encephalopathy, Congenital microcephaly	ID	-
P62	F	Spinal rigidity, Scoliosis, Neonatal respiratory distress, Motor delay, Intrauterine growth retardation, Generalized muscle weakness, Feeding difficulties, Fatigable weakness of respiratory muscles, Dysphagia, Decreased body weight	Mito	-
P63	M	Mitochondrial myopathy, Global developmental delay, Choreoathetosis, Abnormality of the basal ganglia	Mito	-
P64	F	Upper limb hypertonia, Truncal ataxia, Sleep myoclonus, Seizures, Ptosis, Muscular hypotonia of the trunk, Hemiparesis, Global developmental delay, Encephalopathy, Blepharophimosis	ID	-

F – Female; M – Male; ID – Intellectual disability; Mito – Mitochondrial; Neuro – Neuropathies; Skeletal – Skeletal dysplasia; MCA – Multiple congenital anomalies;

Table XVIII – Clinical overview of the 64 re-analyzed patients

Gender
Males **39 (61%)**
Females **24 (37%)**
Foetus **1 (2%)**
Consanguinity
Consanguineous **13 (20%)**
Non-consanguineous **45 (70%)**
Not determined **6 (10%)**
Gene panel applied
Intellectual disability **33 (52%)**
Mitochondrial disorders **14 (22%)**
Neuropathies **9 (14%)**
Multiple congenital anomalies **2 (3%)**
Vision disorders **2 (3%)**
Skeletal dysplasias **2 (3%)**
Skin disorders **2 (3%)**

Yield of the re-analysis

After reanalysing 64 patients in 2016, we were able to find pathogenic variants in 3 patients, likely pathogenic variants in 9 and variants of uncertain significance in 10 cases, leaving only 33 cases still fully unsolved (figure 1; table IV).

Table XIX – Number and type of variants found in the re-analyzed patients

Variant classification according to ACMG	**Variant type**			
	Recessive Homozygous	Recessive Compound heterozygous	Heterozygous	**TOTAL**
Pathogenic	0	1	2	3 (5%)
Likely pathogenic	3	3	3	9 (14%)
Uncertain significance	1	5	4	10 (16%)
TOTAL	4	9	9	22 (35%)

The use of a virtual panel to filter the variants for each patient proved to be a useful strategy when re-analyzing cases, retrieving a total of 17% of the cases solved. For the ID patients, the panels allowed the detection of a candidate variant for 9 cases (which corresponds to 27% of the total ID patients), while for the mitochondrial disorders and multiple congenital anomalies only one case was solved in each panel (50% in each) (table IV).

Whenever a panel application didn't retrieve a candidate variant worth pursuing, the analysis was carried "outside the panel" considering OMIM genes first and only when negative moving for non-

OMIM genes. Considering the global yield achieved used the combined approach of genes "in a panel" and "outside a panel" we could solve a total of the 34% of the patients re-analysed 2-3 years down the road (table IV and figure 1). Looking back at the year of the first publication of the newly identified genes, 4 genes were firstly published in 2016, 4 genes in 2015, 2 genes in 2014 and 3 genes in 2013 (figure 2).

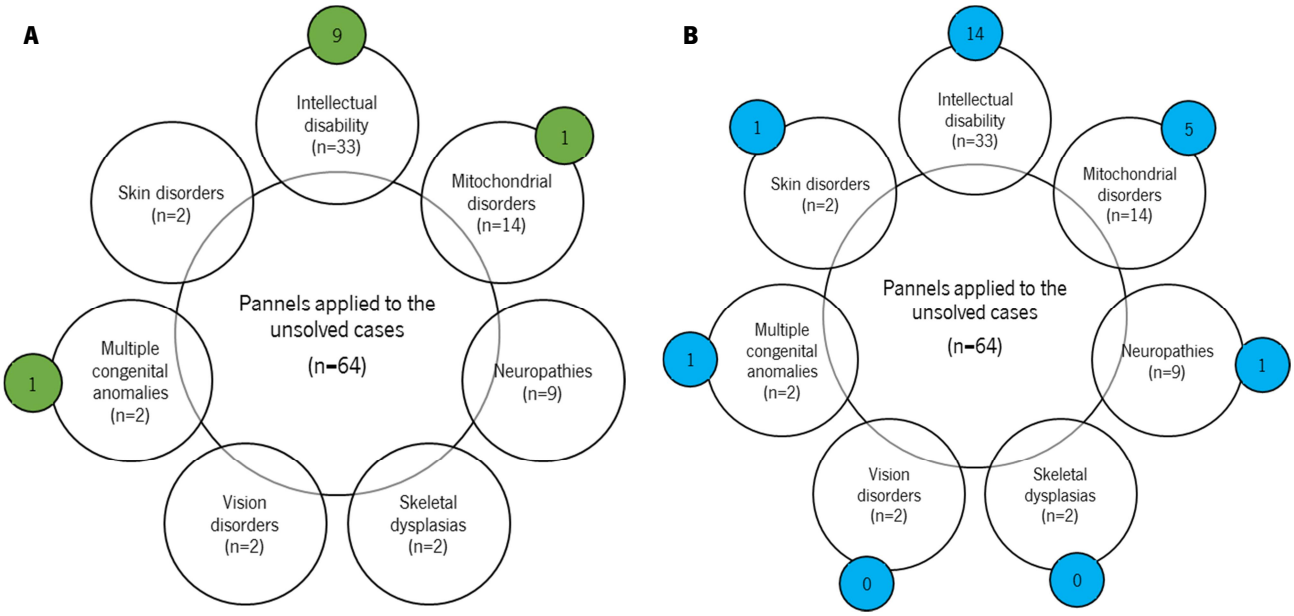


Figure 39 – Yield of re-analysis for the 64 patients. (A) Number of cases for which a candidate gene was found using the panel (green circles). (B) Number of cases for which a candidate gene was found using the combination of panel and, whenever negative, the Orphanet genes (blue circles).

Table XX - Clinical division of the patients according with the panels applied

Phenotypes	No. Patients	Unsolved	Solved				Yield (panel) %	Yield (total) %
			Detected in the panel	Detected outside the panel	Exomiser	Total		
Intellectual disability	33	19	9	4	1	14	27%	42%
Mitochondrial disorders	14	9	1	4	0	5	7%	36%
Multiple congenital anomalies	2	1	1	0	0	1	50%	50%
Neuropathies	9	8	0	1	0	1	0%	11%
Vision disorders	2	2	0	0	0	0	0%	0%
Skeletal dysplasia	2	2	0	0	0	0	0%	0%
Skin disorders	2	1	0	1	0	1	0%	50%
TOTAL	**64**	**42**	**11**	**10**	**1**	**22**	**17%**	**34%**

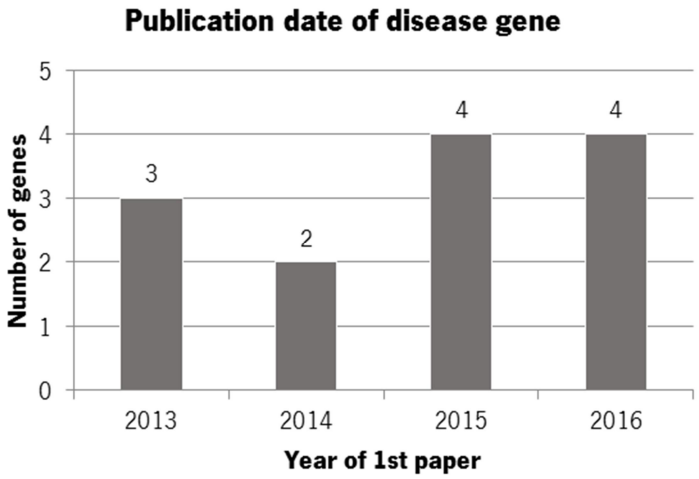


Figure 40 – Number of genes identified during the re-analysis which were described as disease associated for the first time in the last four years.

Variants found in the re-analysed patients

For all the patients the ACMG variant classification was used (Richards *et al.*, 2015). The 22 patients for which at least a variant classified as (I) pathogenic, (II) likely pathogenic or (III) of uncertain significance was found are listed in table V and a detailed description of the clinic and interpretation of the findings is provided for each.

Table XXI – List of re-analyzed patients with variants classified as pathogenic, likely pathogenic or uncertain significance (according to the ACMG guidelines)

Patient	Sex	Panel applied	Detected in panel	Gene	Zygosity	NM	Variant(s)	ACMG classification	Frequency (ExAC)	ClinVar	Bioinformatic prediction	
											SIFT	PolyP.
P1	M	ID, Mito	Yes (Mito)	*FBXL4*	Het Het	NM_012160	c.G1304A:p.R435Q, c.C1303T:p.R435X	Pathogenic	0,00002475, 0,00004949	NP, Pathogenic	0,3, -	1, -
P2	M	ID	No	*KMT2B*	Het	NM_014727	c.C1348T:p.Q450X	Pathogenic	-	NP	-	-
P3	F	MCA	Yes	*MAGEL2*	Het	NM_019066	c.1996dupC p.Q666Pfs	Pathogenic	0,0001	Pathogenic	-	-
P4	F	Skin	No	*DNAJC21*	Homo	NM_001012339	c.100A>G (p.K34E)	Likely pathogenic	-	NP	0	1
P5	F	ID	No	*MYOD1*	Homo	NM_002478	c.G697T; p.E233X	Likely pathogenic	-	NP	-	-
P6	M	Neuro	No	*PAX7*	Homo	NM_002584	c.433C>T (p.R145X)	Likely pathogenic	-	NP	0	1
P7	M	Mito	No	*NBAS*	Het Het	NM_015909	c.T1754C:p.L585P c.686dupT:p.F229fs	Likely pathogenic	0,000008253; 0,0005	NP; NP	0; -	0,998; -
P8	M	ID	Yes	*RARS2*	Het Het	NM_020320	c.G1026A:p.M342I, c.G998A:p.R333Q	Likely pathogenic	0,0002, 0,00003346	NP	0,51, 0	0,241, 1
P9	M	Mito	No	*SPATA5*	Het Het	NM_145207	c.A1763T p.Asp588Val; c.A2189G p.Asp730Gly	Likely pathogenic	-	NP	0; 0	0,956; 1
P10	F	ID	Yes	*AUTS2*	Het	NM_015570	c.C1483T:p.R495X	Likely pathogenic	-	Pathogenic	-	-
P11	F	Mito	No	*MYH7*	Het	NM_000257	c.G1606A:p.E536K	Likely pathogenic	-	NP	0	0,96

(Cont.)

P12	M	ID	Yes	*SMARCB1*	Het	NM_003073	c.G110A:p.R37H	Likely pathogenic	-	Pathogenic	0,13	0,99
P13	M	Mito	No	*VARS*	Homo	NM_006295	c.G2840A:p.R947H	Uncertain significance	0,00002484 (het), not present in homo	NP	0	1
P14	M	ID	Yes	*ASPM*	Het Het	NM_018136	c.G7084A:p.V2362I, c.C3053T:p.S1018L	Uncertain significance	0,00001651, 0,00001648	NP	0,52, 0,10	0,93, 0,96
P15	F	ID	No	*CEP104*	Het Het	NM_014704	c.A626G:p.E209G, c.G550C:p.E184Q	Uncertain significance	0,0003, 0,0003	NP, NP ;	0,09. 0,17	1, 0,49
P16	F	ID	Yes; Yes	*WDR62*	Het Het	NM_001083961	c.G2527T:p.D843Y, c.C3748G:p.R1250G	Uncertain significance	0,000008616, 0,000008267	NP, NP	0; 0,37	1, 0,151
P17	F	ID	No	*SLC6A9*	Het Het	NM_201649	c.A1750T:p.I584F, c.T326G:p.F109C	Uncertain significance	-	NP	0,02, 0	0,99, 1
P18	M	ID	No; No	*ZNF335*	Het Het	NM_022095	c.G1665A:p.P555P, c.C1516T:p.R506C	Uncertain significance	0, 0,000008343; -	NP; NP	-, 0,05	-, 1
P19	F	ID	Yes	*ASXL3*	Het	NM_030632	c.A6308G:p.Y2103C	Uncertain significance	0,000008304	NP	0,04	1
P20	F	ID	Yes	*CREBBP*	Het	NM_004380	c.C4709T:p.A1570V	Uncertain significance	-	NP	0,12	0,99
P21	M	ID	Yes	*GATAD2B*	Het	NM_020699	c.A1546G:p.S516G	Uncertain significance	-	NP	0,43	0,13
P22	M	ID	Yes	*TUBB2B*	Het	NM_178012	c.G1228A:p.E410K	Uncertain significance	-	NP	0	1

F -Female; M - Male; ID – Intellectual disability; Het – Heterozygous; Homo – Homozygous; NP – Not present

Patients with variants classified as pathogenic

Patient P1 is 14 years old and presented moderate ID, seizures, attention deficit hyperactivity disorder and aggressive behavior. Additionally he had neutropenia, lactic acidosis, hypopigmented skin patches, hyperalaninemia, hypertrophic cardiomyopathy and ascending aortic dilation. Exome analysis revealed two compound heterozygous variants (a missense c.G1304A; p.R435Q and a nonsense c.C1303T; p.R435X) in the F-Box And Leucine Rich Repeat Protein 4 (*FBXL4*) gene (MIM 605654, RefSeq accession number NM_012160), both of which previously reported in patients with mitochondrial encephalomyopathy (Huemer *et al.*, 2015). Because both variants occur in adjacent nucleotides there was the chance of both occuring in the same allele. Detailed read analysis determined that both variants were never present in the same read and Sanger sequencing confirmed the presence of both variants in the index (parents not yet tested). *FBXL4* encodes a protein located in mitochondria responsible for maintaining mtDNA integrity and stability (Gai *et al.*, 2013). The parents are not consanguineous and both variants are present in ExAC only in heterozygosity, with frequencies of 0,0000248 and 0,0000495. The previously described data and the fact that the patient has several phenotypic feature overlapping with previously reported cases with mutation in *FBXL4* (ID, lactic acidosis, seizures) (Gai *et al.*, 2013) lead us to believe that this is the cause of the disease, even though the parents are not available for study.

Patient P2 is a 19 years old male with borderline ID, microcephaly, motor delay, developmental regression, speech apraxia and severe expressive language delay. He also presents decreased body weight, dystonia, dysphasia and reduced tendon reflexes. Exome analysis revealed a heterozygous stop variant (c.C1348T; p.Q450X) in the Lysine Methyltransferase 2B (*KMT2B*) gene (MIM 606834, RefSeq accession number NM_014727). Recessive mutations in *KMT2B* were once reported in patients with a Kleefstra syndrome-like phenotype (Agha *et al.*, 2014). More recently a publication came out reporting LoF mutations in a group of patients with early-onset generalized dystonia (Zech *et al.*, 2016). Although the p.Q450X variant was never described before (either in patients or controls), it occurs *de novo* in the patient and is located in the same exon as two other known LoF pathogenic variants (exon 3). The patients described by Zech and colleagues present striking similarities with patient P2: motor delay (only one case), MIC, variable mild to absence ID, dystonia and speech delay. Later in the same month, another description of patients harboring deletions affecting *KMT2B,* LoF and missense mutations was published describing 27 unrelated individuals with complex progressive childhood-onset dystonia

with clinical features very similar to the previously reported and with a characteristic brain imaging pattern affecting mostly the globus pallidus (Meyer *et al.*, 2016). The authors also showed that the LoF mutations lead to reduced expression of *KMT2B* in fibroblasts and that these mutations did not appear to have an impact in the global histone H3K4 methylation levels (Meyer *et al.*, 2016). Considering the currently available evidence we believe that the present variant is causing the disease in the patient.

Patient P3 is a female with myopathy, microcephaly, brain atrophy, growth hormone deficiency and central adrenal insufficiency. She also presents hearing impairment, nystagmus, bronchiectasis, hydronephrosis and multiplex congenital arthrogryposis. Exome analysis revealed a heterozygous frameshift variant (c.1996dupC; p.Gln666Profs) in the MAGE Family Member L2 (*MAGEL2*) gene (MIM 605283, RefSeq accession number NM_019066), inherited from the healthy father. Dominant truncating mutations in the paternally inherited allele of *MAGEL2* cause Schaaf-Yang syndrome, a disorder sharing characteristics with Prader-Willi syndrome and presenting different levels of severity, ranging from fetal akinesia (possibly lethal) to developmental delay/ID, hypotonia, feeding problems and autism (Fountain *et al.*, 2017). This same variant was previously described as pathogenic when inherited from the father (Fountain *et al.*, 2017). In fact, the c.1996dupC variant is the most frequently reported variant so far, with a total of 13 patients described in the literature and the majority of which showing arthrogryposis, also present in our patient (Mejlachowicz *et al.*, 2015, p. 2; Fountain *et al.*, 2017).

Patients with variants classified as likely pathogenic

Patient P4 presented a very intriguing clinical picture, composed by an encephalopathy with global developmental delay, hypotonia, delayed myelination and cerebral cortical atrophy. He also presents pancytopenia, scoliosis, osteopenia, mild neurosensory hearing impairment and hepatosplenomegaly. Exome analysis revealed a homozygous nonsense variant (c.A100G; p.K34E) in DnaJ Heat Shock Protein Family (Hsp40) Member C21 (*DNAJC21*) gene (MIM 617048, RefSeq accession number NM_001012339). This variant is located in a splice site (third nucleotide of exon 2), so the cDNA from the patient's (from lymphocytes) was sequenced, leading us to conclude that there was no alteration in splicing (figure 3). Parents were not available for testing but Sanger sequencing confirmed that the brother and two cousins (all similarly affected) also carried the variant in homozygosity.

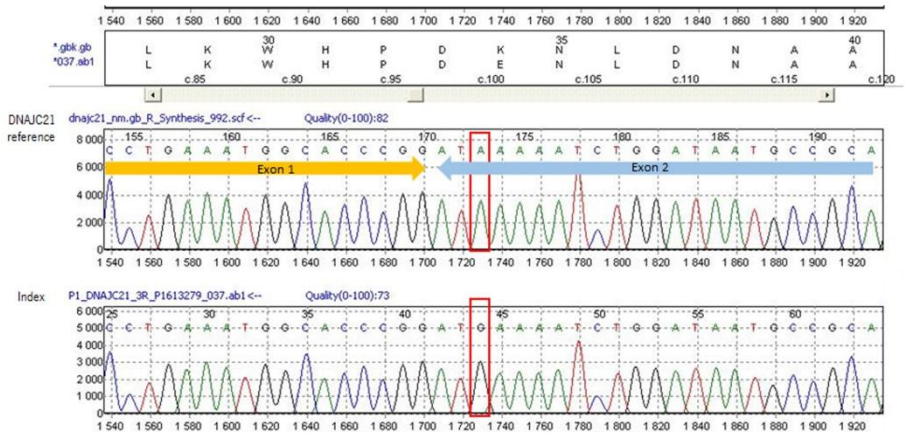


Figure 41 – cDNA Sanger sequencing for the patient with the *DNAJC21* variant. Is possible to observe that there isn´t an alteration in splicing between exon 1 and 2 and therefore the variant c.A100G must act solely as a missense.

Because the clinical presentation was so peculiar, another family with a similar presentation was picked from the same exome project at CHUSJ, in which we found that the affected patients (also two siblings) carried the exact same variant in the same gene. After reviewing the cases and discussion with the responsible physicians we realized that both families share the same ancestry of Algonquin speakers from the First Nations of Canada. This lead us to hypothesize if there could be a founder effect in this population and, to test this hypothesis, we sequenced a pool of 87 controls (24 Ojibwas, 32 Crees and 32 Algonquin speakers).The results revealed the presence of the variant in heterozygosity in 7 of the 32 Algonquin speakers where it is absent from the 24 Ojibwas and 32 Crees controls. Unfortunately, it was not possible to determine if the controls carrying the variant are related (an important factor specially because there is a high inbreeding rate in the population), and there is a small number of available controls to study, making it impossible to determine a true carrier rate. In January 2017 the p.K34E variant (the exact same one as in our patient) was reported in two siblings born from a consanguineous couple of Canadian First Nations ancestry (Dhanraj *et al.*, 2017, p. 21). This new information, together with the fact that this specific variant was never reported before (to our knowledge) and is not present in ExAC database, make us to believe that this could be due to a founder effect of the p.K34E variant in First Nations Canadian populations (Dhanraj *et al.*, 2017). *DNAJC21* encodes a protein thought to be involved in rRNA biogenesis in the nucleolar compartment and in the shuttling of PA2G4 (Proliferation-Associated 2G4) between the cytoplasm and the nucleus (Tummala *et al.*, 2016, p. 21). Variants in *DNAJC21* have been associated with bone marrow failure and Shwachman-Diamond syndrome (which includes bone marrow failure, pancreatic dysfunction and behavioral alterations) (Tummala *et al.*, 2016; Dhanraj *et al.*, 2017). The

patients reported in these works share many clinical features with the patients observed at St Justine Hospital, such as pancytopenia, developmental delay, skeletal and liver anomalies.

Patient P5 (a girl) is the third child to a first-cousin consanguineous couple where the two other siblings (boy and a girl) are healthy (figure 4). She presented to the clinic with respiratory insufficiency due to muscle weakness (need for respiratory assistance when in dorsal decubitus), myopathic facies (triangular), mandibular prognathia, downslanted palpebral fissures and fatigable weakness of swallowing muscles (which resulted if feeding difficulties when younger). Thorax X-ray showed bilateral high diaphragm domes but no lung hypoplasia and abdominal ultrasonography revealed bilateral small kidney (-2DS) without other anomalies. Exome re-analysis in 2016 revealed a homozygous nonsense variant (c.G697T; p.E233X) in Myogenic Differentiation 1 (*MYOD1)* gene (MIM 159970, RefSeq accession number NM_002478). This variant is not described in ExAC database and no stop mutations are described in homozygous state in *MYOD1* gene in ExAC. The zygosity status for this variant was confirmed by Sanger sequencing in the patient (homozygous), parents, brother and sister (all heterozygous carriers) (figure 4 C).

MYOD1 presented itself as the most interesting candidate due to the parental consanguinity, the gene's biological role and the recent description of a homozygous stop variant in humans (Watson *et al.*, 2016). *MYOD1* belongs to a family of four transcription regulators involved in muscle lineage-determination – *MYOD1* (MIM 159970), *MYF5* (MIM 159990), *MYOG* (MIM 159980) and *MYF6* (MIM 159991) (Sabourin and Rudnicki, 2000). The fact that *Myod1* null mice have a normal muscular phenotype is due to MYF5 protein was overexpressed in the mice model revealing the existence of a compensatory mechanism and redundant function between *Myod1* and *Myf5* (Rudnicki *et al.*, 1993). Later on, it was described that these two genes have different roles in different myogenic lineages that originate epaxial (paraspinal and intercostal) and hypaxial (limb, diaphragm and abdominal) musculature: *Myod1* null embryos exhibit a two day delay in the development of hypaxial musculature while *Myf5* null embryos exhibit the same type of delay but in the epaxial musculature (Kablar *et al.*, 1997). If we take in consideration these mouse studies and try to translate them to the present individual, we can observe that her most severely affected musculature is the one from the diaphragm (a hypaxial muscle) and limbs.

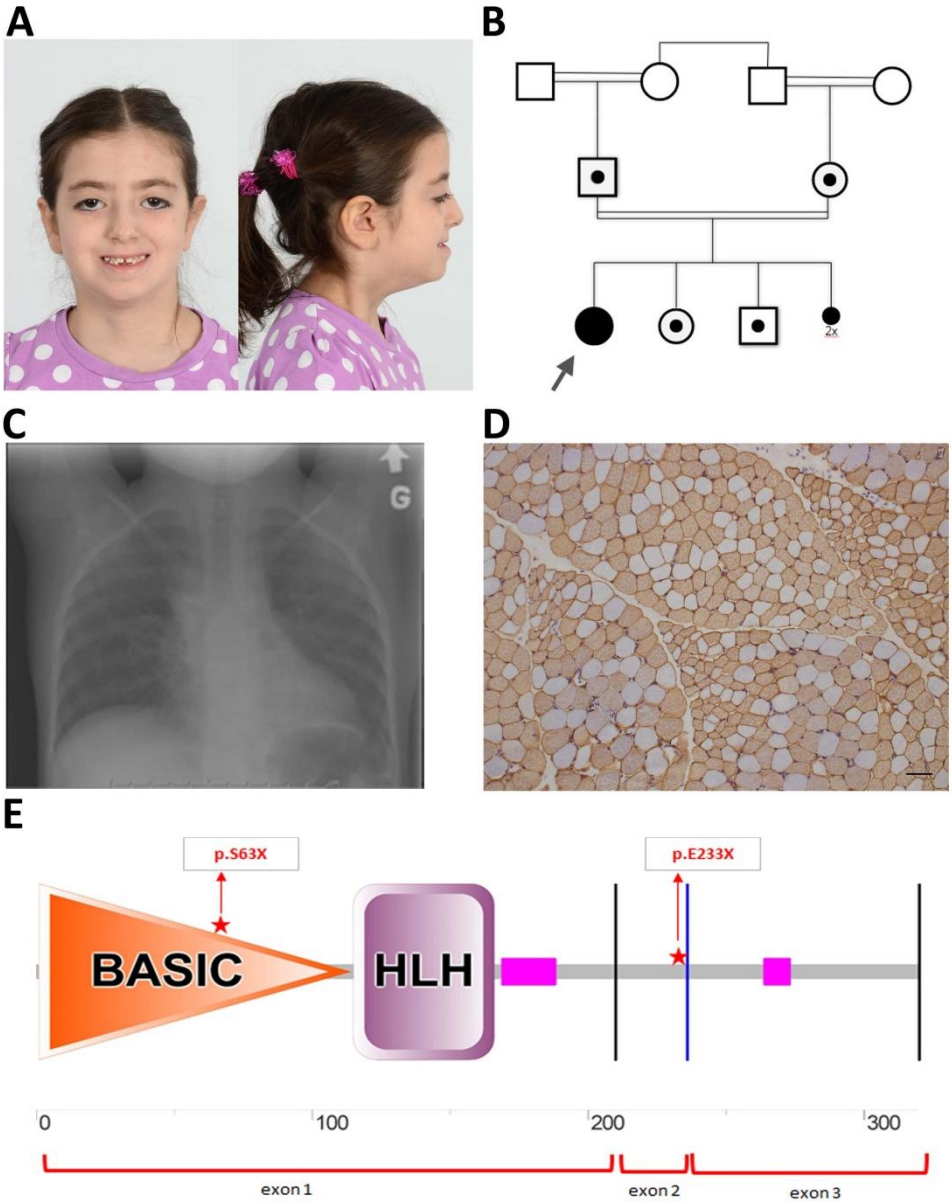


Figure 42 – (A) Frontal and lateral facial appearance at the age of 7.5 years presenting downslanted palpebral fissures, ptosis, proptosis, mandibular prognathia and hypotonic face; **(B)** Family pedigree; **(C)** lung X-ray took at 6 years showing the diaphragm high domes **(D)** muscle biopsy (at 2 years) shows an unusual patter of fiber atrophy/hypotrophy involving the tips of several fascicules and fibre size and shape variations (cryostat section coimmunostained for fast myosin heavy chain and laminin alpha 2; **(E)** schematic representation of the MYOD1 protein domains. The domain structure was determined based on SMART webtool. The variant identified in this gene in the present study (E233X) and in the study by Watson C, 2015 (S63X) are indicated by a red star with its respective location at exon and protein levels (aa).

The specific role of MYOD1 in the diaphragm was addressed in a study by Staib J. et al in 2002, in which the authors observed a downward shift in the force-velocity relationship, a reduced maximal isometric force production and a reduced diaphragmatic maximal velocity in *Myod1* null mice (Staib *et al.*, 2002). Compound mutant embryos for *Myod1* and *Dmd* present the essential

MYOD1 and MYF5-dependent musculature, but their diaphragm is very thin and not functional, leading to lung hypoplasia and neonatal death. The authors also concluded that the lung hypoplasia is likely due to the absence of mechanical forces originated by the diaphragm and that, even though in other tissues MYF5 seems to be able to replace MYOD1, that is not the case for diaphragm development (Inanlou *et al.*, 2003; Inanlou and Kablar, 2003).

MYOD1 can affect diaphragmatic development and contractile properties through several mechanisms. One of those is by the reduced expression of desmin, seen in skeletal muscle of *Myod1* null mice (Yablonka-Reuveni *et al.*, 1999). Desmin is a structural protein of the extrasarcomeric cytoskeleton important for the sarcomere proper function, reduction of which can affect muscle contraction (Staib *et al.*, 2002). Another possible mechanism is the reduced response of *Myod1* null mice fast muscle fibers to Ca^{2+} activation due to differences in troponin T isoform expression (Metzger *et al.*, 1995).

In summary, although the contribution of MYOD1 for muscle development has been widely studied for decades, only recently the first patients with germline *MYOD1* mutation were reported (Watson *et al.*, 2016) and we are presenting the second instance of such mutation. As happened many times before, when a new gene is associated with a disease the core phenotype and variability is more precisely defined as the number of patients described increases. Although the severity of the phenotype is remarkably different, there are still many similarities between the reported and the present individual (detailed clinical comparison in table VI). All the individuals have a triangular face, generalized muscle weakness, renal anomalies and alterations in diaphragm function with consequent respiratory insufficiency. One explanation for the difference in severity could be the site of the protein where the mutation occurs. In the previously reported individuals the mutation occurred in exon 1, affecting the basic motif of the bHLH protein domain, while in the present case it occurred on exon 2, outside of any (known) core protein domain (figure 4C). While the HLH motif of the bHLH domain functions in the interaction with other proteins by forming complexes, the basic motif is the responsible for the binding of the protein to the DNA (Jones, 2004). Taking this in consideration, we could hypothesize that the variant described here could affect the protein in a less severe manner, leading only to the ablation of MYOD1 terminal region and possibly keeping intact the bHLH domain.

Table XXII - Clinical comparison of the four individuals with *MYOD1* homozygous variants.

Clinical	Present individual	Watson et al, 2016		
		III.1	III.2	III.4
Gender	♀	♂	♂	♀
Family history				
Consanguinity	+	+	+	+
Prenatal and perinatal history				
Prenatal anomalies	-	cystic hygroma polyhydramnios	polyhydramnios	cystic hygroma
Birth	At term	35+5	35+1	37
Apgar	ND	1-1	1-1	1-1
Neonatal death	-	+	+	+
Cranio facial symptoms				
Triangular face	+	+	+	+
Downslanted palpebral fissures	+	+	ND	ND
Ptosis	+	ND	ND	ND
Proptosis	+	ND	ND	ND
Mandibular prognathia	+	ND	ND	ND
Cleft palate	Ogival	+	+	+
Dental malocclusion	+	ND	ND	ND
Respiratory Symptoms				
Respiratory insufficiency due to muscle weakness	+	+	+	+
Diaphragm	high domes	right-sided eventration	very high domes	extremely high domes
Lung hypoplasia	-	+	+	+
Musculoskeletal symptoms				
Generalized muscle weakness	+	+	+	+
Fatigable weakness of swallowing muscles	+	ND	ND	ND
Clinodactyly/ digit overlapping	+	+	+	-
Cutaneous symptoms				
Congenital, generalized hypertrichosis	+	ND	ND	ND
Genitourinary symptoms				
cryptorchidism	NA	+	unilateral	NA
renal anomaly	Small kidneys (-2SD)	bilateral renal pelvis distension	hydronephrosis duplex kidney	renal hypoplasia

Legend: + present; - absent; NA not applicable; ND not described; SD standard deviation; ♀ female; ♂ male

P6 is 6 years and the only son of consanguineous parents and presents symptomatology of congenital myopathy with progressive skeletal muscle atrophy, generalized hypotonia, congenital bilateral ptosis, dysphagia, scoliosis and delayed CNS myelination. Additionally she also presents genitourinary anomalies such as hydronephrosis, abnormal renal corticomedullary differentiation

and cryptorchidism. Exome analysis revealed a homozygous nonsense variant (c.C433T; p.R145X) in Paired Box Gene 7 (*PAX7)* (MIM 167410, RefSeq accession number NM_013945). This variant is not described in ExAC database (heterozygous nor homozygous state). The zygosity status for this variant was confirmed by Sanger sequencing in the individual (homozygous) but the parents were not available for testing. *PAX7* gene function has been widely studied in mouse models, and it is known to be a transcription factor involved in the maintenance and renewal of satellite cells in muscle (Olguín and Pisconti, 2012, p. 7). However, to the best our knowledge *PAX7* mutations in humans were never reported. Pax7-/- mice exhibit muscle weakness, abnormal gait and splayed hind limbs (Seale *et al.*, 2000). On histological analysis, *Pax7* KO mouse muscles show a normal histological organization but reduced fiber diameter, possibly due to *Pax7* being essential for the renewal and maintenance of the muscle satellite cells, and its absence leading to an impairment in muscle mass maintenance and regeneration (Oustanina *et al.*, 2004). Considering the mouse data regarding the biological role of *Pax7* and the severity of the variant is possible that *PAX7* loss of function may result in the progressive skeletal atrophy and increasingly deteriorating condition of the patient.

Patient P7 presented since the neonatal period symptoms of mitochondrial depletion, hypoglycemia, vomiting and chronic hepatitis, which lead to hepatic failure and consequent death. Exome analysis revealed two compound heterozygous variants (c.T1754C; p.L585P and c.686dupT; p.F229fs) in the Neuroblastoma Amplified Sequence (*NBAS*) gene (MIM 608025, RefSeq accession number NM_015909). Both parents are heterozygous carriers for the variants. Until the middle of 2015 *NBAS* gene had only been associated once in the literature with a syndrome characterized by short stature with optic atrophy and Pelger-Huët anomaly (SOPH syndrome) (Maksimova *et al.*, 2010). It was only in 2015 (after his death) that biallelic mutations in *NBAS* were first described in patients with acute liver failure and consequent encephalopathy (Haack *et al.*, 2015), an overlapping clinical presentation with the one present in our patient. Meanwhile, more publications have reported patients with autosomal recessive mutations in the *NBAS* gene with different phenotypic outcomes: hepatic encephalopathy due to liver failure (Capo-Chichi *et al.*, 2015; Staufner *et al.*, 2016), SOPH syndrome without symptoms of severe liver disease (Kortüm *et al.*, 2017) and patients with a wider phenotypic spectrum of a multisystemic disease involving liver, eye, skeletal, connective tissue and immune system (Segarra *et al.*, 2015). The findings in the last years suggest that the phenotype associated with *NBAS* recessive mutations is wider than expected in 2015. *NBAS* encodes a protein that is a

component of the syntaxin 18 complex, which is involved in ER membrane fusion and the transport from the Golgi complex to the endoplasmatic reticulum (Aoki *et al.*, 2009). In zebrafish, the depletion of *NBAS* orthologue 8 (*nbas*) results in developmental defects and embryonic lethality (Anastasaki *et al.*, 2011).

Patient P8 is a 4 years old male that presented global developmental delay together with lactic acidosis, hypoglycemia, infantile spasms, tachypnea, strabismus and hypermetropia. Exome analysis revealed two compound heterozygous nonsense variants (c.G1026A; p.M342I and c.G998A; p.R333Q) in Arginyl-TRNA Synthetase 2 (*RARS2*) gene (MIM 611524, RefSeq accession number NM_020320). *RARS2* encodes the mitochondrial arginyl-tRNA synthetase 2 protein, important for protein synthesis in mitochondria, and has been considered for many years as causative of pontocerebellar hypoplasia type 6 (Edvardson *et al.*, 2007). More recently, *RARS2* recessive mutations have been reported in patients with mitochondrial encephalopathy without the key neuroimaging abnormalities of pontocerebellar hypoplasia type 6 (such as cerebellar hypoplasia and pontocerebellar and cortical atrophy (van Dijk *et al.*, 2016; Lühl *et al.*, 2016, p. 2). We believe that our patient might fit in the same category of cases without brain structural anomalies. Both variants are predicted to be pathogenic and present in ExAC database in heterozygosity at frequencies of 0,0002 and 0,00003346, respectively. Sanger sequencing was not possible to perform in the parents.

Patient P9 is an eight years old male with an epileptic encephalopathy presenting cerebral atrophy, delayed myelination, seizures, optic atrophy with blindness and sensorineural hearing impairment. He also presents a motor delay, spastic tetraplegia and muscular hypotonia of the trunk. Exome analysis revealed two compound heterozygous variants (c.A1763T; p.Asp588Val and a c.A2189G; p.Asp730Gly) in Spermatogenesis Associated 5 (*SPATA5*) gene (MIM 613940, RefSeq accession number NM_145207). *SPATA5* encodes an ATPase protein that is ubiquitously expressed (including in all brain regions) and works in the post-translational modification of mitochondrial morphogenesis during early spermatogenesis (Liu *et al.*, 2000). Functional studies are necessary to determine that exact function of SPATA5 in the brain but the number of patients with recessive mutations and concordant phenotype described make it its contribution in neuronal development very likely (Tanaka *et al.*, 2015). Recessive mutations in *SPATA5* are associated with an infantile encephalopathy characterized by ID, MIC, epilepsy, hearing loss and altered muscular tonus (Kurata *et al.*, 2016), a phenotype highly overlapping with the present patient. Although none of the variants in the patient has been previously described (no entrance

in ClinVar) both are predicted to be pathogenic and were confirmed by Sanger in the index (parents not available for testing).

Patient P10 is a seven years old female with global developmental delay, congenital microcephaly, plagiocephaly, muscular hypotonia of the trunk, limb hypertonia and intention tremor. She carries a heterozygous stop variant (c.C1483T; p.R495X) in the Autism Susceptibility Candidate 2 (*AUTS2*) gene (MIM 607270, RefSeq accession number NM_015570). Even though the parents are not available for testing and so the segregation pattern or *de novo* origin cannot be determined, this variant is described as pathogenic in ClinVar (single submitter, ID372310). *AUTS2* encodes a cytoplasmatic protein that regulates Rho family GTPases Rac1 and Cdc42 to induce lamellipodia and suppress filopodia (respectively) in cortical development and neuritogenesis (Hori *et al.*, 2014, p. 2). *AUTS2* mutations are known to cause ID (of variable severity), microcephaly, feeding problems and behavioral alterations (obsessive behavior), a comparable clinical presentation to that of the present patient (Beunders *et al.*, 2016, p. 2).

Patient P11 is a 12 years old female with symptoms of mitochondrial myopathy together with developmental regression, cerebral atrophy, hyporeflexia, hepatomegaly and decreased body weight. She also presents a series of cardiac problems such as dilated cardiomyopathy, ventricular septal defect and coarctation of the aorta. Exome analysis revealed heterozygous missense variant (c.G1606A; p.E536K) in Myosin Heavy Chain 7 (*MYH7*) gene (MIM 160760, RefSeq accession number NM_000257), which, even it though is unlikely to justify the entire presentation of the patient, might be contributing for some of the features. *MYH7* encodes a slow/b-cardiac myosin heavy chain and is expressed in slow fibers of cardiac and skeletal muscles (Tajsharghi *et al.*, 2003). Mutations in the so called neck domain are usually associated with hypertrophic cardiomyopathy whereas mutations in the more distal regions often lead to a broader phenotype (skeletal myopathies with or without cardiac muscle alterations) (Fiorillo *et al.*, 2016). Even though the variant p.E536K has not been described so far in any other patient, another missense variant in the same amino acid residue was reported as likely pathogenic in ClinVar and the variant itself occurs in the motor (neck) domain, is not present in ExAC and is predicted to be pathogenic. This led us to hypothesize about its role in the dilated cardiomyopathy symptomatology of the patient. Information about the carrier status of both parents and their healthy/affected condition would be of great value to the final interpretation of the variant significance, but unfortunately they are not available for study.

Patient P12 is a 29 years old male with ID, encephalopathy, choreoathetosis, macrocephaly, hydrocephalus, scoliosis, hypotonia, hypertelorism, choreoathetosis and dysphagia. He also presents automutilation features that lead to the initial suspicion of Lesch-Nyhan syndrome. WES revealed a heterozygous variant (c.G110A; p.R37H) in SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily B, Member 1 (*SMARCB1*) gene (MIM 601607, RefSeq accession number NM_003073) previously described in patients with ID, hydrocephalus and myopia (also present in our patient) (Santen *et al.*, 2013). The p.R37H variant was previously described in a patient with ID, hypotonia and brachycephaly but without behavioral alterations (Kleefstra *et al.*, 2012, p. 1). This is in accordance with the report of only 50% of the patients carrying *SMARCB1* mutations presenting self-injury behavior (Santen *et al.*, 2013). Even though the parents are not available for Sanger confirmation there is strong evidence that the variant is causing the disease in the patient.

Patients with variants classified as uncertain significance

Patient P13 is an 11 years old boy with severe developmental delay and ID, microcephaly, cerebral white matter agenesis, cerebral atrophy, aplasia/hypoplasia of the cerebellum with vermis atrophy, and corpus callosum atrophy. He presented episodic metabolic acidosis and intermittent lactic academia, and also presents myopathy, axial hypotonia and spasticity. The patient carries a homozygous missense variant (c.G2840A, p.R947H) in the Valyl-tRNA synthetase (*VARS*) gene (MIM 192150, RefSeq accession number NM_006295). The parents were not available for testing. The first time *VARS* was associated with disease was in November 2015, when it was reported in three patients with microcephaly, DD/ID, seizures and cortical atrophy, a similar clinical presentation to the one in patient P13 (Karaca *et al.*, 2015). Due to the predicted pathogenicity of the variant, the fact that the parents are consanguineous and the overlapping clinic with the reported cases we believe that *VARS* is causing the disease in the patient. *VARS* encodes a protein that belongs to the aminoacyl-tRNA synthetases group. These enzymes are essential for the first step of protein translation – charging tRNA molecules with cognate amino acids (Simons *et al.*, 2015). Although *VARS* mutations were only described once in the literature, there is a significant amount of recently described mutations in other aminoacyl-tRNA synthetase genes in patients with neurological disorders, namely involving the cerebelum, evidencing once again the important contribution of RNA metabolism pathways for brain development (Diodato *et al.*, 2014; Taylor *et al.*, 2014; Simons *et al.*, 2015; Jiang *et al.*, 2016).

Patient P14 is a seven years old male with global developmental delay, microcephaly, abnormal cortical gyration and CNS myelination, seizures and spastic tetraparesis. He also shows feeding difficulties (dysphagia) with decreased body weight and breathing anomalies (dyspnea). Exome analysis revealed two compound heterozygous variants (c.G7084A; p.V2362I and c.C3053T; p.S1018L) in the Abnormal Spindle Microtubule Assembly (*ASPM*) gene (MIM 605481, RefSeq accession number NM_018136). Even though the parents' DNA was not available for testing, the combination of both variants can be accounting for the disease in the patient. APSM recessive mutations were described in patients with microcephaly together with simplified gyral pattern and brain malformations, a presentation fitting that of our patient (Nicholas *et al.*, 2009; Passemard *et al.*, 2009). In 2009, Saadi *et al* described a consanguineous family where individuals affected with compound heterozygous *ASPM* variants had ID, MIC, simplified cortical gyration and low birth weigh, a phenotype similar to the present patient (Saadi *et al.*, 2009). The *ASPM* gene encodes a protein known to be essential for maintenance of neural progenitors at early stages of corticogenesis and for neuronal migration at later stages (Buchman *et al.*, 2011). Also, as mentioned before, ASPM works together with WDR62 in the neuronal apical complex orientation and fate (Jayaraman *et al.*, 2016, p. 62). Both variants are present in ExAC database in heterozygosity at frequencies of ~2/120 000 which is consistent with the hypothesis of both parents being carriers, even though in this case the parents are non-consanguineous.

Patient P15 is a 15 years old female with moderate ID, seizures, short stature and obesity. Exome analysis revealed two compound heterozygous nonsense variants (c.A626G:p.E209G, c.G550C:p.E184Q) in the Centrosomal Protein 104 (*CEP104*) gene (MIM 616690, RefSeq accession number NM_014704). CEP104 was described mutated in patients with Joubert syndrome, a clinical presentation where the patient doesn´t seem to fit (Srour *et al.*, 2015). However, only nonsense and splice site mutations were described in the patients with Joubert syndrome, the effect of recessive compound missense variants being still unknown. Nevertheless, *CEP104* encodes a protein involved in cilia formation and elongation and which is thought to bind to tubulin putting it at an interesting position from the functional point of view as an ID-causing candidate gene (Al-Jassar *et al.*, 2017). Both variants have a relatively high frequency in ExAC database in heterozygosity (0,0003) and are not described in ClinVar. Unfortunately, the parents were not available for testing, therefore the inheritance is not known, an important point for determining the eventual association of *CEP104* with the patient´s phenotype.

Patient P16 is a nine years old daughter of consanguineous parents and has severe ID, seizures, corpus callosum and pons hypoplasia, cerebellar vermis aplasia and abnormality of the periventricular white matter. Additionally she also has muscular hypotonia and obesity. She carries two compound heterozygous nonsense variants (c.G2527T:p.D843Y, c.C3748G:p.R1250G) in WD Repeat Domain 62 (*WDR62*) gene (MIM 613583, RefSeq accession number NM_001083961). Recessive mutations in *WDR62* are described to cause microcephaly together with other structural brain malformations (e.g. cerebellar hypoplasia, lissencephaly, corpus callosum hypoplasia) (Bilgüvar *et al.*, 2010, p. 62). The WDR62 protein is involved in neurogenesis through the mitotic progression of neural progenitor cells (NPCs) an its absence leads to abnormal cortical neuron development due to the decreased number of NPCs (Chen *et al.*, 2014, p. 62). Recently it was described that *WDR62* and *ASPM* (another microcephaly associated gene) work together in the correct localization of the apical neuronal structures (Jayaraman *et al.*, 2016, p. 62). The patient does not have microcephaly, even though all the *WDR62* recessive mutation carrying patients described do. However, microcephaly is not the only outcome of *WDR62* mutations, as many other structural brain anomalies has also been reported in *WDR62* mutated patients (Farag *et al.*, 2013). Is fact, our patient shows overlapping brain anomalies with previously reported cases, such as cerebellar hypoplasia resulting in balance impairment (Bilgüvar *et al.*, 2010; Farag *et al.*, 2013). Both variants in the *WDR62* gene are present in the ExAC database only in heterozygosity and none is yet listed in ClinVar database. Only one of the variants is present in the mother, but the father was not possible to study. Nevertheless, the biological role of the gene leads us to suspect that *WDR62* compound heterozygous variants might be contributing for the patient´s phenotype.

Patient P18 is a six years old male who presents microcephaly with ventriculomegaly and widened subarachnoid space. He also has short stature, brachycephaly, spastic tetraplegia, single umbilical artery and facial dysmorphisms (protruding ears and mandibular prognathia). He carries two compound heterozygous variants (a splice site synonymous c.G1665A; p.P555P and a nonsense c.C1516T; p.R506C) in the Zinc Finger Protein 335 (*ZNF335*) gene (MIM 610827, RefSeq accession number NM_022095). Both variants are predicted to be pathogenic, the first one being a possible splice site altering change. *ZNF335* encodes a nuclear zinc finger protein which interacts with H3K4 methyltransferase complexes and binds to the REST promotor, a key regulator of neurogenesis and neuronal differentiation in human (Yang *et al.*, 2012). Additionally, Znf335 KO mice die at the embryonic stage and the conditional knockout presents signs of

microcephaly (Yang *et al.*, 2012). Recessive mutations have been described in patients with microcephaly, developmental delay, limb spasticity, invisible basal ganglia, brainstem hypoplasia and cerebellar atrophy (Nishida *et al.*, 2016; Sato *et al.*, 2016). The functional role of the gene and the overlapping phenotype with the previous reported patients lead us to believe that the *ZNF335* recessive variants are accounting for the disease in the patient. Unfortunately parents were not available for testing.

Patient P19 is a five years old female with global developmental delay, abnormality of the globus pallidus, short stature and obesity. She also presents spastic tetraparesis, hyperactive deep tendon reflexes, dysphagia and eczema. Exome analysis revealed a heterozygous variant (c.A6308G; p.Y2103C) in the Additional Sex Combs Like 3 (*ASXL3*) gene (MIM 615115, RefSeq accession number NM_030632). *ASXL3* encodes a transcriptional regulator expressed in brain for which only loss-of-function mutations have been described, to date, associated with Bainbridge-Ropers syndrome (BRPS, MIM 615485). Patients with BRPS present ID, severe speech impairment, motor delay, feeding problems in infancy, hypotonia and characteristic facial dysmorphisms (long face with poor expression, arched eyebrows, synophrys and downslanting palpebral fissures) (Hori *et al.*, 2016, p. 3; Srivastava *et al.*, 2016, p. 3; Kuechler *et al.*, 2017, p. 3). In the ClinVar database, all the missense variants in *ASXL3* are described as of uncertain significance. This variant is also present in the ExAC database with a frequency of 0,000008304. This, together with the absence of reports of unequivocally pathogenic missense in BRPS, make the determination of inheritance a key factor in the conclusion to take for the present case. Unfortunately, the parents were not available for testing.

Patient P20 is a ten years old female with global developmental delay, mild ID, hypotonia, scoliosis and hypothyroidism. She also presents widened subarachnoid space, seizures, optic atrophy, myopia, microtia and congenital sensorineural hearing impairment. Exome analysis revealed a heterozygous variant (c.C4709T; p.A1570V) in the CREB Binding Protein (*CREBBP*) gene (MIM 600140, RefSeq accession number NM_004380). Dominant mutations in *CREBBP* are known to cause Rubinstein-Taybi syndrome (MIM 180849), a autosomal dominant disease characterized by ID, growth retardation, broad thumbs and halluces and a distinctive facial presentation (Wincent *et al.*, 2016). However, recently mutations in *CREBBP* were described for the first time in a group of patients without Rubinstein-Taybi syndrome that presented with developmental delay, hearing and feeding problems, microcephaly and variable facial dysmorphisms (Menke *et al.*, 2016). All the variants described in these patients occurred in exon

30 and 31, while the variant present in our patient occurs in exon 28 of the gene, corresponding to the HAT (histone acetyltransferase) domain. Sanger analysis confirmed the presence of the variant in heterozygosity in the index and its absence in the mother (father not available for testing). The variant is predicted to be pathogenic. It is possible to establish some overlap between the patients described by Menke and colleagues and the present case (developmental delay, ID and hearing impairment), therefore this variant might contributing for the disease (Menke *et al.*, 2016).

Patient P21 is a six years old male with global developmental delay, brain atrophy and muscular hypotonia. He presents unilateral cryptorchidism, hypothyroidism and episodes of febrile seizures. He carries a heterozygous variant (c.A1546G; p.S516G) in GATA Zinc Finger Domain Containing 2B (*GATAD2B*) gene (MIM 614998, RefSeq accession number NM_020699). The variant is not present in ExAC or ClinVar, and the bioinformatics tools are not concordant regarding its predicted pathogenicity. *GATAD2B* encodes a protein that is a subunit of the transcription repressor complex MeCP1-Mi2–NuRD, involved in gene repression due to deacetylation of methylated nucleosomes (Willemsen *et al.*, 2013). Until now, only dominant loss of function mutations in *GATAD2B* have been described (Willemsen *et al.*, 2013; Hamdan *et al.*, 2014; Luo *et al.*, 2017) as causative of ID, which is not the case of the present variant, a missense with conflicting predictions between *in silico* tools, that occurs outside the CR2 domain (important for the interaction with the methyl binding domain protein 3-MBD3-subunit). Nevertheless, the variant is not present in ExAC database and the phenotype is partially overlapping with that of previous cases (ID and hypotonia). The inheritance of the variant is a key element for the interpretation of these findings; unfortunately, this was not possible since the parents were not available.

Patient P22 is a 15 years old male with global developmental delay, microcephaly and several brain malformations: cerebral white matter hypoplasia, lissencephaly, hypoplasia of the corpus callosum, hypoplasia of the brainstem, hydrocephalus and ventriculomegaly. Additionally, he also has microphthalmia and bilateral coxa valga. He carries a heterozygous variant (c.G1228A, p.E410K) in the Tubulin Beta 2B Class IIb (*TUBB2B*) gene (MIM 612850, RefSeq accession number NM_178012). *TUBB2B* encodes a neuronally expressed tubulin for which dominant mutations were initially associated with polymicrogyria (Guerrini *et al.*, 2012). Currently a wider spectrum of cortical malformations has been reported in patients carrying *TUBB2B* mutations such as lissencephaly and microcephaly (Cushion *et al.*, 2013; Romaniello *et al.*, 2014). The

variant is predicted to be pathogenic, is not present in the ExAC database and, although never described before in another patient, occurs nearby an amino acid where pathogenic variants have occurred. The parents were not available for testing.

Exomiser performance assessment

The reanalysis of previously studied cases is a time consuming task that is not always possible to performe in a timely manner in a diagnostic laboratory. Also, as all the exome analysis, is still a task undertaken by analysis and therefore subjected to human error. For these reasons the use of an automated tool to analyse exomes at together with the manual analysis can be of great advantage. Because the Exomiser software was never used in the lab there was the need to perform an optimization test in a group of 33 already "solved" patients (i.e. with a previously identified pathogenic or likely pathogenic variant). This group was divided in two subgroups: (I) patients with causal variations in genes with more than 10 individual descriptions in the literature (named ">10 patients") and (II) patients with causal variations in genes with less than 10 individual descriptions in the literature (named "<10 patients"). The use of Exomiser in 33 previously solved cases retrieved the known causal variants with a yield of 53%. We were able to see that Exomiser's performance rose up to 65% if we considered only the genes with 10 or more patients previously described in the literature, while for the genes with less patient descriptions in the literature this percentage was as low as 33% (figure 5).

Exomiser run in re-analyzed cases

Exomiser was also run in all the 64 re-analysed cases in order to assess if there was any good candidate gene missed in the manual analysis. After manually revising the three best Exomiser hits we observed that Exomiser was able to pick two compound heterozygous in the *SLC6A9* gene in a patient classified in the ID panel and with an overlapping phenotype to the described for mutations in this genes, the co-occurrence of which had been missed in the manual analysis. Patient P17 is a six years old female with an encephalopathy accompanied by episodes of nonketotic hyperglycinemia and microcephaly. She also had neonatal hypotonia and currently presents motor delay, delayed speech and language development, and episodic tachypnea. She has anotia with severe conductive hearing impairment and hypertelorism.

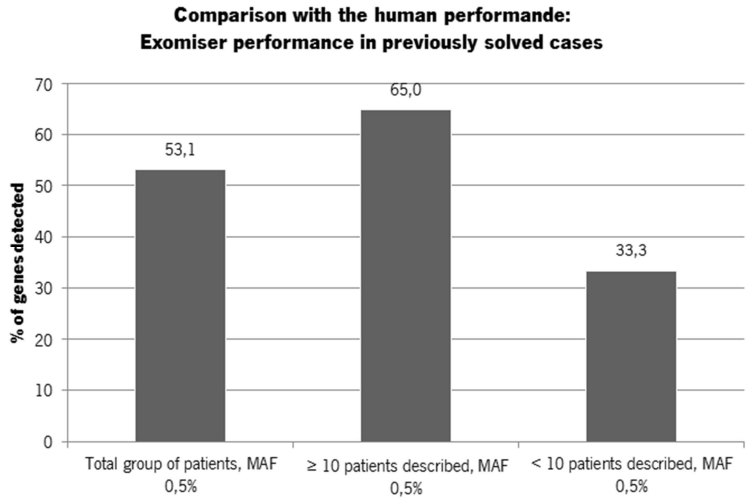


Figure 43 - Exomiser performance is previously "solved" cases. Percentage of genes that Exomiser was able to correctly identify among the genes though to be causing the disease in 33 cases solved in 2013/2014. The column to the left shows the overall percentage of success; both columns to the right show the percentage of success in the cases for which there were < than 10 or ≥ than 10 patients described in the literature bearing mutations in the gene.

The patient carries two compound heterozygous variants (c.A1750T; p.I584F and c.T326G; p.F109C) in the Solute Carrier Family 6 (Neurotransmitter Transporter, Glycine), Member 9 (*SLC6A9*) gene (MIM 601019, RefSeq accession number NM_201649), each of the parents carrying one of the variants. *SLC6A9* encodes a calcium-dependent transporter of the amino acid glycine (an inhibitory neurotransmitter in the central nervous system), also called *GLYT1* (Betz *et al.*, 2006). Glycine is an inhibitory neurotransmitter and a co-agonist at the N-methyl-D-aspartate (NMDA) receptor at synapse where it plays an important role in motor function, and its reuptake into nerve terminals and glial cells is done by GlyT-1 and GlyT-2 (Eulenburg *et al.*, 2005), that regulate glycine concentrations in CNS, playing an crucial role in inhibitory *vs* excitatory neurotransmission balance (Cioffi and Guzzo, 2016, p. 1). Until 2016 only mutations in components of the glycine cleavage system had been associated with glycine encephalopathy, also known as nonketotic hyperglycinemia (NKH), when recessive mutations in *SLCA69* were described in patients with NKH (Applegarth and Toone, 2006; Alfadhel *et al.*, 2016; Kurolap *et al.*, 2016). NKH patients usually present with ID, respiratory alterations, seizures, hypotonia and microcephaly, a phenotype overlapping to that of our patient. Even though none of the variants here identified is yet described in other patients (ClinVar database) both are predicted to be pathogenic and very likely account for the disease in the patient.

Discussion

In this study we re-analyzed WES data from cases previously studied and left unsolved, in the same clinical context, 2-3 years after the first assessment/analysis. The enrolled patients belong to a clinically heterogeneous cohort and have been extensively studied, but still a 34% pathogenic/likely pathogenic variant identification yield was achieved in this re-analysis, a very satisfactory result since many of these patients were at the end of the diagnostic odyssey. We decide to combine a whole exome search approach with a previous virtual panel filtering step. Although the use of a panel based amplification approach would increase the mean coverage and reduce the costs of the analysis, it would only solve cases which had the causative gene included in the selected panel. In our population this would not be effective since we were dealing with a clinically heterogeneous population and many times were confronted with the possibility of more than one panel being possible to apply to a patient. Also, there is always room for fallibility in the selection of the gene panel to sequence to each case, which can be due to clinical heterogeneity and atypical presentations in patients with already known syndromes (Sawyer *et al.*, 2016). Furthermore, new genes keep being identified which means panels quickly become obsolete. In the light of these facts, the use of a virtual panel filtering to WES data as a first step proved to be a better solution. The panels used for filtering are frequently updated with new disease related genes and, as time passes, they became more complete. Also, whenever a panel failed the identification of a good candidate gene we easily moved outside the panel and looked at the rest of the genes (first looking at genes with an OMIM entrance and finally the ones without an entrance in OMIM). In a recent study by the Saudi Mendeliome Group the authors reported that a panel based approach retrieves higher yields in disorders with a broader phenotypic spectrum when compared with WES (Saudi Mendeliome Group, 2015). We believe that this might be mainly due to the fact that panels for broader phenotypes contain a much higher number of genes, making them more "similar" to as of a whole-exome approach. Another important aspect explaining the high yield of this study (43%) is the fact that the studied population is enriched in consanguineous families. However, considering only the autosomal variants the yield of the study was 24%, a very similar percentage to WES studies in outbred populations (Saudi Mendeliome Group, 2015). For the 22 patients with a variant discovered in our re-analysis, half of the variants were found in the panels and the other half outside. Even though the panels are updated regularly, the speed at which new disease associated genes are discovered is just too high and not feasible to accompany at the same rate by the laboratories. Illustrating this, the 11 variants

found outside the panel in the 2016 re-analysis were located in genes which were still not integrated in the applied panel (which is a situation similar to what happen in the study of the Saudi Mendeliome group) (Saudi Mendeliome Group, 2015).

With this work we can conclude that an unsolved case is classified as so at a specific moment in time and that this status might be only temporary for many patients, due to interpretative limitations or knowledge base limitations at the time of analysis. The question of *timing* of analysis and the available literature information at the moment is essential for diagnosis. In fact, we have witnessed its impact in more than one patient during the course of this work. Good examples of this situation are patients P5 (*MYOD1* homozygous nonsense variant) and for patient P4 (*DNAJC21* homozygous missense variant) for which the re-analysis was performed just weeks after the publication of the first reports of mutations in the *DNAJC21* gene in humans. The same situation occurred for patients P13 and P12. Patient P13 revealed a variant in *VARS* (homozygous missense), a gene only associated with human disease in November 2015, when it was reported in three patients with severe microcephaly, DD/ID and cortical atrophy (Karaca *et al.*, 2015). A similar situation happened for patient P7. Even though exome studies were initiated before, it was only after he passed away that *NBAS* was reported mutated in patients with autosomal recessive hepatic encephalopathy due to liver failure (Capo-Chichi *et al.*, 2015; Staufner *et al.*, 2016), a highly similar phenotype to the one he presented. On a positive note, the parents currently know that they are heterozygous carriers and have now the possibility of specific prenatal diagnosis. Another factor contributing for missing a candidate variant is the phenotypic variability present in several syndromes. This was the case for patient P12 whose initial presentation was of ID, encephalopathy, choreoathetosis, macrocephaly with hydrocephalus and automutilation features (which lead to the initial suspicion of Lesch-Nyhan syndrome). In this patient, WES revealed a *de novo SMARCB1* variant previously described in patients with ID, hydrocephalus and myopia (also present in our patient) (Kleefstra *et al.*, 2012). *SMARCB1* is one of the genes in which *de novo* mutations cause Coffin-Siris syndrome but only 50% of the patients carrying *SMARCB1* mutations present self-injury behavior (Santen *et al.*, 2013). The auto-mutilation behavior considered in the beginning as a main component of his phenotype is, therefore, a variably expressed feature of the disease. These examples demonstrate that re-analysis of unsolved exomes should be integrated into clinical WES programs, although the optimal frequency of doing such re-analysis is yet to be determined.

Although systematic novel gene discovery search was not the main purpose of this analysis, we still integrated it in the pipeline. In one consanguineous family, among all rare homozygous variants there was a LoF variant in *PAX7* gene. Even though there are (that we know of) no other patients with *PAX7* mutations, based on the function of the gene and the clinical similarities of the patients with the mouse KO model we believe that *PAX7* is a very strong candidate for the progressive skeletal atrophy and increasingly deteriorating condition of patient P6.

When looking in detail into the year in which the candidate gene was first reported as mutated we found that only 13 genes were discovered between the years of 2013 and 2016 (three in 2013, two in 2014, four in 2015 and four in 2016). This leaves us with nine genes that were possibly known to cause disease before the exome analysis was first done in the patients. Nevertheless, is still legitimate that these nine genes have been missed due to timing issues within the year (as mentioned above), lack of enough interpretative literature at the time of analysis and even human failure in the interpretation. This constitutes another piece of evidence to state that re-analysis should be routinely implemented in the WES diagnostic context.

Certainly, in many laboratories the most limiting factor in the implementation of periodic re-analysis of unsolved cases is the allocation of resources for the task. In order to try to find a possible solution for this problem, we decided to test the Exomiser bioinformatics tool to see if this could be efficiently used in a clinical setting. However, the results were not very promising since after optimization and implementation in the lab it was only able to correctly detect up to 65% of the previously solved cases (and this is the best case scenario). One contributor for Exomiser´s poor performance in our patients is the quality of their phenotypic description. Exomiser performs better if five to ten HPO terms are feed to the algorithm, and if the selected terms are not similar to the ones described in OMIM gene entrance (which can be a result of phenotypic variability) the interpretation power of the software is limited. Additionally, it is unlikely that Exomiser would be able to pick recently discovered genes that still don't have an OMIM disease track. Despite all these less positive aspects, we ran Exomiser in the 42 left unsolved cases after re-analysis in order to check if something was missed. Much to our surprise, a compound heterozygous variant in *SLC6A9* for patient P17 was found by Exomiser, which after phenotypic correlation and Sanger validation, we believe was in fact the cause of the disease in the patient. Upon re-evaluation of the manual analysis done previously we concluded that mutations in *SLC6A9* gene were first reported in patients in August 2016 and the final report for the patient was done in June that year. Is important to mention that this patient was first

analyzed in 2013 and re-analyzed in 2016 and still the diagnosis was missed by two months, highlighting again the importance of re-visiting older exome data. Even though we consider this missed case mainly a "timing" issue, our approach also had a "research" analysis component. Because of this, even in June 2016 this gene should have been highlighted as a good candidate since components of the cleavage system of this transporter were reported mutated in patients with a similar phenotype (Kure *et al.*, 2006). This error raises awareness for the possibility of human error always associated with WES data interpretation making Exomiser still a useful tool to run in the unsolved cases of a laboratory.

An important limitation of this work is the time consuming task of analysing only index exomes. Even though trio sequencing adds to the cost of initial sequencing, in the absence of parental exomes a greater number of candidate variants must be re-validated using traditional PCR-based Sanger sequencing, which is also costly. Moreover, parental exome data allows more rapid filtering of proband exome variant sets, reducing the time required for manual review of candidate variants for validation. Thus the relative cost-differential for routine parental WES requires a complex calculation, and will be increasingly favored as WES costs continue to decrease, whereas labor and Sanger sequencing costs are by now relatively fixed. The utility of parental exomes is based on the fact that most of our probands have unaffected parents, with a family history consistent with either recessive or dominant de novo modes of transmission of high penetrance variants. We can´t disregard the fact that in the current study a significant portion of the studied patients come from consanguineous parents (13 out of 64 patients; equivalent to 20%) but some of these variants may be homozygous in either parent as well, de-prioritizing them as candidate causal mutations.

Adding to the limitations of sequencing just the index lays the fact that, in the clinical context, the collection of samples from parents is often only done when something is found in the patient. Because only after the WES analysis (which is a time consuming task itself) the parents are called back and their sample is collected this adds even more time to the data retrieval. Additionally, we must mention that for many cases the return of the parents back to the clinic is difficult, takes a long time and is often not possible. This happened in many of the re-analysed cases, leaving us with many variants of uncertain significance.

On a more positive note, we concluded in this work that a way to tune down the negative effect of this approach (very time consuming) is the use of virtual panels in the initial step of analysis. This

significantly reduces the number of variants for analysis and, whenever a panel retrieves a positive detection, the time of analysis for each patient is reduced.

References

Agha Z, Iqbal Z, Azam M, Ayub H, Vissers LELM, Gilissen C, et al. Exome sequencing identifies three novel candidate genes implicated in intellectual disability. PloS One 2014; 9: e112687.

Alfadhel M, Nashabat M, Qahtani HA, Alfares A, Mutairi FA, Shaalan HA, et al. Mutation in SLC6A9 encoding a glycine transporter causes a novel form of non-ketotic hyperglycinemia in humans. Hum. Genet. 2016; 135: 1263–1268.

Al-Jassar C, Andreeva A, Barnabas DD, McLaughlin SH, Johnson CM, Yu M, et al. The Ciliopathy-Associated Cep104 Protein Interacts with Tubulin and Nek1 Kinase. Struct. Lond. Engl. 1993 2017; 25: 146–156.

Amendola LM, Jarvik GP, Leo MC, McLaughlin HM, Akkari Y, Amaral MD, et al. Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. Am. J. Hum. Genet. 2016; 98: 1067–1076.

Anastasaki C, Longman D, Capper A, Patton EE, Cáceres JF. Dhx34 and Nbas function in the NMD pathway and are required for embryonic development in zebrafish. Nucleic Acids Res. 2011; 39: 3686–3694.

Aoki T, Ichimura S, Itoh A, Kuramoto M, Shinkawa T, Isobe T, et al. Identification of the neuroblastoma-amplified gene product as a component of the syntaxin 18 complex implicated in Golgi-to-endoplasmic reticulum retrograde transport. Mol. Biol. Cell 2009; 20: 2639–2649.

Applegarth DA, Toone JR. Glycine encephalopathy (nonketotic hyperglycinemia): comments and speculations. Am. J. Med. Genet. A. 2006; 140: 186–188.

Bao R, Huang L, Andrade J, Tan W, Kibbe WA, Jiang H, et al. Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing. Cancer Inform. 2014; 13: 67–82.

Betz H, Gomeza J, Armsen W, Scholze P, Eulenburg V. Glycine transporters: essential regulators of synaptic transmission. Biochem. Soc. Trans. 2006; 34: 55–58.

Beunders G, van de Kamp J, Vasudevan P, Morton J, Smets K, Kleefstra T, et al. A detailed clinical analysis of 13 patients with AUTS2 syndrome further delineates the phenotypic spectrum and underscores the behavioural phenotype. J. Med. Genet. 2016; 53: 523–532.

Bilgüvar K, Oztürk AK, Louvi A, Kwan KY, Choi M, Tatli B, et al. Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. Nature 2010; 467: 207–210.

Bonafe L, Cormier-Daire V, Hall C, Lachman R, Mortier G, Mundlos S, et al. Nosology and classification of genetic skeletal disorders: 2015 revision. Am. J. Med. Genet. A. 2015; 167A: 2869–2892.

Buchman JJ, Durak O, Tsai L-H. ASPM regulates Wnt signaling pathway activity in the developing brain. Genes Dev. 2011; 25: 1909–1914.

Capo-Chichi J-M, Mehawej C, Delague V, Caillaud C, Khneisser I, Hamdan FF, et al. Neuroblastoma Amplified Sequence (NBAS) mutation in recurrent acute liver failure: Confirmatory report in a sibship with very early onset, osteoporosis and developmental delay. Eur. J. Med. Genet. 2015; 58: 637–641.

Chen J-F, Zhang Y, Wilde J, Hansen KC, Lai F, Niswander L. Microcephaly disease gene Wdr62 regulates mitotic progression of embryonic neural stem cells and brain size. Nat. Commun. 2014; 5: 3885.

Cioffi CL, Guzzo PR. Inhibitors of Glycine Transporter-1: Potential Therapeutics for the Treatment of CNS Disorders. Curr. Top. Med. Chem. 2016; 16: 3404–3437.

Cushion TD, Dobyns WB, Mullins JGL, Stoodley N, Chung S-K, Fry AE, et al. Overlapping cortical malformations and mutations in TUBB2B and TUBA1A. Brain J. Neurol. 2013; 136: 536–548.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 2011; 43: 491–498.

Dhanraj S, Matveev A, Li H, Lauhasurayotin S, Jardine L, Cada M, et al. Biallelic mutations in DNAJC21 cause Shwachman-Diamond syndrome. Blood 2017

van Dijk T, van Ruissen F, Jaeger B, Rodenburg RJ, Tamminga S, van Maarle M, et al. RARS2 Mutations: Is Pontocerebellar Hypoplasia Type 6 a Mitochondrial Encephalopathy? JIMD Rep. 2016

Diodato D, Melchionda L, Haack TB, Dallabona C, Baruffini E, Donnini C, et al. VARS2 and TARS2 mutations in patients with mitochondrial encephalomyopathies. Hum. Mutat. 2014; 35: 983–989.

Edvardson S, Shaag A, Kolesnikova O, Gomori JM, Tarassov I, Einbinder T, et al. Deleterious mutation in the mitochondrial arginyl-transfer RNA synthetase gene is associated with pontocerebellar hypoplasia. Am. J. Hum. Genet. 2007; 81: 857–862.

Eulenburg V, Armsen W, Betz H, Gomeza J. Glycine transporters: essential regulators of neurotransmission. Trends Biochem. Sci. 2005; 30: 325–333.

Farag HG, Froehler S, Oexle K, Ravindran E, Schindler D, Staab T, et al. Abnormal centrosome and spindle morphology in a patient with autosomal recessive primary microcephaly type 2 due to compound heterozygous WDR62 gene mutation. Orphanet J. Rare Dis. 2013; 8: 178.

Fiorillo C, Astrea G, Savarese M, Cassandrini D, Brisca G, Trucco F, et al. MYH7-related myopathies: clinical, histopathological and imaging findings in a cohort of Italian patients. Orphanet J. Rare Dis. 2016; 11: 91.

Fountain MD, Aten E, Cho MT, Juusola J, Walkiewicz MA, Ray JW, et al. The phenotypic spectrum of Schaaf-Yang syndrome: 18 new affected individuals from 14 families. Genet. Med. Off. J. Am. Coll. Med. Genet. 2017; 19: 45–52.

Gai X, Ghezzi D, Johnson MA, Biagosch CA, Shamseldin HE, Haack TB, et al. Mutations in FBXL4, encoding a mitochondrial protein, cause early-onset mitochondrial encephalomyopathy. Am. J. Hum. Genet. 2013; 93: 482–495.

Guerrini R, Mei D, Cordelli DM, Pucatti D, Franzoni E, Parrini E. Symmetric polymicrogyria and pachygyria associated with TUBB2B gene mutations. Eur. J. Hum. Genet. EJHG 2012; 20: 995–998.

Haack TB, Staufner C, Köpke MG, Straub BK, Kölker S, Thiel C, et al. Biallelic Mutations in NBAS Cause Recurrent Acute Liver Failure with Onset in Infancy. Am. J. Hum. Genet. 2015; 97: 163–169.

Haendel MA, Vasilevsky N, Brush M, Hochheiser HS, Jacobsen J, Oellrich A, et al. Disease insights through cross-species phenotype comparisons. Mamm. Genome Off. J. Int. Mamm. Genome Soc. 2015; 26: 548–555.

Hamdan FF, Srour M, Capo-Chichi J-M, Daoud H, Nassif C, Patry L, et al. De novo mutations in moderate or severe intellectual disability. PLoS Genet. 2014; 10: e1004772.

Hori I, Miya F, Ohashi K, Negishi Y, Hattori A, Ando N, et al. Novel splicing mutation in the ASXL3 gene causing Bainbridge-Ropers syndrome. Am. J. Med. Genet. A. 2016; 170: 1863–1867.

Hori K, Nagai T, Shan W, Sakamoto A, Taya S, Hashimoto R, et al. Cytoskeletal regulation by AUTS2 in neuronal migration and neuritogenesis. Cell Rep. 2014; 9: 2166–2179.

Huemer M, Karall D, Schossig A, Abdenur JE, Al Jasmi F, Biagosch C, et al. Clinical, morphological, biochemical, imaging and outcome parameters in 21 individuals with mitochondrial maintenance defect related to FBXL4 mutations. J. Inherit. Metab. Dis. 2015; 38: 905–914.

Inanlou MR, Dhillon GS, Belliveau AC, Reid GAM, Ying C, Rudnicki MA, et al. A significant reduction of the diaphragm in mdx:MyoD-/-(9th) embryos suggests a role for MyoD in the diaphragm development. Dev. Biol. 2003; 261: 324–336.

Inanlou M-R, Kablar B. Abnormal development of the diaphragm in mdx:MyoD-/-(9th) embryos leads to pulmonary hypoplasia. Int. J. Dev. Biol. 2003; 47: 363–371.

Jayaraman D, Kodani A, Gonzalez DM, Mancias JD, Mochida GH, Vagnoni C, et al. Microcephaly Proteins Wdr62 and Aspm Define a Mother Centriole Complex Regulating Centriole Biogenesis, Apical Complex, and Cell Fate. Neuron 2016; 92: 813–828.

Jiang P, Jin X, Peng Y, Wang M, Liu H, Liu X, et al. The exome sequencing identified the mutation in YARS2 encoding the mitochondrial tyrosyl-tRNA synthetase as a nuclear modifier for the phenotypic manifestation of Leber's hereditary optic neuropathy-associated mitochondrial DNA mutation. Hum. Mol. Genet. 2016; 25: 584–596.

Jones S. An overview of the basic helix-loop-helix proteins. Genome Biol. 2004; 5: 226.

Kablar B, Krastel K, Ying C, Asakura A, Tapscott SJ, Rudnicki MA. MyoD and Myf-5 differentially regulate the development of limb versus trunk skeletal muscle. Dev. Camb. Engl. 1997; 124: 4729–4738.

Karaca E, Harel T, Pehlivan D, Jhangiani SN, Gambin T, Coban Akdemir Z, et al. Genes that Affect Brain Structure and Function Identified by Rare Variant Analyses of Mendelian Neurologic Disease. Neuron 2015; 88: 499–513.

Kleefstra T, Kramer JM, Neveling K, Willemsen MH, Koemans TS, Vissers LELM, et al. Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. Am. J. Hum. Genet. 2012; 91: 73–82.

Kortüm F, Marquardt I, Alawi M, Korenke GC, Spranger S, Meinecke P, et al. Acute Liver Failure Meets SOPH Syndrome: A Case Report on an Intermediate Phenotype. Pediatrics 2017; 139

Kuechler A, Czeschik JC, Graf E, Grasshoff U, Hüffmeier U, Busa T, et al. Bainbridge-Ropers syndrome caused by loss-of-function variants in ASXL3: a recognizable condition. Eur. J. Hum. Genet. EJHG 2017; 25: 183–191.

Kurata H, Terashima H, Nakashima M, Okazaki T, Matsumura W, Ohno K, et al. Characterization of SPATA5-related encephalopathy in early childhood. Clin. Genet. 2016; 90: 437–444.

Kure S, Kato K, Dinopoulos A, Gail C, DeGrauw TJ, Christodoulou J, et al. Comprehensive mutation analysis of GLDC, AMT, and GCSH in nonketotic hyperglycinemia. Hum. Mutat. 2006; 27: 343–352.

Kurolap A, Armbruster A, Hershkovitz T, Hauf K, Mory A, Paperna T, et al. Loss of Glycine Transporter 1 Causes a Subtype of Glycine Encephalopathy with Arthrogryposis and Mildly Elevated Cerebrospinal Fluid Glycine. Am. J. Hum. Genet. 2016; 99: 1172–1180.

Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, et al. Clinical exome sequencing for genetic identification of rare Mendelian disorders. JAMA 2014; 312: 1880–1887.

Liu Y, Black J, Kisiel N, Kulesz-Martin MF. SPAF, a new AAA-protein specific to early spermatogenesis and malignant conversion. Oncogene 2000; 19: 1579–1588.

Lühl S, Bode H, Schlötzer W, Bartsakoulia M, Horvath R, Abicht A, et al. Novel homozygous RARS2 mutation in two siblings without pontocerebellar hypoplasia - further expansion of the phenotypic spectrum. Orphanet J. Rare Dis. 2016; 11: 140.

Luo X, Zou Y, Tan B, Zhang Y, Guo J, Zeng L, et al. Novel GATAD2B loss-of-function mutations cause intellectual disability in two unrelated cases. J. Hum. Genet. 2017

Maksimova N, Hara K, Nikolaeva I, Chun-Feng T, Usui T, Takagi M, et al. Neuroblastoma amplified sequence gene is associated with a novel short stature syndrome characterised by optic nerve atrophy and Pelger-Huët anomaly. J. Med. Genet. 2010; 47: 538–548.

Meienberg J, Bruggmann R, Oexle K, Matyas G. Clinical sequencing: is WGS the better WES? Hum. Genet. 2016; 135: 359–362.

Mejlachowicz D, Nolent F, Maluenda J, Ranjatoelina-Randrianaivo H, Giuliano F, Gut I, et al. Truncating Mutations of MAGEL2, a Gene within the Prader-Willi Locus, Are Responsible for Severe Arthrogryposis. Am. J. Hum. Genet. 2015; 97: 616–620.

Menke LA, van Belzen MJ, Alders M, Cristofoli F, DDD Study, Ehmke N, et al. CREBBP mutations in individuals without Rubinstein-Taybi syndrome phenotype. Am. J. Med. Genet. A. 2016; 170: 2681–2693.

Metzger JM, Rudnicki MA, Westfall MV. Altered Ca2+ sensitivity of tension in single skeletal muscle fibres from MyoD gene-inactivated mice. J. Physiol. 1995; 485 (Pt 2): 447–453.

Meyer E, Carss KJ, Rankin J, Nichols JME, Grozeva D, Joseph AP, et al. Mutations in the histone methyltransferase gene KMT2B cause complex early-onset dystonia. Nat. Genet. 2016

Nicholas AK, Swanson EA, Cox JJ, Karbani G, Malik S, Springell K, et al. The molecular landscape of ASPM mutations in primary microcephaly. J. Med. Genet. 2009; 46: 249–253.

Nishida T, Mihara T, Horiki T, Ka K. Anaesthesia and orphan disease: primary autosomal recessive microcephaly-10 caused by a mutation in the ZNF335 gene. Eur. J. Anaesthesiol. 2016; 33: 543–545.

Olguín HC, Pisconti A. Marking the tempo for myogenesis: Pax7 and the regulation of muscle stem cell fate decisions. J. Cell. Mol. Med. 2012; 16: 1013–1025.

Oustanina S, Hause G, Braun T. Pax7 directs postnatal renewal and propagation of myogenic satellite cells but not their specification. EMBO J. 2004; 23: 3430–3439.

Passemard S, Titomanlio L, Elmaleh M, Afenjar A, Alessandri J-L, Andria G, et al. Expanding the clinical and neuroradiologic phenotype of primary microcephaly due to ASPM mutations. Neurology 2009; 73: 962–969.

Richards CS, Bale S, Bellissimo DB, Das S, Grody WW, Hegde MR, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. Genet. Med. Off. J. Am. Coll. Med. Genet. 2008; 10: 294–300.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. Off. J. Am. Coll. Med. Genet. 2015; 17: 405–424.

Robinson PN, Köhler S, Oellrich A, Sanger Mouse Genetics Project, Wang K, Mungall CJ, et al. Improved exome prioritization of disease genes through cross-species phenotype comparison. Genome Res. 2014; 24: 340–348.

Romaniello R, Arrigoni F, Cavallini A, Tenderini E, Baschirotto C, Triulzi F, et al. Brain malformations and mutations in - and -tubulin genes: a review of the literature and description of two new cases. Dev. Med. Child Neurol. 2014; 56: 354–360.

Rudnicki MA, Schnegelsberg PN, Stead RH, Braun T, Arnold HH, Jaenisch R. MyoD or Myf-5 is required for the formation of skeletal muscle. Cell 1993; 75: 1351–1359.

Saadi A, Borck G, Boddaert N, Chekkour MC, Imessaoudene B, Munnich A, et al. Compound heterozygous ASPM mutations associated with microcephaly and simplified cortical gyration in a consanguineous Algerian family. Eur. J. Med. Genet. 2009; 52: 180–184.

Sabourin LA, Rudnicki MA. The molecular regulation of myogenesis. Clin. Genet. 2000; 57: 16–25.

Santen GWE, Aten E, Vulto-van Silfhout AT, Pottinger C, van Bon BWM, van Minderhout IJHM, et al. Coffin-Siris syndrome and the BAF complex: genotype-phenotype study in 63 patients. Hum. Mutat. 2013; 34: 1519–1528.

Sato R, Takanashi J-I, Tsuyusaki Y, Kato M, Saitsu H, Matsumoto N, et al. Association Between Invisible Basal Ganglia and ZNF335 Mutations: A Case Report. Pediatrics 2016; 138

Saudi Mendeliome Group. Comprehensive gene panels provide advantages over clinical exome sequencing for Mendelian diseases. Genome Biol. 2015; 16: 134.

Sawyer SL, Hartley T, Dyment DA, Beaulieu CL, Schwartzentruber J, Smith A, et al. Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. Clin. Genet. 2016; 89: 275–284.

Seale P, Sabourin LA, Girgis-Gabardo A, Mansouri A, Gruss P, Rudnicki MA. Pax7 is required for the specification of myogenic satellite cells. Cell 2000; 102: 777–786.

Segarra NG, Ballhausen D, Crawford H, Perreau M, Campos-Xavier B, van Spaendonck-Zwarts K, et al. NBAS mutations cause a multisystem disorder involving bone, connective tissue, liver, immune system, and retina. Am. J. Med. Genet. A. 2015; 167A: 2902–2912.

Simons C, Griffin LB, Helman G, Golas G, Pizzino A, Bloom M, et al. Loss-of-function alanyl-tRNA synthetase mutations cause an autosomal-recessive early-onset epileptic encephalopathy with persistent myelination defect. Am. J. Hum. Genet. 2015; 96: 675–681.

Smedley D, Jacobsen JOB, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat. Protoc. 2015; 10: 2004–2015.

Srivastava A, Ritesh KC, Tsan Y-C, Liao R, Su F, Cao X, et al. De novo dominant ASXL3 mutations alter H2A deubiquitination and transcription in Bainbridge-Ropers syndrome. Hum. Mol. Genet. 2016; 25: 597–608.

Srour M, Hamdan FF, McKnight D, Davis E, Mandel H, Schwartzentruber J, et al. Joubert Syndrome in French Canadians and Identification of Mutations in CEP104. Am. J. Hum. Genet. 2015; 97: 744–753.

Staib JL, Swoap SJ, Powers SK. Diaphragm contractile dysfunction in MyoD gene-inactivated mice. Am. J. Physiol. Regul. Integr. Comp. Physiol. 2002; 283: R583-590.

Staufner C, Haack TB, Köpke MG, Straub BK, Kölker S, Thiel C, et al. Recurrent acute liver failure due to NBAS deficiency: phenotypic spectrum, disease mechanisms, and therapeutic concepts. J. Inherit. Metab. Dis. 2016; 39: 3–16.

Sun Y, Ruivenkamp CAL, Hoffer MJV, Vrijenhoek T, Kriek M, van Asperen CJ, et al. Next-generation diagnostics: gene panel, exome, or whole genome? Hum. Mutat. 2015; 36: 648–655.

Tajsharghi H, Thornell L-E, Lindberg C, Lindvall B, Henriksson K-G, Oldfors A. Myosin storage myopathy associated with a heterozygous missense mutation in MYH7. Ann. Neurol. 2003; 54: 494–500.

Tanaka AJ, Cho MT, Millan F, Juusola J, Retterer K, Joshi C, et al. Mutations in SPATA5 Are Associated with Microcephaly, Intellectual Disability, Seizures, and Hearing Loss. Am. J. Hum. Genet. 2015; 97: 457–464.

Taylor RW, Pyle A, Griffin H, Blakely EL, Duff J, He L, et al. Use of whole-exome sequencing to determine the genetic basis of multiple mitochondrial respiratory chain complex deficiencies. JAMA 2014; 312: 68–77.

Tummala H, Walne AJ, Williams M, Bockett N, Collopy L, Cardoso S, et al. DNAJC21 Mutations Link a Cancer-Prone Bone Marrow Failure Syndrome to Corruption in 60S Ribosome Subunit Maturation. Am. J. Hum. Genet. 2016; 99: 115–124.

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38: e164.

Watson CM, Crinnion LA, Murphy H, Newbould M, Harrison SM, Lascelles C, et al. Deficiency of the myogenic factor MyoD causes a perinatally lethal fetal akinesia. J. Med. Genet. 2016

Willemsen MH, Nijhof B, Fenckova M, Nillesen WM, Bongers EMHF, Castells-Nobau A, et al. GATAD2B loss-of-function mutations cause a recognisable syndrome with intellectual disability and are associated with learning deficits and synaptic undergrowth in Drosophila. J. Med. Genet. 2013; 50: 507–514.

Wincent J, Luthman A, van Belzen M, van der Lans C, Albert J, Nordgren A, et al. CREBBP and EP300 mutational spectrum and clinical presentations in a cohort of Swedish patients with Rubinstein-Taybi syndrome. Mol. Genet. Genomic Med. 2016; 4: 39–45.

Yablonka-Reuveni Z, Rudnicki MA, Rivera AJ, Primig M, Anderson JE, Natanson P. The transition from proliferation to differentiation is delayed in satellite cells from mice lacking MyoD. Dev. Biol. 1999; 210: 440–455.

Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, et al. Molecular Findings Among Patients Referred for Clinical Whole-Exome Sequencing. JAMA 2014; 312: 1870–1879.

Yang YJ, Baltus AE, Mathew RS, Murphy EA, Evrony GD, Gonzalez DM, et al. Microcephaly gene links trithorax and REST/NRSF to control neural stem cell proliferation and differentiation. Cell 2012; 151: 1097–1112.

Zech M, Boesch S, Maier EM, Borggraefe I, Vill K, Laccone F, et al. Haploinsufficiency of KMT2B, Encoding the Lysine-Specific Histone Methyltransferase 2B, Results in Early-Onset Generalized Dystonia. Am. J. Hum. Genet. 2016; 99: 1377–1387.

CHAPTER 4

General discussion

Discussion

This chapter makes a summary of the main findings, strengths and limitations of this thesis. In this work two different genome wide approaches were used to identify genetic causes underlying neurodevelopmental spectrum diseases: aCGH and MPS.

For the patients and families obtaining a molecular diagnosis for the disease is the most important outcome from this type of studies. The clinical utility of a genetic diagnosis is determined by the impact that it has on outcomes relevant for the patients and families. In the context of NDD we can consider that several outcomes are of relevance, such as the possibility of recurrence risk estimation and its impact of reproductive decisions, the early and appropriate access to education and behavioral therapies and the avoidance of additional consultations, samples collection and molecular tests.

The importance of a molecular diagnosis

Often, the families face the uncertainty of a molecular diagnosis for many years paired with feelings of anger, guilt and misunderstanding of the process. The identification of a genetic cause can help provide relieve and release from feelings of guilt. A comparative study between parents of children with ID with and without an etiological diagnosis revealed that parents felt the need for a molecular diagnosis more intensely when the disease started to show in the child, which goes in line with the initial shock of the disease and eventual feelings of guilt (Makela *et al.*, 2009). Also, the parents reported that validation was a very important aspect of the molecular diagnosis since it offered legitimacy for the child´s behavior (Makela *et al.*, 2009). When a patient has a molecular confirmation, this might allow the families to have a better prognosis for the health of their relative as well as the possibility for recurrence risks calculation allowing genetic counselling, an extremely important service to offer to couples with affected progeny that wish to have more children. The diagnosis can also be seen as a right of the patients and, in the case of intellectually impaired children or adults, their families or caretakers. In fact, the diagnosis can retrieve carrier information which is also important for the family members, and they can claim the right to know in order to make fully informed reproductive and life decisions (Fulda, 2006).

The contribution of genome wide techniques for unraveling genomic basis of NDD

Because NDDs present a high clinical variability and are genetically heterogeneous, whole genome or whole exome techniques present a suitable approach to diagnosis. aCGH is

recommended as the first technique to be applied to patients with ID or MCA without a clinical presentation that justifies the suspicion of specific syndromes (Miller *et al.*, 2010). In **sub-chapter 2.1** we have used aCGH to investigate disease etiology in Portuguese patients with unexplained ID. Overall, we have identified clinically significant CNVs (considering known pathogenic and likely pathogenic variants) in 15.3% of the patients and variants of unknown significance in 14.7% of the patients. Making a stricter analysis and considering only the variants associated (or likely associated) with disease we can consider that this yield is comparable with several other similar studies, in which percentages ranging between 8.5% and 16% were achieved (Rosenberg *et al.*, 2006; Lu *et al.*, 2007; D'Amico and Bertini, 2008; Buysse *et al.*, 2009; Sagoo *et al.*, 2009; Xiang *et al.*, 2010). Other studies have been published revealing higher detection rates, such as Cappuccio and colleagues in 2016, that mentioned the presence of 30% of pathogenic CNVs in patients with ID, and/or ASD and/or MCA, one of the highest yields published so far (Cappuccio *et al.*, 2016). One of the questions that we wanted to address was whether there were any differences between the genomic findings detected by aCGH in two groups of patients collected in different contexts in Portugal. All the patients included in the research cohort (RC) had been extensively studied from the genetic perspective, the most common genomic aberrations (for instance very large CNVs and subtelomeric variants) being therefore excluded. This is reflected in the lower yield for known and likely pathogenic variants in this cohort (8.5% and 5.3%, respectively obtained in the RC, compared with 11.7% and 5.8% in the clinical cohort (CC)). The CC included a much more heterogeneous set of patients (and of aCGH platforms applied) and a much more variable panel of techniques performed before the patient entered the aCGH study. One of the most similar studies in the literature is the one from Rodríguez-Revenga and colleagues in 2013, where the authors compared the yields of two different aCGH platforms in a set of Spanish patients with ID. Overall the study had a yield of 15% (which is similar to our results) and the vast majority of the positive findings were detected using a custom array (KaryoArray V3.0), one of the platforms used in the CC (Rodríguez-Revenga *et al.*, 2013). Other similar studies in the literature also reported higher yields (23%) than ours (Männik *et al.*, 2011; Qiao *et al.*, 2014), but we believe that this might be due to the nature of the cohorts in which only for a low number of cases it wasn't possible to study the parents, leading to a lower number of VOUS than that obtained in this work.

The advent of aCGH technologies allowed the delineation of new microdeletion and microduplication syndromes and the detailed clinical description and establishment of the core

phenotype associated with those. Even though many of these syndromes have been identified in many patients and widely characterized from the clinical point of view, the phenotypic presentation of these can still be unspecific and variable, making it difficult to recognize (Slavotinek, 2008). However, as time goes by and the application of aCGH technologies grows wider, the rate at which these syndromes were described slowed down and gradually less distinct microdeletion and microduplication syndromes were published (van Ravenswaaij-Arts and Kleefstra, 2009).

In this work, we contributed with the description of more patients carrying known pathogenic CNVs, namely: deletions - 1p36.23-p36.21, 1q43-q44, Coffin-Siris, Cri-du-Chat, 7q11.23, 8p23.1, 9p13.1-p11.2, Jacobsen, 12q24.21-q24.22, 16p11.2, 17q21.31, 22q11 and 22q13.3, Xp22.33-p22.31; duplications - 9q34.13-q34.3, 12q24.21, 13q12.12-q34, 15q11.2-q13.1, 21q11.2-q22.11 and *MECP2* duplication; as well as risk associated *loci* - 1q21.1 and 16p13.3. Additionally, we also contributed with the identification of novel likely pathogenic CNVs in 18 patients, namely CNVs, namely: deletions – 2q11.2-q12.2, 7q33, 10q26.3, 17p11.2, 20q13.12-q13.13; duplications - 1p22.1-p21.3, 2p15, 2q11.2, 7q33, 9q33.2-q33.3, 12p13.33, Xq24 and Xq26.3. We also contribute with a list of VOUS that correspond to 14.7% of the cases, the clinical significance of which can be addressed by other groups or in future studies.

The problem with very rare genomic variants

When working with techniques that retrieve data at the exome/genome level, we are faced with extremely rare and many times private genomic variants that are of extreme importance to share within the scientific community. In fact, the replication of the finding in other patients with overlapping characteristics is a criteria to establish the pathogenicity of a variant or gene. However, often these variants are difficult to pair with someone else. Because of the diagnostic challenges with the identification of novel private or of unknown significance genomic variants, the use of patient databases, such as Decipher and GeneMatcher (Firth *et al.*, 2009; Sobreira *et al.*, 2015) or even direct researcher contacts is often a good resource to find more patients with overlapping CNVs. Good examples of these situations are described in sub-chapter 2.3 and 2.4 for CNVs in 2p13.2 and 7q33, respectively.

In **sub-chapter 2.3** we described two patients with overlapping deletions in chromosome 2p13.2 (one deletion has 0.78 Mb and the other 4 Mb in size) who share some similar phenotypic features, such as ID with a specific compromise of the language, hyperactivity, ear

malformation and skeletal abnormalities (at the vertebral and craniofacial level). These two patients carry extremely rare deletions and were only paired together thanks to the Decipher database. Even though for the patient with the 4Mb *de novo* deletion it is very likely that the variant was the responsible for the disease, the same was not the case for the patient with the smaller 0.78Mb *de novo* deletion since the hypothesis for this to be a random non disease-causing event was present. Additionally, for the patient with the larger deletion (affecting dozens of genes) it was not clear which one(s) was (were) key for the clinical symptoms. The pairing with another patient with a smaller deletion and overlapping phenotype resulted in the narrowing down of the list to only two genes: *EXOC6B* and *CYP26B1*. This finding allowed the possibility of functional studies at the cellular level and a better understanding the biological consequences of this genomic abnormality (reduced expression of *EXOC6B* at mRNA and protein level) (Wen *et al.*, 2013). The molecular studies comprised important evidence that *EXOC6B* haploinsufficiency might play an important role in the molecular pathogenesis of ID, something until then never reported. In fact, it was only one year after our publication that another case with a 2p13.2 deletion was reported in the literature (Evers *et al.*, 2014). The authors report a mosaic 0.46Mb deletion affecting the same genes (although *EXOC6B* was only partially deleted) in a boy with DD, speech delay and dysmorphic features, reinforcing the possible contribution of the exocyst for Notch signaling and consequently of its dysfunction in ID. This again highlights the importance of sharing information about patients with unique and almost private CNVs in relevant databases and in the literature. We can hypothesize that if the first patient hadn't been introduced in Decipher there would still be considerable doubt about the pathogenicity of his deletion, the publication presented in sub-chapter 2.3 would not have been possible and consequently the same for the report by Evers and colleagues. Similar situations were found when we were analyzing the exome sequencing data.

In **sub-chapter 3.1** a cohort of 19 patients with Rett syndrome-like features was studied by aCGH and exome sequencing (the last one using a trio-analysis approach). Six patient revealed variants in gene previously associated with NDDs and in five new NDD candidate genes. Considering the new candidate genes category, *GABBR2* and *SMARCA1* genes are good examples of the speed at which MPS technologies allow the discovery of new genes. At the time of publication, *GABBR2* had been only described as mutated in two patients with epileptic encephalopathy (EuroEPINOMICS-RES Consortium *et al.*, 2014). While we were preparing the

publication we used the GeneMatcher database (Sobreira *et al.*, 2015) to help us find other researchers with interest in this gene but we didn't get any connection. It was only after the publication, that we were contacted by other laboratories (either via direct email under the subject of the publication or by GeneMatcher) and exchange information about the patients. A similar situation happened for *SMARCA1* gene where two contacts were meanwhile made via GeneMatcher regarding patients with very distinct phenotypes: one with ID and epilepsy and the other with a CNS tumor but without any compromise of the intelligence and mental development (Fadi Hamdan; Bryan Krock, personal communication). Meanwhile, *SMARCA1* variants affecting brain structure and function were also reported by others (Karaca *et al.*, 2015), including in patients with psychosis (Homann *et al.*, 2016) revealing that the spectrum of *SMARCA1* implications in disease might be wider than initially thought. From this work we could observe that, after filtering and validation of the variants in the progenitors, we can still be left with a variant in a gene yet to be associated with disease (or even any known cellular function). In these cases there is always the possibility that these changes will turn out not to be causative of the disease, in which case the discovery of more patients or healthy controls with the variant as well as functional studies could help solve that problem.

Sub-chapter 2.4 describes seven patients (from a total of three independent families) with overlapping CNVs at the 7q33 citoband. All the patients present ID, dysmorphic features and behavioral alterations. To the best of our knowledge, the collection from sub-chapter 2.4 comprises the first group of patients with small and overlapping genomic imbalances in this region reported in the literature (as 7q33 deletions seem to be quite rare) that leads to a possible definition of key genes for the phenotype. The genes commonly affected in all the cases are *CALD1* and *AGBL3*, however *EXOC4* and *CNOT4* (also in the region and deleted in four of the patients) comprise good candidate genes for the phenotype. In this work the Decipher database played an important role in the sharing and discussion with other laboratories that found patients with overlapping CNVs. When using this strategy to conclude about the CNV at 7q33 region one of the limitations that we found was the reduced clinical description often uploaded for each patient. As of January 2017, there were 15 patients in Decipher with relatively small and overlapping CNVs. Of those 15 patients, four had no clinical description and three were only described as having ID. Additionally, for eight patients the inheritance of the variant was not reported. This lack of information made it difficult to compare the cases and use the findings for

interpretation. We believe that this is a challenge faced by many Genetics laboratories. The reality in the diagnostic context is that many times the clinical information complementing the genetic study request is reduced. The most likely scenario is that those patients (listed in Decipher) were also collected in a similar context to the one of our CC (sub-chapter 2.1) and the laboratory simply wasn´t given the information. This might be the motive leading to the lack of data in databases, showing again the importance of complete clinical description accompanying the molecular studies. Equally important is obtaining the information about the clinical status of the progenitors and their DNA samples, which can be extremely helpful in the interpretation of inherited genomic imbalances. However, often the clinician faces difficulties in assessing the clinical status of the parents. This might be due to the lack of resources (often the consultation occurs in the Pediatrics context and the formal evaluation of adults might not be possible) or the social background of the families (many times the clinician has doubts about the mental health of the parents and even if this is biologically or socially driven or both). During the practical work of this thesis we were faced with doubts about the father/mother's health status many times, especially in chapter 2 when evaluating the contribution of CNVs to NDD, and this hampered our ability to conclude about the pathogenicity of several of the variants found. In fact, the ACMG guidelines for interpretation and classification of CNVs mention the crucial importance of a thorough medical evaluation of a carrier progenitor, as this can be associated with incomplete penetrance, variable expressivity, parent of origin imprinted genes, mosaicism or the presence of a second genetic variant (Kearney *et al.*, 2011). Another important point mentioned in the ACMG guidelines is the paternity/maternity confirmation using genetic markers and the fact that this is very seldom feasible in the clinical context. In fact paternity testing is not recommended unless there is a compelling clinical reason to make this assessment and explicit informed consent is obtained. The nonpaternity possibility should be disclaimed in all reports (Kearney *et al.*, 2011).

The phenotype variability within patients and across time

The use of aCGH also lead to a shift in the approach used to group patients sharing clinic and genetic anomalies. The phenotype-first approach was increasingly dropped out throughout the years and replaced with a "reverse dysmorphology" or genotype-first approach in which a common genetic cause is found first for a group of patients and only after the shared clinical phenotype is determined (Slavotinek, 2008). A good example of this exercise is presented in **sub-chapter 2.2** where a small group of patients with CNVs affecting the known syndrome 1q43-

q44 is discussed. The 1q43-q44 microdeletion syndrome was first reported in 1976 by conventional karyotype in a girl with ID, MIC, epilepsy and severe dysmorphic features (Mankinen *et al.*, 1976). With the advent of aCGH technologies several smaller deletions affecting the 1q43-q44 citoband were discovered, allowing a better delineation of the critical genomic region and phenotype associated and making it a good example of the reverse dysmorphology approach allowed by aCGH technology (Ballif *et al.*, 2012). In this chapter we describe a collection of five cases in which one parent carries a quadruplication (that should lead to MAC) while asymptomatic, questioning the contribution of copy number gains to MAC. This is not the first case not in agreement with the genotype-phenotype correlation model to be reported, since two other cases in the literature are described not to have MIC while carrying a disrupting *AKT3* CNV (Ballif *et al.*, 2012). The delineation of the phenotype and the consequent phenotypic variability observed in the region reveals that, as for the known susceptibility *loci* (reviewed by (Torres *et al.*, 2016)), incomplete penetrance and variable expressivity is found many times in other (less frequently affected) genomic regions, and is possible to observe even in smaller cohorts of patients.

Another important aspect to consider is the variability of the phenotype across time. The way that the phenotype associated with deletion/duplication syndromes is often described in the literature and databases highlight an important problematic: the need for longitudinal studies. Often it is possible to observe a collection of patients with a new identified genomic variant and subsequent (cross-sectional) reports adding other patients. However, the re-evaluation of the same patients several years after is not common in the literature. One factor related to this can be the context in which the patient was diagnosed. If it was done in the clinical context is likely that the clinician/laboratory lost contact with the patient in the following years after the diagnosis and scientific report.

Challenges in the interpretations of genetic variants: the uncertain significance

Currently, the labs that routinely use aCGH and MPS face the reality of finding likely pathogenic variants or VOUS that are unique for that patient or family, leading to difficulties in the interpretation of the significance of the variant for the disease. Several aspects should be taken into account when considering the impact of a novel CNV in the patient phenotype:

- The variant: in the case of CNVs if it is a deletion or duplication (as well as the size) and, in the case of SNVs, if it is a loss of function variant (frameshift, splice site altering, stop) or a missense variant;
- The gene content of the variant or the gene affected: gene content is a point that might lead to conclusion bias since many times the genes involved are not yet fully characterized (maybe their function in nervous systems or even in any systems is still unknown) or were never implicated in a human disease before, which might lead to the false interpretation of unlikely pathogenicity;
- If the variant is frequently described in controls, and therefore likely polymorphic, or if the CNV overlaps significantly with a polymorphic region;
- If the variant is described in other affected patients or affected family members;
- If the variant is inherited from a healthy parent: the existence of a CNV inherited from a parent without evident clinical presentation of symptoms might not mean it is benign since a more complex mechanism behind the chromosomal anomaly may be involved (such as the occurrence of another variation in another allele or even in a different chromosomal location can be present affecting the clinical outcome in the patient) (Mencarelli *et al.*, 2008). The hypothesis that a CNV is not fully penetrant in the parents should also be considered as it has been proven to be the case in recurrent CNVs conferring increased risk for neurodevelopmental and psychiatric disorders (reviewed in (Torres *et al.*, 2016)). Additionally, there is also the hypothesis of the symptomatology being associated with a complex inheritance pattern caused by an extra modifier, such as parental imprinting effect, parental mosaicism or variable expressivity, which might bias the interpretation (Klopocki *et al.*, 2007).

Both in chapter 2 and 3, we were faced with the challenge of the definitive interpretation of a genomic variant significance in the child with ID when the parental DNA is not available. Many times, when the one or even both parents are not available, the interpretation of the relevance of the variants found is not possible and the molecular diagnosis is compromised. As mentioned before, the use of genome/exome wide techniques for genetic analysis shifted the clinical evaluation from a "phenotype delineation first" to a "reverse phenotyping" approach after the determination of the genetic cause of disease. When we are faced with a situation of a genomic alteration which is very well described in many patients and with a clear, well proven pathogenic

association with disease, the conclusion of causality in the patient can be easily done. However, often the impact of the detected genetic finding is not very well established and there is the need for clinical re-evaluation of the patient and/or families in order to look for specific symptoms that could previously have been missed. In the specific context of clinical practice, often the families have received the clinical diagnosis for quite a while, and coming back to the consultation for clinical re-evaluation may represent a difficult and time consuming task. Also, the time available for consultation often doesn't allow the collection of an extensive clinical description.

Looking back at the number of variants that we classified as VOUS in sub-chapters 2.1 and 3.2, we can state that many of them would be moved either to the "pathogenic" or "benign" categories if the parents had been available for analysis.

This problematic, in the research context, has implications in the final classification of the variant and consequently in the final yield obtained in the studies and possibility of discovering a new disease associated variant/gene. However, in the clinical context the implications go beyond that: it may cause the inability to provide genetic counseling. If a couple that wants a prenatal test for a variant could in theory be "forced" (or strongly encouraged) to provide their DNA samples and to allow their own clinical assessments, this might not the case if one of the progenitors has a new partner or if a sibling (or another family member) is taking care of the patient.

The importance of re-visiting unsolved cases

The cumulative and fast-evolving nature of the available knowledge on genes and their relationship to human disease is one of the most interesting components of the advent of MPS application to the clinic. Consequently, it was just a matter of time until the need for re-evaluating unsolved cases occurred. In **sub-chapter 3.2** we re-analyzed a cohort of 64 patients for which exome sequencing data had previously been generated and for which that effort resulted in an unsolved situation. Considering the pathogenic, likely pathogenic and uncertain significance variants found at the time of re-analysis jointly we ended out with 34% of the cases with at least one variant of those previously identified possibly explaining the phenotype, but only recognized as such 2-3 years after the first evaluation. The strategy followed started with the application of a virtual gene panel, selected accordingly to the patients' symptoms, then moving outside the panel to the analysis of all genes only when the first approach revealed no interesting results. When comparing this work to a similar one published recently, we can observe that the yield

obtained by others was around 10%, much lower than ours (Wenger *et al.*, 2017). We believe that this is due to two main factors: (I) we described the results in a less stringent manner considering all the variants for the calculations (pathogenic, likely and uncertain) while Wenger and colleagues considered only certainly causative variants; (II) in our cohort we still miss a lot of parents' samples to confirm the inheritance status. If we consider only the variants that according to the ACMG guidelines we can classify as pathogenic we reached a yield of 4.7%, while if we also add the likely pathogenic variants this raises to 18.8%, both more similar numbers as the one from Wenger et al. An important factor possibly creating a bias in our cohort is the fact that a significant portion of the patients (20%) come from consanguineous families, which may increase our yield.

In the context of the NDDs re-evaluation of the same group of patient's years after the first report is generally lacking, comprising a limitation in the field. A good example of this situation occurred in sub-chapter 2.1. The patient carrying a 10q26.3 deletion was first analysed in 2011 and at the time this genomic variant was considered as VOUS. This classification was due to the fact that, even though it occurred *de novo* and the entire region was not covered by deletions in the controls, a portion of the *EBF3* was deleted in three controls. Even though at the time we thought that the *EBF3* gene could have an important function role in the nervous system, there were no reports of patients with *EBF3* CNVs or SNVs. It was only in 2016 that were two publications came out reporting *EBF3* point mutations in patients with ID. For this reason, and in the light of recently disclosed data, the classification of the variant could be altered for likely pathogenic. The new data in the literature lead to a clinical re-evaluation of the patient five years after the initial analysis, making it possible to observe the development of a patient with *EBF3* haploinsufficiency in that period. Longitudinal studies should be performed for the syndromes described in the last decades with the advent of genome-wide diagnostic techniques. For this to be possible, the patients should be followed in the same genetic center and reevaluated at each new consultation. However, in reality this good practice might not be possible for many genetic outpatient clinics. Unfortunately, often the families lose contact with the genetic department as their child grows older and recurrent consultations are not always possible for the families. Also, the time available for each geneticist consultation might be reduced for a detailed patient re-evaluation.

Functional studies in the discovery of new ID-associated genes

When a new gene is found as possibly associated with ID, the real impact of the variant in the gene needs to be determined by functional studies. However, at the pace that new disease associated genes are discovered is (more than 700 new genes until 2015) is impossible to keep pace with functional assays for all of them and even more impossible to establish *in vitro* or *in vivo* models (Vissers *et al.*, 2016).

Because of the nature of the key tissue affected in ID (neuronal) the use of specific cellular assays to test the variant impact is more difficult. One promising tool to overcome this limitation is the use of induced pluripotent stem cells (iPSCs). New reprogramming approaches allow the transformation of an adult cell into a pluripotent embryonic stem cell-like condition by forcing the expression of four specific transcription factors (Shi *et al.*, 2017). Afterwards, the cells can be differentiated in the desired cell type, which in the ID context would be neurons. With the establishment of iPSCs for patients carrying a specific variant, their individual genetic background is still present and the impact of the variant in the neurons can be studied (Mora *et al.*, 2017). One example of this was the iPCSc model for Williams-Beuren Syndrome (WBS) that determined significant differences in electrophysiological defects in WBS patients derived neurons (Khattak *et al.*, 2015).

Outside the *in vitro* context, the impact of a variant can also be studied in *in vivo* though the use of animal models. For many years, the establishment of a mouse model for a given disease was the goal standard to study its pathogenesis. However, this is not feasible in a disease such as ID for which many different genes are currently known. The time and cost of establishment a mouse model is not compatible with the heterogeneity currently recognized for ID. Because of this problem, simpler animal models have been used such as the fruit fly (*Drosophila melanogaster*), zebrafish (*danio rerio*) and *Caenorhabditis elegans* (*C. elegans).* Zebrafish presents itself as a good model to surrogate human pathologies due to its smaller evolutionary distance to humans (when compared with *C. elegans*), the orthology of organs and signaling processes (it has a brain), embryos being transparent and allowing medium and high throughput phenotypic screenings (Davis *et al.*, 2014). This model has been widely used in the study of microcephaly/macrocephaly and ciliopathies (Dzafic *et al.*, 2015; Guemez-Gamboa *et al.*, 2015, p.; Song *et al.*, 2016). Recently it was used to determine the contribution of possible key genes in a new 2p15p16.1 microdeletion syndrome, revealing that knockdown of four of the genes

involved in the 2p15p16.1 region resulted in microcephaly with brain anomalies and alterations in body growth and development (Bagheri *et al.*, 2016). The use of zebrafish to model the haploinsufficiency of these genes allowed to confirm the dosage deficit already observed in human lymphocytes cell line as well as to modulate the impact of these genes deficiency in an *in vivo* model (Bagheri *et al.*, 2016).

The *Drosophila melanogaster* model allows the use of several paradigms to study of its behavior as well as the study of brain organization during development (Bellen *et al.*, 2010). The fruit fly allows the study of functional modules in which ID genes work with sufficient throughput to allow the necessary functional studies for genes identified by MPS diagnostics, identification of reversible ID related phenotypes and for drug discovery efforts (van der Voet *et al.*, 2014). For several years the impact of genes' haploinsufficiency has been modulated in *Drosophila* models. This is the case of *EHTM1*, the Kleefstra syndrome gene, for which the modeling in the drosophila revealed it to be key regulator of cognition (Kramer *et al.*, 2011) as well as for *GATAD2B* which knockdown in neurons leads to alterations in synapse morphology and learning deficits in fly (Willemsen *et al.*, 2013).

Caenorhabditis elegans is considered to be a model positioned between the simplicity of a cell model (it still can be grown in Petri dish) and the complexity of a mouse model (it has a nervous system composed by 302 neurons). This model is also transparent at all ages, which allows the visualization of specific neurons under the microscope or magnifying glass using fluorescent markers (Kinser and Pincus, 2016). Is a inexpensive animal to maintain in laboratory and allows high-throughput screening studies in the nervous system (Teixeira-Castro *et al.*, 2011; reviewed by Bessa *et al.*, 2013). In the context of NDDs, *C. elegans* is often used to dissect the function of a given gene. This was the case for instance of the *PQBP1* gene, in which the study of its conserved homolog *in C. elegans* was crucial to understand the role of *PQBP1* in NDDs (Takahashi *et al.*, 2009).

Limitations of this thesis

In research, we are often pressure to think of the limitations of our work as negative points. However, in this thesis we believe that the definition of the limitations comprises such an important finding as the identification of new disease causative gene, as they help us to pinpoint the lacks in our work and to try to overcome those in future works. Despite the advances in

technology and in the ability to achieve a molecular diagnosis, there is still progress to be made in the integration of all the data currently available and how to put it to good use in the clinics and in the laboratories. Due to MPS, in the last years the scientific community studying ID and ID-related disorders was able to identify more than 700 candidate causative genes (data of 2015) (reviewed by Vissers *et al.*, 2016). This increase was not only observed in ID, as OMIM had (at the end of 2015) 4,570 diseases with a documented gene associated while eleven years before that (in 2004), OMIM listed only 1,636 disorders in the same conditions (Wenger *et al.*, 2017). At this growth rate it is absolutely imperative that previously unsolved cases are revised at the light of new data, something that is already in the awareness of the Genetics community and is possible to observe with the improvement and generation of public databases such as ClinVar, GeneMatcher and ExAC.

The limitations listed in this chapter are an important as any other. The speed at which whole genome/exome technologies allows the discovery of new variants is faster than the ability to manage the information obtained. Is not that this need is not perceived within the scientific community, but instead the pace at which the data is acquired outpaces the ability to update and properly curate databases. Nevertheless, it is important to highlight that in the last years the available databases have significantly improved, and greatly facilitated our work.

Final remarks

In this thesis we used the power of aCGH and MPS techniques for the understanding of the genetic variants (CNVs and SNVs) underlying a clinically and genetically heterogeneous group of disorders. We can foresee that in time MPS will be established as the "one-in-all" solution at diagnostic labs. All the samples would enter the MPS pipeline and only after the raw data is generated specific bioinformatics approaches would be applied according with the desired analysis workflow. For instance, if the patient presents some overlapping clinical features with a certain microdeletion syndrome it would make sense to apply first a bioinformatics pipeline for the detection of CNVs and only after (and if the CNV analysis was normal) move to a pipeline for SNV detection (which could be first applied in a gene panel, secondly in the whole exome and lastly in the whole genome). Whole genome applications of MPS will allow the discovery of aneuploidies, mosaicism, CNVs, inversions, small indels and point mutations (Martin and Warburton, 2015). Several groups have already used WES and WGS in the noninvasive prenatal

setting to detect through the detection of cell-free fetal DNA in maternal peripheral blood (Gregg *et al.*, 2014; Xu and Shi, 2014; Minarik *et al.*, 2015). In this method the fetal DNA is enriched for the chromosome aneuploidy screening analysis, avoiding the need for invasive procedures such as amniocentesis or chorionic villus biopsy. Apart from the use in noninvasive prenatal diagnosis, MPS can also be applied to the detection of copy number variation (CNVs) in the genome since it can present several advantages over aCGH: it allows the detection of very small CNVs, can detect structural anomalies (inversions) and can estimate the breakpoints more precisely. However, this particular application of MPS still requires improvement, since the targeted nature of the technology introduces bias, requiring specialized bioinformatic analysis (Abel and Duncavage, 2013). With the improvement of algorithms and methodology, this soon will be applicable in the scanning for smaller CNVs across the genome, in a similar way as aCGH (Martin and Warburton, 2015).

From the MPS approaches, WES is the most used in Mendelian disorders since it allows the detection of exonic (coding) and splice site variants while requiring only about 2% of the sequencing load that would be required when sequencing the genome (Petersen *et al.*, 2017). Nevertheless, whole exome analysis would be a more complete approach as it allows a genome-wide characterization of, for example, all of the protein binding sites in the genome, methylated loci, mapping of open vs closed chromatin and the detection of positional alterations of genomic fragments (inversions, insertions) (Mardis, 2009).

In 2014 a Dutch group used WGS to study 59 patients with severe ID that previously tested negative for aCGH and WES. The author ended out focusing solely in coding regions (being able to find cause of disease in only 20 patients) and realizing that most of the variants found in the WGS had been missed in the previous WES analysis due to low coverage and technical limitations (Gilissen *et al.*, 2014). The main challenge with WGS (in addition to lowering the cost) will be the interpretation of the data generated in a correct and time-efficient manner. The way such genome wide data would have to be interpreted would need to take into account the proposed model of inheritance (often uncertain) and the genetic cause hypothesis (monogenic vs polygenic or multifactorial). If today it is challenging to determine the real impact of a variant in a patient's phenotype in a supposedly monogenic disease (due to incomplete penetrance, phenotypic variability, etc), this would compose a much greater challenge in oligogenic and/or multifactorial disorders, as may be the case, in many instances, for ID and other NDDs. In this line of thought, we need to stress the crucial importance of parental DNA availability and a

complete and consistent clinical description of the patient and family history whenever a technique like aCGH or MPS is to be used. This challenge was faced many times throughout the elaboration of this thesis, given the limited clinical information many times provided by the clinician requesting the exam. This limitation was more striking in the patients collected in the context of the CC presented in sub-chapter 2.1 (these patients were referred for aCGH though a diagnostic lab, and only after a positive results were the physician and families contacted in order to enter the research study) and in sub-chapter 3.2 (where the parents DNA was not collected at the same time as the patient's).

Main findings and conclusions of this thesis

1. aCGH application in two cohorts of Portuguese ID patients retrieved a diagnostic yield of 14%, similar to that obtained in previous published studies in other populations;
2. 12 new candidate *loci* for ID were put forward and the most likely relevant genes within these *loci* pinpointed: deletions - 2q11.2-q12.2 (*POU3F3*), 7q33 (*CALD1*, *EXOC4*, *CNOT4*), 10q26.3 (*EBF3*), 17p11.2 (*EPN2*, *RNF112*, *ULK2*), 20q13.12-q13.13 (*KCNB1*, *ARFGEF2*); duplications - 1p22.1-p21.3 (*FAM69A*), 9q33.2-q33.3 (*FBXW2*, *LHX2*, *LHX6*, *DENND1A*), 12p13.33 (*CACNA1C*, *TULP3*, *TSPAN9*), Xq24 (*CUL4B*) and Xq26.3 (*ARHGEF6*);
3. The effect of small CNVs at the 1q43-q44 region affecting *AKT3* on brain size was shown to be mostly toward microcephaly in the case of deletion, however their phenotypic impact in NDD should be interpreted with care;
4. Deletions in the 2p13.2 citoband were shown to lead to haploinsufficiency of *EXOC6B* (involved in exocytosis and Notch signaling) and *CYP26B1* (involved in retinoic acid metabolism) and consequently cause an NDD with associated behavior, craniofacial and skeletal features;
5. Small CNVs at 7q33 region were shown to be a very rare genomic imbalance, causing ID and behavioral alterations, *CALD1* being the most likely candidate gene for the cognitive phenotype;
6. Deletions in *10q26.3* region encompassing the *EBF3* gene were shown to be associated with ID, hypotonia, characteristic facial dysmorphisms (triangular looking facies) and behavior alterations, reinforcing the haploinsufficiency role of *EBF3* in disease.

7. Rett syndrome-like phenotypes were shown to be caused by structural anomalies impairing *MECP2* gene methylation and by CNV or SNVs affecting other genes that work in the same pathways as *MECP2*, *CDKL5* and *FOXG1*. These novel candidate disease genes are *HTT*, *SMARCA1*, *GABBR2*, *RHOBTB2* and *EIF4G1*;
8. The re-analysis of previously unsolved cases studied by MPS was shown to be a very important point to be considered in diagnostic laboratories, and was shown to solve a significant portion of the cases already sequenced. Pathogenic variants were identified in the *FBXL4*, *KMT2B* and *MAGEL2* genes; novel likely pathogenic variants were identified in the *DNAJC21*, *MYOD1*, *PAX7*, *NBAS*, *RARS2*, *SPATA5*, *AUTS2*, *MYH7* and *SMARCAB1* genes.

In a global manner we can say that we identified new *loci* and genes likely relevant for NDDs and established genotype-phenotype correlations.

Future perspectives

In this thesis we applied genome-wide techniques to unravel the genetic causes underlying NDDs. Although we focused our description in the positive findings of this thesis, we should not forget that a significant portion of the patients we screened still remain unsolved. We believe that this is mainly due to the limitations in technical approaches used. When a patient with NDDs has negative report this doesn't mean that there is no mutation underlying the disease but just that we still lack the means to determine it or its effect. In the near future we would like to perform WES or WGS in all the negative cases from chapter 2 (i.e., in all the patients negative by aCGH). Additionally, we intend to re-evaluate also all the cases carrying variants classified as VOUS in the future. With this analysis, we believe we would be able to clarify some of the cases as, in time, a variant can move in classification to benign or pathogenic.

Still concerning the negative cases, we aim to re-evaluate all the negative cases left from the WES studies presented in chapter 3. Maybe meanwhile a gene left behind due to its apparently less interesting function is meanwhile reported in patients with an overlapping clinical presentation, leading to a change in perspective about its role in NDD. Also, we aim to keep in contact with other groups and keep using databases such as GeneMatcher in order try to solve more cases.

As stated before, one of the biggest lacks in genetics of ID is the absence of longitudinal studies for mic/dup syndromes and specific gene associated disorders. In a similar manner, the majority

of the patients included in this work had a clinical description performed at a certain age which was used for publication. This is the particular case of the patients with particular CNVs described, such as the patients with 2p13.2 deletions and 7q33 del/dup here described for the first time. It would be of particular interest to re-evaluate these same patients in a few years in determine how their development occurs entering teenage years and adulthood. Similarly, it would be of great clinical value to perform the same re-evaluation in all the patients presenting known and already known and novel likely pathogenic variants.

Finally, we would like to perform functional studies comprising several of the variants found in order to clarify the functional contribution of the variants for neurodevelopment. This type of studies are often lacking when describing new candidate variants since, at the pace that new genes are currently discovered using genome-wide techniques, it is not feasible to develop an mouse model for each gene. However, there are other models available that provide a better balance between time of model development and functional outputs that could be used (such as *C. elegans*, iPSCs, etc).

References

Abel HJ, Duncavage EJ. Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches. Cancer Genet. 2013; 206: 432–440.

Bagheri H, Badduke C, Qiao Y, Colnaghi R, Abramowicz I, Alcantara D, et al. Identifying candidate genes for 2p15p16.1 microdeletion syndrome using clinical, genomic, and functional analysis. JCI Insight 2016; 1: e85461.

Ballif BC, Rosenfeld JA, Traylor R, Theisen A, Bader PI, Ladda RL, et al. High-resolution array CGH defines critical regions and candidate genes for microcephaly, abnormalities of the corpus callosum, and seizure phenotypes in patients with microdeletions of 1q43q44. Hum. Genet. 2012; 131: 145–156.

Bellen HJ, Tong C, Tsuda H. 100 years of Drosophila research and its impact on vertebrate neuroscience: a history lesson for the future. Nat. Rev. Neurosci. 2010; 11: 514–522.

Bessa C, Maciel P, Rodrigues AJ. Using C. elegans to Decipher the Cellular and Molecular Mechanisms Underlying Neurodevelopmental Disorders. Mol. Neurobiol. 2013

Buysse K, Delle Chiaie B, Van Coster R, Loeys B, De Paepe A, Mortier G, et al. Challenges for CNV interpretation in clinical molecular karyotyping: lessons learned from a 1001 sample experience. Eur. J. Med. Genet. 2009; 52: 398–403.

Cappuccio G, Vitiello F, Casertano A, Fontana P, Genesio R, Bruzzese D, et al. New insights in the interpretation of array-CGH: autism spectrum disorder and positive family history for intellectual disability predict the detection of pathogenic variants. Ital. J. Pediatr. 2016; 42: 39.

D'Amico A, Bertini E. Congenital myopathies. Curr. Neurol. Neurosci. Rep. 2008; 8: 73–79.

Davis EE, Frangakis S, Katsanis N. Interpreting human genetic variation with in vivo zebrafish assays. Biochim. Biophys. Acta 2014; 1842: 1960–1970.

Dzafic E, Strzyz PJ, Wilsch-Bräuninger M, Norden C. Centriole Amplification in Zebrafish Affects Proliferation and Survival but Not Differentiation of Neural Progenitor Cells. Cell Rep. 2015; 13: 168–182.

EuroEPINOMICS-RES Consortium, Epilepsy Phenome/Genome Project, Epi4K Consortium. De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. Am. J. Hum. Genet. 2014; 95: 360–370.

Evers C, Maas B, Koch KA, Jauch A, Janssen JWG, Sutter C, et al. Mosaic deletion of EXOC6B: further evidence for an important role of the exocyst complex in the pathogenesis of intellectual disability. Am. J. Med. Genet. A. 2014; 164A: 3088–3094.

Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am. J. Hum. Genet. 2009; 84: 524–533.

Fulda KG. Ethical issues in predictive genetic testing: a public health perspective. J. Med. Ethics 2006; 32: 143–147.

Gilissen C, Hehir-Kwa JY, Thung DT, van de Vorst M, van Bon BWM, Willemsen MH, et al. Genome sequencing identifies major causes of severe intellectual disability. Nature 2014; 511: 344–347.

Gregg AR, Van den Veyver IB, Gross SJ, Madankumar R, Rink BD, Norton ME. Noninvasive prenatal screening by next-generation sequencing. Annu. Rev. Genomics Hum. Genet. 2014; 15: 327–347.

Guemez-Gamboa A, Nguyen LN, Yang H, Zaki MS, Kara M, Ben-Omran T, et al. Inactivating mutations in MFSD2A, required for omega-3 fatty acid transport in brain, cause a lethal microcephaly syndrome. Nat. Genet. 2015; 47: 809–813.

Homann OR, Misura K, Lamas E, Sandrock RW, Nelson P, McDonough SI, et al. Whole-genome sequencing in multiplex families with psychoses reveals mutations in the SHANK2 and SMARCA1 genes segregating with illness. Mol. Psychiatry 2016; 21: 1690–1695.

Karaca E, Harel T, Pehlivan D, Jhangiani SN, Gambin T, Coban Akdemir Z, et al. Genes that Affect Brain Structure and Function Identified by Rare Variant Analyses of Mendelian Neurologic Disease. Neuron 2015; 88: 499–513.

Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants: Genet. Med. 2011; 13: 680–685.

Khattak S, Brimble E, Zhang W, Zaslavsky K, Strong E, Ross PJ, et al. Human induced pluripotent stem cell derived neurons as a model for Williams-Beuren syndrome. Mol. Brain 2015; 8: 77.

Kinser HE, Pincus Z. High-throughput screening in the C. elegans nervous system. Mol. Cell. Neurosci. 2016

Klopocki E, Schulze H, Strauss G, Ott C-E, Hall J, Trotier F, et al. Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome. Am. J. Hum. Genet. 2007; 80: 232–240.

Kramer JM, Kochinke K, Oortveld MAW, Marks H, Kramer D, de Jong EK, et al. Epigenetic regulation of learning and memory by Drosophila EHMT/G9a. PLoS Biol. 2011; 9: e1000569.

Lu X, Shaw CA, Patel A, Li J, Cooper ML, Wells WR, et al. Clinical implementation of chromosomal microarray analysis: summary of 2513 postnatal cases. PloS One 2007; 2: e327.

Makela NL, Birch PH, Friedman JM, Marra CA. Parental perceived value of a diagnosis for intellectual disability (ID): a qualitative comparison of families with and without a diagnosis for their child's ID. Am. J. Med. Genet. A. 2009; 149A: 2393–2402.

Mankinen CB, Sears JW, Alvarez VR. Terminal (1)(q43) long-arm deletion of chromosome no. 1 in a three-year-old female. Birth Defects Orig. Artic. Ser. 1976; 12: 131–136.

Männik K, Parkel S, Palta P, Zilina O, Puusepp H, Esko T, et al. A parallel SNP array study of genomic aberrations associated with mental retardation in patients and general population in Estonia. Eur. J. Med. Genet. 2011; 54: 136–143.

Mardis ER. New strategies and emerging technologies for massively parallel sequencing: applications in medical research. Genome Med. 2009; 1: 40.

Martin CL, Warburton D. Detection of Chromosomal Aberrations in Clinical Practice: From Karyotype to Genome Sequence. Annu. Rev. Genomics Hum. Genet. 2015; 16: 309–326.

Mencarelli MA, Katzaki E, Papa FT, Sampieri K, Caselli R, Uliana V, et al. Private inherited microdeletion/microduplications: implications in clinical practice. Eur. J. Med. Genet. 2008; 51: 409–416.

Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am. J. Hum. Genet. 2010; 86: 749–764.

Minarik G, Repiska G, Hyblova M, Nagyova E, Soltys K, Budis J, et al. Utilization of Benchtop Next Generation Sequencing Platforms Ion Torrent PGM and MiSeq in Noninvasive Prenatal Testing for Chromosome 21 Trisomy and Testing of Impact of In Silico and Physical Size Selection on Its Analytical Performance. PloS One 2015; 10: e0144811.

Mora C, Serzanti M, Consiglio A, Memo M, Dell'Era P. Clinical potentials of human pluripotent stem cells. Cell Biol. Toxicol. 2017

Petersen B-S, Fredrich B, Hoeppner MP, Ellinghaus D, Franke A. Opportunities and challenges of whole-genome and -exome sequencing. BMC Genet. 2017; 18: 14.

Qiao Y, Mercier E, Dastan J, Hurlburt J, McGillivray B, Chudley AE, et al. Copy number variants (CNVs) analysis in a deeply phenotyped cohort of individuals with intellectual disability (ID). BMC Med. Genet. 2014; 15: 82.

van Ravenswaaij-Arts CMA, Kleefstra T. Emerging microdeletion and microduplication syndromes; the counseling paradigm. Eur. J. Med. Genet. 2009; 52: 75–76.

Rodríguez-Revenga L, Vallespín E, Madrigal I, Palomares M, Mur A, García-Miñaur S, et al. A parallel study of different array-CGH platforms in a set of Spanish patients with developmental delay and intellectual disability. Gene 2013; 521: 82–86.

Rosenberg C, Knijnenburg J, Bakker E, Vianna-Morgante AM, Sloos W, Otto PA, et al. Array-CGH detection of micro rearrangements in mentally retarded individuals: clinical significance of imbalances present both in affected children and normal parents. J. Med. Genet. 2006; 43: 180–186.

Sagoo GS, Butterworth AS, Sanderson S, Shaw-Smith C, Higgins JPT, Burton H. Array CGH in patients with learning disability (mental retardation) and congenital anomalies: updated systematic review and meta-analysis of 19 studies and 13,926 subjects. Genet. Med. Off. J. Am. Coll. Med. Genet. 2009; 11: 139–146.

Shi Y, Inoue H, Wu JC, Yamanaka S. Induced pluripotent stem cell technology: a decade of progress. Nat. Rev. Drug Discov. 2017; 16: 115–130.

Slavotinek AM. Novel microdeletion syndromes detected by chromosome microarrays. Hum. Genet. 2008; 124: 1–17.

Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum. Mutat. 2015; 36: 928–930.

Song Z, Zhang X, Jia S, Yelick PC, Zhao C. Zebrafish as a Model for Human Ciliopathies. J. Genet. Genomics Yi Chuan Xue Bao 2016; 43: 107–120.

Takahashi K, Yoshina S, Masashi M, Ito W, Inoue T, Shiwaku H, et al. Nematode homologue of PQBP1, a mental retardation causative gene, is involved in lipid metabolism. PloS One 2009; 4: e4104.

Teixeira-Castro A, Ailion M, Jalles A, Brignull HR, Vilaça JL, Dias N, et al. Neuron-specific proteotoxicity of mutant ataxin-3 in C. elegans: rescue by the DAF-16 and HSF-1 pathways. Hum. Mol. Genet. 2011; 20: 2996–3009.

Torres F, Barbosa M, Maciel P. Recurrent copy number variations as risk factors for neurodevelopmental disorders: critical overview and analysis of clinical implications. J. Med. Genet. 2016; 53: 73–90.

Vissers LELM, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. Nat. Rev. Genet. 2016; 17: 9–18.

van der Voet M, Nijhof B, Oortveld MAW, Schenck A. Drosophila models of early onset cognitive disorders and their clinical applications. Neurosci. Biobehav. Rev. 2014; 46 Pt 2: 326–342.

Wen J, Lopes F, Soares G, Farrell SA, Nelson C, Qiao Y, et al. Phenotypic and functional consequences of haploinsufficiency of genes from exocyst and retinoic acid pathway due to a recurrent microdeletion of 2p13.2. Orphanet J. Rare Dis. 2013; 8: 100.

Wenger AM, Guturu H, Bernstein JA, Bejerano G. Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. Genet. Med. Off. J. Am. Coll. Med. Genet. 2017; 19: 209–214.

Willemsen MH, Nijhof B, Fenckova M, Nillesen WM, Bongers EMHF, Castells-Nobau A, et al. GATAD2B loss-of-function mutations cause a recognisable syndrome with intellectual disability and are associated with learning deficits and synaptic undergrowth in Drosophila. J. Med. Genet. 2013; 50: 507–514.

Xiang B, Zhu H, Shen Y, Miller DT, Lu K, Hu X, et al. Genome-wide oligonucleotide array comparative genomic hybridization for etiological diagnosis of mental retardation: a multicenter experience of 1499 clinical cases. J. Mol. Diagn. JMD 2010; 12: 204–212.

Xu L, Shi R. Noninvasive prenatal diagnosis using next-generation sequencing. Gynecol. Obstet. Invest. 2014; 77: 73–77.