U.PORTO

CICLO DE ESTUDOS

PDPGM

Genomic approach in idiopathic intellectual disability

Maria de Fátima e Costa Torres

D

2018

U.PORTO

Autor: Maria de Fátima e Costa Torres

Genomic approach in idiopathic intellectual disability

D.ICBAS **2018**

Genomic approach in idiopathic intellectual disability

Maria de Fátima e Costa Torres

SEDE ADMINISTRATIVA

INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR

FACULDADE DE MEDICINA

MARIA DE FÁTIMA E COSTA TORRES

GENOMIC APPROACH IN IDIOPATHIC INTELLECTUAL DISABILITY

Tese de Candidatura ao grau de Doutor em Patologia e Genética Molecular, submetida ao Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto

Orientadora – Doutora Patrícia Espinheira de Sá Maciel

Categoria – Professora Associada

Afiliação – Escola de Medicina e Ciências da Saúde da Universidade do Minho

Coorientadora – Doutora Maria da Purificação Valenzuela Sampaio Tavares

Categoria – Professora Catedrática

Afiliação – Faculdade de Medicina Dentária da Universidade do Porto

Coorientadora – Doutora Filipa Abreu Gomes de Carvalho

Categoria – Professora Auxiliar com Agregação

Afiliação – Faculdade de Medicina da Universidade do Porto

U.PORTO

DECLARAÇÃO Dissertação/Tese

Identificação do autor

Nome completo _Maria de Fátima e Costa Torres_

N.º de identificação civil _07718822 **N.º de estudante** __ 198600524___

Email institucional _fatima.torres@gcgenetics.com OU: up198600524@icbas.up.pt _

Email alternativo _mfctorres@gmail.com _ **Tlf/Tlm** _918197020_

Ciclo de estudos (Mestrado/Doutoramento) _Patologia e Genética Molecular__

Faculdade/Instituto _Instituto de Ciências Biomédicas de Abel Salazar_

Identificação da publicação

Dissertação de mestrado [ ] **Tese de doutoramento** [X]

Título completo _Abordagem genómica no atraso mental idiopático /Genomic approach in idiopathic intelectual disability_

Orientador _Patrícia Espinheira de Sá Maciel_

Coorientador(es) _Maria da Purificação Valenzuela Sampaio Tavares e Filipa Abreu Gomes de Carvalho

Data da conclusão (defesa pública) _11 de Junho de 2019

Especialidade /Ramo de Conhecimento __Genética _

Palavras-chave _Perturbações do neurodesenvolvimento_; défice intelectual_; _epilepsia_; __aCGH__; _WES_

Declaro, sob compromisso de honra, que

[ ] **O exemplar da publicação corresponde à versão final aprovada pelo júri**

[ ] **Autorizo a disponibilização imediata do texto integral no Repositório da U.Porto**

Autorizo a disponibilização do texto integral no Repositório da U.Porto no prazo de:

[ ] **6 Meses**

[X] **12 Meses**

Justificação para a não autorização imediata _Há estudos ainda em curso que se prevê serem finalizados e publicados no decorrer do próximo ano

Mantenho todos os direitos de uso em trabalhos futuros.

Data _____

Assinatura _____

Aos meus pais e à minha família, por tudo;
ao meu pai, que nos deixou a 19-02-2018.

Agradecimentos

Há um provérbio Africano que diz:
"*Se queres ir depressa, vai sozinho. Se queres ir longe, vai acompanhado*".

Seria impossível ter realizado todo este trabalho, condensado agora nesta tese, sem o contributo e o apoio de várias pessoas a quem gostaria de agradecer publicamente:

À Professora Doutora Patrícia Maciel, por ter aceite orientar este trabalho, mesmo sendo eu uma *outsider* da Universidade do Minho. Patrícia, muito obrigada por todo o trabalho extra (reuniões, discussões, artigos, telefonemas e emails) que tiveste durante estes anos! Foi um prazer trabalhar contigo e participar ao longo deste tempo nos vários projetos que conduziram a esta tese!

À Professora Doutora Purificação Tavares, pela disponibilidade demonstrada, quer pessoal, ao ser coorientadora do trabalho, quer institucional, ao disponibilizar o uso de dados do CGC para serem usados como base deste trabalho.

À Professora Doutora Filipa Carvalho, pela disponibilidade demonstrada na coorientação deste trabalho e pelo apoio dado ao longo deste percurso.

À Fátima L., pelas muitas experiências, reuniões, discussões, artigos, telefonemas e emails que partilhámos, tanto nos trabalhos de aCGH como no estudo de epilepsia/WES: muito obrigada por tudo, e em particular pelas tuas providenciais fórmulas nas folhas de Excel ☺ e pela ajuda nesta fase final com os dados de WES. Foi um prazer trabalhar contigo e espero que o contacto se mantenha. Desejo-te muitas felicidades e o maior sucesso para a tua vida, pessoal e profissional.

À Dra. Paula Rendeiro, Directora do Laboratório de Citogenética do CGC, pela colaboração e partilha de dados de aCGH, bem como pela contribuição na realização de análises suplementares de FISH, que serviram de complemento a alguns casos.

Ao Dr. Jorge Pinto-Basto, Director do Laboratório de Diagnóstico Molecular e Genómica Clínica do CGC, por ter possibilitado a realização de extrações de DNA/RNA para além, nesta fase final do trabalho, das sequenciações de Sanger necessárias à complementação de certos estudos.

A todos os colegas do CGC que de alguma forma contribuíram para este trabalho, em particular:
- À equipa da ARCA, por todos os registos extra que fizeram;
- À equipa da EXTRACÇÃO (Catarina Ribeiro, Marta e Aleksandra), pelas amostras adicionais de DNA e RNA que processaram;
- Aos colegas "voluntários" que contribuíram com colheitas de sangue para os estudos de expressão (Cláudia, Catarina Varela, Diana, Sara Teixeira, Cristiano, Marcos, Paulo Brochado), bem como à Ana Paula e à Clara, que fizeram as colheitas;
- Às meninas que trabalham diretamente comigo (Alexandra Lopes, Liliana Gonçalves, Lígia Lameiras, Carolina e, também, à Inês Eulálio): muito obrigada pelo interesse e pelo apoio demonstrados; um agradecimento especial à Carolina, pela ajuda nesta fase final, com as sequenciações de Sanger.

E, *last but not the least*, à minha enorme família e aos meus (também enormes) amigos, por todo o apoio e amizade!

O meu muito obrigada a todos!

Abstract

Neurodevelopmental disorders (NDs) are a large group of clinical entities caused by the disruption of brain development, that impact personal, social, academic and occupational functioning. Intellectual disability (ID), formerly referred to as mental retardation, is one of the most common NDs, affecting nearly 1-3% of the world population, with onset in infancy or in the early childhood, being characterized by significant limitations both in intellectual functioning and in adaptive behavior as expressed in conceptual, social and practical adaptive skills.

Epilepsies are common complex neurological diseases, characterized by recurrent (two or more) spontaneous seizures, that constitute the third leading contributor to the global burden of disease for neurological disorders. Epilepsy is a frequent feature of NDs, being commonly associated with ID.

The etiology of NDs, in general, and of ID and epilepsy, in particular, is complex and often results from the combination of environmental or acquired (non-genetic) and genetic factors. Genetic causes are thought to be responsible for 17-50% of the cases of ID, while in about 70% of the epilepsy cases, genetic factors play a major role, polygenic cases being predominant. Both epilepsy and ID represent a major burden not only for the individual, but also to his/her immediate family. Being so, the clarification of etiology is important to provide information regarding prognosis or expected clinical course, being fundamental to guide an adequate intervention.

The determination of the precise molecular cause (genotype) that explains the clinical features of a particular disease (phenotype) may be a considerable challenge, either due to an incomplete knowledge about a particular disease or due to the variability associated with certain conditions. However, the rapid development in genetic tools and sequencing technology in the past few years had a particular impact in the understanding of the genetics of NDs. Advances in chromosomal microarray technology have allowed for the analysis of copy number variations (CNVs) in very large case–control cohorts and highlighted the biological relevance of these structural variants. On the other hand, the great advances of massive parallel sequencing (MPS) technologies expanded the sequencing coverage of the genome and potentiated the capacity to identify variants that explain many Mendelian diseases in both known and novel disease genes.

In this work, we have used both aCGH and MPS techniques in order to unravel the genetic etiology of neurodevelopmental disease in a large series of patients of Portuguese origin. aCGH was performed in a large group of patients with idiopathic ID and in a large family with epilepsy, while WES analysis was performed in selected patients from this group. These techniques were complemented, whenever necessary, with quantitative polymerase chain reaction (qPCR) and qRT-PCR or RT-PCR plus Sanger sequencing.

With this approach, we were able to identify pathogenic or likely pathogenic causal variants in 13.2% of the patients of the ID group, and to pinpoint 12 new candidate *loci* for ID. Moreover, we confirmed the *AKT3* gene as a key gene for microcephaly, although recognizing the role of other factors in the manifestation of the phenotype. We showed that microduplications in the 2p15p16.1 region, although rare, are consistently associated with ID and specific morphological anomalies. We have proposed the *CALD1* gene as the most likely candidate for the core phenotype (i.e. ID and behavioral alterations) associated with small CNVs at 7q33 region. We have observed an increased expression consistent with the change in copy number in several genes within duplications and proposed the increased expression of the *CUL4B* gene in a patient carrying a Xq24 duplication, as the most likely disease contributor for patient´s phenotype. Finally, the WES analysis performed in three patients of a large Portuguese family with epilepsy allowed us to pinpoint the *FERMT2* gene as the most likely candidate to explain the disease phenotype.

In summary, this work has contributed to the elucidation of the genetic aetiology of disease in several patients and families, as well as to the identification of new candidate *loci* and genes for ID and epilepsy.

Resumo

As perturbações do neurodesenvolvimento constituem um grande grupo de entidades clínicas causadas pela perturbação do desenvolvimento do cérebro, que têm impacto a nível do funcionamento pessoal, social, académico e ocupacional. O défice intelectual, anteriormente denominado de atraso mental, é uma das perturbações do neurodesenvolvimento mais comuns, que afeta cerca de 1-3% da população mundial, e que se manifesta usualmente ao longo da infância, sendo caracterizada por limitações significativas no funcionamento intelectual e no comportamento adaptativo, expresso nas capacidades a nível conceptual, social e prático.

As epilepsias são doenças neurológicas comuns caracterizadas por convulsões recorrentes e espontâneas, e constituem o terceiro grande grupo no que respeita à contribuição para o impacto das doenças neurológicas na sociedade. A epilepsia é uma característica frequente das perturbações do neurodesenvolvimento, e está frequentemente associada ao défice intelectual.

A etiologia das perturbações do neurodesenvolvimento, em geral, e do défice intelectual e da epilepsia, em particular, é complexa e normalmente resulta da combinação de fatores não genéticos (ambientais ou adquiridos) e genéticos. As causas genéticas estão na origem de 17-50% dos casos de défice intelectual, e desempenham um papel importante em cerca de 70% dos casos de epilepsia, onde os casos poligénicos são predominantes. Quer a epilepsia quer o défice intelectual representam um grande peso não só para o indivíduo, mas também para a sua família. Assim sendo, a clarificação etiológica da doença é importante na obtenção de informação relativa ao prognóstico e evolução clínica, sendo fundamental para uma intervenção adequada.

A determinação da causa molecular (genótipo) que explica as características de uma dada doença (fenótipo) pode, por vezes, ser um desafio considerável, não só devido ao conhecimento incompleto relativo à doença particular, mas também por causa da variabilidade fenotípica associada a certas condições. Contudo, o rápido desenvolvimento, nos últimos anos, de ferramentas na área da genética e nas tecnologias de sequenciação teve um impacto particular no conhecimento da genética das perturbações do neurodesenvolvimento. Os avanços na tecnologia do estudo de anomalias cromossómicas recorrendo a arrays de hibridização genómica comparativa (aCGH) permitiram a análise de variações de número de

cópia (CNVs) em séries muito grandes de indivíduos (afetados vs indivíduos controlo) e evidenciaram a importância biológica destas variantes. Por outro lado, os avanços na tecnologia de sequenciação massiva paralela (MPS) permitiram a expansão na cobertura de sequenciação do genoma e consequentemente deram lugar à identificação de variantes que explicam muitas doenças Mendelianas, quer em genes conhecidos, quer em novos genes.

Neste trabalho, usámos as técnicas de aCGH e de MPS, em particular a sequenciação do exoma (WES), para caracterizar a etiologia genética de perturbações de neurodesenvolvimento num grande grupo de doentes de origem portuguesa. O aCGH foi utilizado tanto no grupo de doentes com défice intelectual, como numa família alargada com diagnóstico de epilepsia, enquanto que o WES foi realizado em doentes selecionados deste grupo. Estas técnicas foram complementadas, sempre que necessário, por PCR quantitativo em tempo real (qPCR), qRT-PCR ou RT-PCR seguido de sequenciação de Sanger.

Com esta abordagem, foi possível identificar variantes causais patogénicas ou provavelmente patogénicas em 13.2% dos doentes com défice intelectual, para além de se terem identificado 12 novos *loci* candidatos, muito possivelmente associados ao défice intelectual. Para além disso, confirmámos que o *AKT3* é um gene chave para a microcefalia, reconhecendo, no entanto, a influência de outros fatores no fenótipo. Mostrámos, ainda, que as microduplicações na 2p15p16.1, apesar de raras, estão consistentemente associadas a défice intelectual e a anomalias morfológicas específicas. Propusemos o gene *CALD1* como o candidato mais provável para explicar o fenótipo base (défice intelectual e alterações de comportamento) associado a CNVs na região 7q33. Detetámos um aumento de expressão, consistente com a variação do número de cópias, em vários genes localizados dentro de duplicações, e propusemos como causa provável para explicar o fenótipo num doente com uma duplicação em Xq24, o aumento de expressão do *CUL4B* observado. Por fim, a sequenciação do exoma realizada em três doentes de uma família com epilepsia permitiu a identificação do gene *FERMT2* como o candidato mais provável para explicar a doença nesta família.

Em resumo, este trabalho contribuiu para a clarificação da etiologia genética em vários doentes e famílias, bem como para a identificação de novos *loci* e genes candidatos para o défice intelectual e a epilepsia.

Table of contents

Abbreviations used throughout this work

AAIDD – American Association on Intellectual and Developmental Disabilities

aCGH – array Comparative Genomic Hybridization

ACMG – American College of Medical Genetics and Genomics

ADHD – Attention Deficit and Hyperactivity Disorder

AE – Absence epilepsy

ASD – Autism Spectrum Disorder(s)

BD – Bipolar Disease

bp – base pair

BFNIS – Benign Familial Neonatal-Infantile Seizures

BMPs – Bone Morphogenic Proteins

CAE – Childhood Absence Epilepsy

cDNA – complementary DNA

CoNIFER – Copy Number Inference From Exome Reads

CNS – Central Nervous System

CNVs – Copy Number Variations

DECIPHER – DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources

DD – Developmental Delay

DDD – Deciphering Developmental Disorders

DDG2P – Developmental Disorders Genotype-Phenotype Database

DGV – Database of Genomic Variants

DSM – Diagnostic and Statistical Manual of Mental Disorders

ECM – Extracellular Matrix

EEG – Electroencephalogram

EGI – Epilepsy Genetics Initiative

ESHG – European Society of Human Genetics

FoSTeS – Fork stalling and template switching

GDD – Global Developmental Delay

GEE – Generalized Epileptiform activity

GEFS+ – Generalized Epilepsy with Febrile Seizures plus

GGEs – Genetic Generalized Epilepsies

GTCSA – Generalized Tonic–Clonic Seizures Alone

GTPases – Guanosine Triphosphatases

GWAS – Genome-wide association studies

HGMD – Human Genome Mutation Database

IAE – Idiopathic Absence Epilepsy

ID – Intellectual Disability
IEGTC – Intractable Epilepsy with Generalized Tonic-Clonic seizures
IEM – Inborn Error of Metabolism
IFs – Incidental Findings
IGEs – Idiopathic Generalized Epilepsies
IGV – Integrative Genomics Viewer
ILAE – International League Against Epilepsy
iPSCs – induced Pluripotent Stem Cells
IQ – Intelligence Quotient
IRDiRC – International Rare Diseases Research Consortium
IRUD – Initiative on Rare and Undiagnosed Diseases
JAE – Juvenile Absence Epilepsy
JME – Juvenile Myoclonic Epilepsy
LCLs – Lymphoblastoid Cell Lines
LCRs – Low-copy repeats
LOH – Loss of Heterozygosity
MCDs – Malformations of Cortical Development
MEM – Mendelian Error Method
MMEJ – Microhomology-Mediated End-Joining
MMBIR – Microhomology-Mediated Break-Induced Replication
MPS – Massive Parallel Sequencing
MPSI – Migrating Partial Seizures of Infancy
mRNA – messenger RNA
mTOR – Mammalian Target of Rapamycin
NAHR – Nonallelic Homologous Recombination
NHEJ – Nonhomologous End Joining
NDs – Neurodevelopmental Disorders
NGS – Next Generation Sequencing
OMIM – Online Mendelian Inheritance In Man
OR – Odds Ratio
PCR – Polymerase Chain Reaction
PPV – Positive Predictive Value
qRT-PCR – quantitative Reverse Transcription and (real-time) PCR
qPCR – quantitative (real-time) PCR
RBPs – mRNA Binding Proteins
SDs – Segmental Duplications
SMEI – Severe Myoclonic Epilepsy of Infancy

SNPs – Single Nucleotide Polymorphisms

SNVs – Single Nucleotide Variants

SvABA – Structural variation and indel Analysis by Assembly

SVs – Structural Variants

SZ – Schizophrenia

TADs – Topological Activated Domains

TH – Thyroid Hormone

TPRT – Target site Primed Reverse Transcription

UPD – Uniparental Disomy

VOUS – Variants of Unknown Clinical Significance

WES – Whole Exome Sequencing

WGS – Whole Genome Sequencing

XHMM – eXome-Hidden Markov Model

Thesis Outline

This dissertation is divided in four chapters and includes both published and unpublished data. The data presented relates to the identification and discussion of the genetic lesions detected in a group of patients with neurodevelopmental disorders (NDs), namely intellectual disability (ID) and epilepsy.

Chapter 1 is the general introduction to the theme of this dissertation. In this chapter an overview of the neurodevelopmental disorders (NDs), their etiological aspects and related pathways, the types of genetic lesions associated with NDs and their detection is provided. Moreover, a brief description of the techniques used, array comparative genomic hybridization (aCGH) and massive parallel sequencing (MPS) techniques, and their contribution to the identification of new disease-associated *loci* and genes is also presented.

Chapter 2 describes the results of the aCGH analysis performed in a group of patients with idiopathic ID and is composed by four sub-chapters.

Sub-chapter 2.1 describes the work performed in a large cohort of Portuguese patients with idiopathic ID, studied by aCGH. A presentation of the main CNVs found, as well as the clinical description of the cohort and a detailed discussion of the genes possibly contributing to the patients' phenotype is provided. The work presented here was submitted for publication.

Sub-chapter 2.2 describes a small report of patients carrying CNVs at the 1q43-q44 region. A comparison of the clinical and genomic imbalances observed in four patients as well as a discussion of the contribution of the *AKT3* gene alterations to the occipital-frontal circumference is provided. The work presented here was submitted for publication.

Sub-chapter 2.3 refers to the collection of four patients from the same family, all with a 2p15 microduplication. A comparison between the clinical and genomic data of these patients with other cases already described in literature is here provided. The work presented here is in preparation for publication.

Sub-chapter 2.4 describes the clinical and genetic features of seven patients with ID, dysmorphisms and behavioral anomalies who carry CNVs at 7q33 cytoband. A comparison of the clinical and genomic imbalances observed in these patients as well as a discussion of the contribution of the *CALD1* gene

alterations to the core phenotype associated with 7q33 CNVs is provided. This work is published in the **Neurogenetics**: Lopes F*, Torres F*, *et al*. (2018) *The contribution of 7q33 copy number variations for intellectual disability*. Neurogenetics. Jan;19(1):27-40. doi: 10.1007/s10048-017-0533-5. Epub 2017 Dec 19. (*both authors contributed equally for this work).

Chapter 3 refers to the aCGH and WES analysis performed in patients from a large family diagnosed with epilepsy. An analysis of aCGH and the WES findings, as well as a discussion of the possible new gene contributors to epilepsy is provided. The work presented here is in preparation for publication.

Chapter 4 is the general discussion of the dissertation. In this chapter an integrated view of the findings is presented, together with the main conclusions and future perspectives of this work.

CHAPTER 1

Introduction

Disclaimer:

This chapter makes direct use of parts of two already published articles:

1. review article published in **Journal of Medical Genetics** (with permission of BMJ Publishing Group Ltd):

Torres F, Barbosa M, Maciel P (2016). Recurrent copy number variations as risk factors for neurodevelopmental disorders: critical overview and analysis of clinical implications. *J Med Genet*. 53(2): 73-90. [DOI: 10.1136/jmedgenet-2015-103366].

2. advanced article published in **Encyclopedia of Life Sciences** (with permission of Wiley & Sons):

Torres, Fátima; Lopes, Fátima; and Maciel, Patrícia (July 2018) Relevance of Copy Number Variation to Human Genetic Disease. In: eLS. John Wiley & Sons, Ltd: Chichester. DOI: 10.1002/9780470015902.a0020226.pub2.

Neurodevelopmental disorders and intellectual disability

Neurodevelopmental disorders (NDs) are a large group of clinical entities caused by the disruption of brain development, that include autism spectrum disorders (ASD), intellectual disability (ID), communication disorders, attention deficit and hyperactivity disorder (ADHD), specific learning disorders and also motor disorders. NDs also include schizophrenia (SZ), that despite manifesting in the adult stage, results from neurodevelopmental disturbances (American Psychiatric Association. 2013). NDs impact personal, social, academic and occupational functioning (Braat and Kooy 2015). Moreover, often several of these entities co-occur, (van Bokhoven 2011), adding further complexity to the health management of people with these conditions (Maulik et al. 2011).

Intellectual disability, formerly referred to as mental retardation, is one of the most common NDs, affecting nearly 1-3% of the world population, with onset in infancy or in the early childhood (Maulik et al. 2011). According to the American Association on Intellectual and Developmental Disability (AAIDD), ID is characterized by significant limitations both in intellectual functioning and in adaptive behavior as expressed in conceptual, social and practical adaptive skills, usually with manifestation before 18 years of age (Moeschler and Shevell 2014). Adaptive behavior comprises conceptual skills (e.g., language, reading, writing, mathematics, reasoning, knowledge, memory and time concepts), social skills (e.g., interpersonal skills, like empathy and social judgment, communication skills, the ability to follow rules, the ability to make and keep friendships, and social problem solving) and practical skills (e.g., activities of daily living, such as personal care, job responsibilities, managing money, occupation, recreation and organizing school and work tasks) (American Psychiatric Association. 2013; Maulik et al. 2011).

These parameters can be evaluated through standardized testing and measures and usually includes the measure of Intelligence quotient (IQ). An IQ score around or below 70 (two standard deviations below the mean of 100 in the population) indicates a limitation in intellectual functioning (American Psychiatric Association. 2013). The severity of ID has been described as “mild,” “moderate,” “severe,” and “profound”, being the IQ score, until recently, the major criteria used for this classification. Nowadays, other assessment criteria include daily skills and the

level of support needed for daily functioning. Table 1.1 describes the distribution of cases within each class and the different criteria used to classify ID.

Table 1.1 – Classification of ID level: proportion of individuals affected, IQ range and functional level in adulthood.

Severity	Proportion of cases	DSM-IV Criteria (classification on the basis of IQ score)	DSM-5 Criteria (classification on the basis of daily skills)	AAIDD Criteria (classification on the basis of intensity of support needed)
Mild	85%	IQ score: 50–69	Independent living, with minimum levels of support	Intermittent support, during transitions or periods of uncertainty
Moderate	10%	IQ score: 36–49	Independent living may be achieved with moderate levels of support, such as those provided in group homes.	Limited support in daily activities
Severe	3.5%	IQ score: 20–35	Daily assistance with self-care activities and safety supervision	Extensive support for daily activities
Profound	1.5%	IQ score<20	24-hour care	Permanent support for every aspect of daily routines.

Borderline intellectual functioning (IQ between 70 and 85) is not considered a disorder but it is an important and frequently unacknowledged comorbid condition. Either genetic liability, or biological causes, such as perinatal difficulties, as well as epigenetic factors such as those resulting from socioeconomic status and maternal stress, can contribute to this condition (Baglio et al. 2014). People with borderline intellectual functioning comprise a vulnerable group: many adults do have problems in adaptive functioning and face difficulties across all areas of ordinary life. Moreover, several studies show that this group is at an increased risk for the development of almost all psychiatric disorders, including personality disorders and substance misuse (Wieland and Zitman 2016).

Global developmental delay (GDD) refers to a significant delay in 2 or more developmental domains, including gross or fine motor skills, speech/language, cognitive and social/personal skills, and activities of daily living, implying deficits in learning and adaptation; it has a prevalence of 1-3%, similar to ID; when the delays are significant, they may predict later ID (Shevell, Ashwal, and Donley 2003). Usually, the term GDD is reserved for younger children, typically younger

than 5 years, whereas ID is usually applied to older children in whom IQ testing was performed (Moeschler and Shevell 2014).

Epilepsy

Epilepsies are common complex neurological diseases, characterized by recurrent (two or more) spontaneous seizures, being the third leading contributor to the global burden of disease for neurological disorders (Devinsky et al. 2018). Epilepsy is a frequent feature of NDs (Heyne et al. 2018) and is commonly associated with ID (Forsgren et al. 2005). The diagnosis of epilepsy is confirmed whenever one of the following conditions is verified: 1) at least two unprovoked (or reflex) seizures occurring >24 hours apart; 2) one unprovoked (or reflex) seizure and a probability of further seizures similar to the general recurrence risk (at least 60%) after two unprovoked seizures, occurring over the next 10 years; 3) a diagnosis of an epilepsy syndrome (Fisher et al. 2014).

Recently, the International League Against Epilepsy (ILAE), has proposed a framework for classification of epilepsies based on 3 levels: 1) seizure type; 2) epilepsy type; 3) inclusion in an epilepsy syndrome. Figure 1.1 summarizes the framework for classification of epilepsies according to ILAE (Scheffer et al. 2017). Whenever possible, all three levels should be sought, as well as the etiology of the individual's epilepsy.

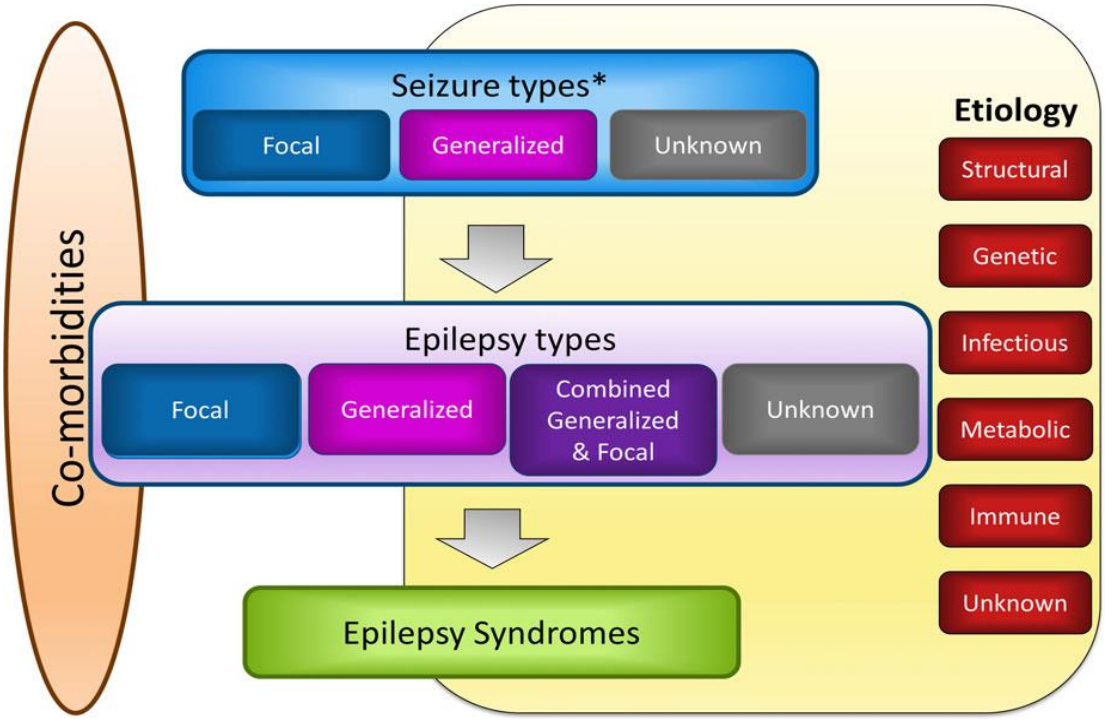


Figure 1.1: Framework for classification of the epilepsies. *Denotes region of onset of seizure [reprinted with permission from the Wiley Press (Scheffer et al. 2017)].

Seizure types

A seizure, as defined by ILAE, is an abnormal electrical perturbation resulting from a network of neurons (Berg et al. 2010). Seizures are defined by onset as: focal (old term “partial”), generalized, unknown, or unclassifiable (Falco-Walter, Scheffer, and Fisher 2018). Focal seizures refer to those that are originated within networks limited to one hemisphere while generalized seizures imply a bilateral hemispheric onset (Sirven 2015). Figure 1.2 summarizes the ILAE 2017 classification of seizure type (Fisher et al. 2017).

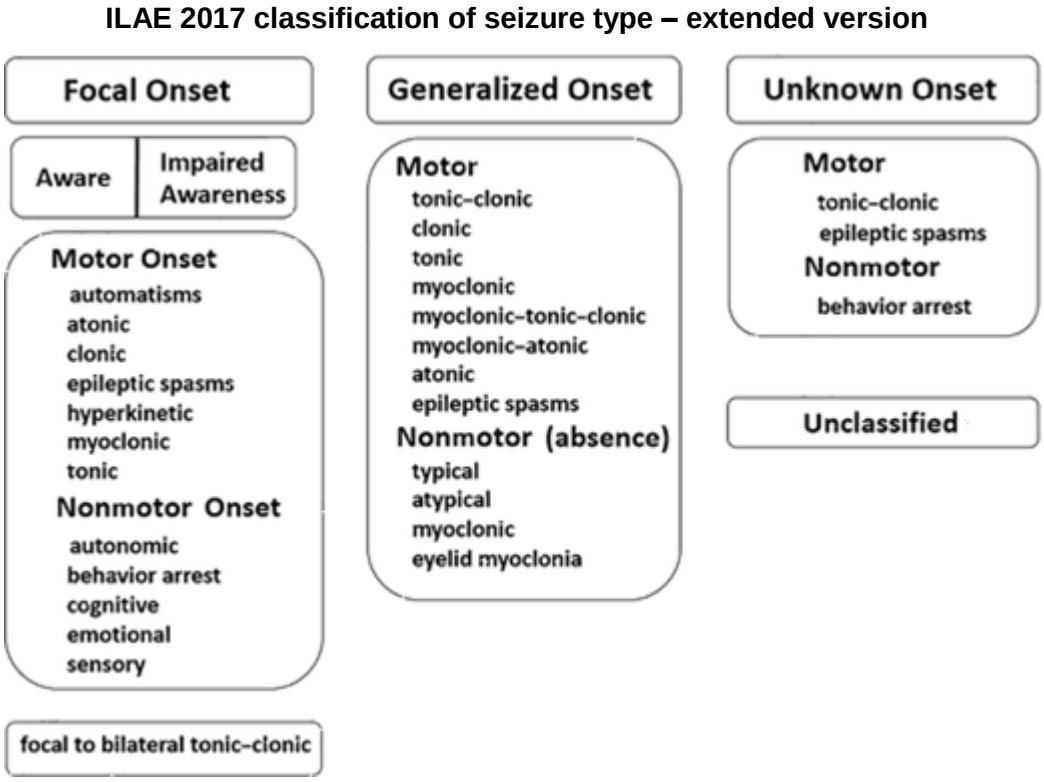


Figure 1.2: The expanded ILAE 2017 classification of seizure types [adapted with permission from (Fisher et al. 2017)]: Seizures should be classified by the earliest prominent feature, some examples being described briefly. Generalized onset seizures – affects both sides of the brain from the start. Focal onset seizures – those that start in one side of the brain, until recently called partial seizures. These seizures can be divided in two groups, according to the awareness (retained awareness meaning that the person is aware of self and environment during the seizure, even if immobile). Focal aware seizures – focal seizures with maintained awareness of the subject, previously called simple partial seizure. Focal impaired awareness seizure – previously called complex partial seizure, when the awareness of what's happening around the person is impaired during any part of the seizure. Focal aware and impaired awareness seizures may be further characterized by one of the motor-onset or non-motor-onset symptoms, reflecting the first prominent sign or symptom in the seizure. Motor seizures – any seizure that involves a change in body movement, may include 1) atonic seizures – sudden loss of muscle tone and strength, manifesting usually by sudden spontaneous falls, usually do not have specified awareness; sometimes called drop attacks; 2) tonic seizures – stiffening of all limbs, without clonic jerking; 3) clonic seizures – rhythmical sustained jerking of limbs and/or head with no tonic stiffening phase; 4) tonic-clonic seizures – loss of awareness, with stiffening of all limbs (tonic phase), followed by sustained rhythmic jerking of limbs and face (clonic phase); the seizure may produce a cry at the start, falling, tongue biting, and incontinence; 5) myoclonic seizures – sudden, short-lasting jerks that can affect some part or the entire body; usually do not affect consciousness; also called myoclonic jerks; 6) myoclonic-tonic-clonic seizures – tonic-clonic seizure, but it is preceded by a

few myoclonic jerks on both sides of the body; 7) epileptic spasms – when he body flexes and extends repeatedly. Non-motor seizures – any seizure that doesn't involve changes in movement, such as 1) cognitive seizures – imply impaired language or other cognitive domains or positive features such as déjà vu, hallucinations, illusions, or perceptual distortions; 2) emotional seizures – involve anxiety, fear, joy, other emotions, or appearance of affect without subjective emotions; 3) absence seizures – manifesting usually by unconsciousness for a few seconds (the person may have the appearance of daydreaming or 'switching off' for people around); may be considered typical or atypical, being the atypical ones characterized by slow onset or termination or significant changes in tone supported by atypical, slow, generalized spike and wave on the electroencephalogram (EEG). Unclassified seizures – refer to seizures that cannot be classified in any of the aforementioned categories.

Epilepsy types

Epilepsy types are classified as: 1) Focal; 2) Generalized; 3) Combined Generalized and Focal; 4) Unknown. This classification considers the possibility of co-occurrence of multiple seizure types in one patient, and incorporates the information about the overall clinical picture, imaging, genetics, laboratory tests, prognoses and comorbidities (Falco-Walter, Scheffer, and Fisher 2018).

Epilepsy syndromes

Epilepsy syndromes refers to clusters of features that occur together (seizure type(s), EEG findings, imaging findings, age-dependent features, triggers and sometimes prognosis). Although there are several well recognized epilepsy syndromes, such as such as Childhood Absence Epilepsy (CAE), West syndrome, or Dravet syndrome, a formally classified list of epilepsy syndromes has been never undertaken by ILAE (Scheffer et al. 2017).

Etiological aspects of ID and epilepsy

The etiology of NDs, in general, and of ID and epilepsy, in particular, is complex and often results from the combination of environmental or acquired (non-genetic) and genetic factors (Reichenberg et al. 2016; Hildebrand et al. 2013).

Among the non-genetic factors contributing to ID there are infections (present at birth or occurring after birth), toxic agents (intrauterine exposure to alcohol, cocaine, amphetamines and other drugs), trauma (before and after birth), nutritional unbalances (such as iodine deficiency, apparently associated with

poorer child cognitive development and educational attainment) (Moeschler and Shevell 2014; Pearce et al. 2016; Huang et al. 2016).

Purely genetic causes are thought to be responsible for 17-50% of the cases of ID (Moeschler 2006; Kaufman, Ayub, and Vincent 2010). Among this category, there is a subgroup constituted by inherited metabolic conditions, or inborn errors of metabolism (IEM), for which an increasing number of treatments has become available: van Karnebeek and Stockler have identified 81 treatable IEM presenting with ID as a major feature (Van Karnebeek and Stockler 2012). An example of developmental improvement is seen in patients with glucose transporter-1 deficiency syndrome, caused by mutations in the *SLC2A1* gene, in whom the ketogenic diet is successful in controlling medicine refractory epilepsy (Leen et al. 2010). Genetically, an important contribution can be attributable to X-linked conditions, which in part explains the fact that almost all the studies of prevalence and incidence of ID show an excess of boys of approximately 40%, rendering testing for X-linked genes in boys with GDD/ID often warranted (Moeschler and Shevell 2014). Additionally, this male over-representation could result from a certain degree of bias: this has been reported at least in ASD, with sex differences in phenotypic presentation, including fewer restricted and repetitive behaviors and externalizing behavioral problems in females, contributing to this male bias (Werling and Geschwind 2013).

As for epilepsies, the scenario is very similar: the etiology of common epilepsies can be defined as a biological continuum, due to the overlap between genetic and acquired cases. Nevertheless, in about 70% of the cases genetic factors play a major role, polygenic cases being predominant (Hildebrand et al. 2013). Figure 1.3 illustrates the relation between phenotypic specificity and genetic heterogeneity associated with some common pediatric conditions, ID and epilepsies included.

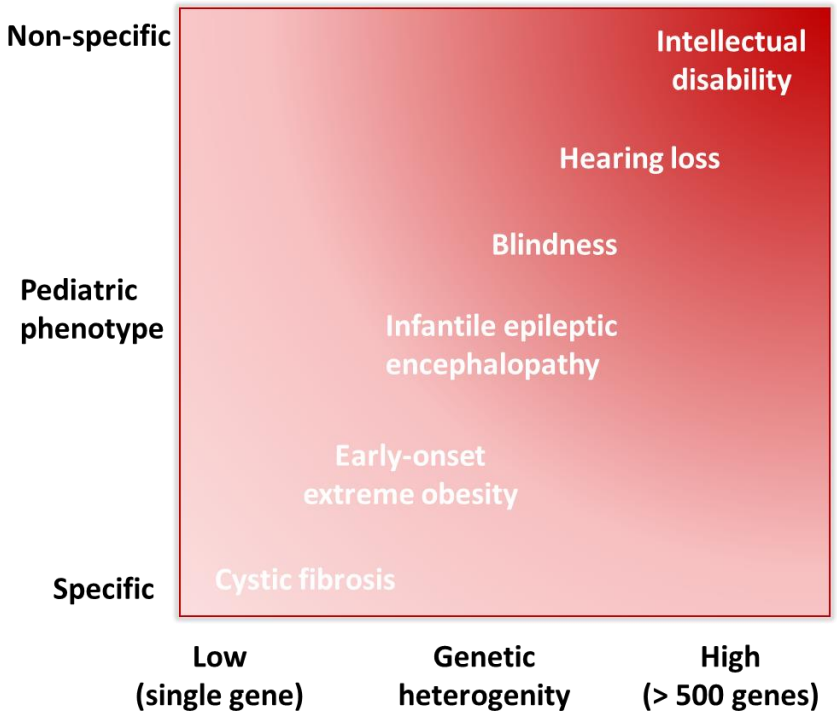


Figure 1.3: Genetic heterogeneity versus phenotypic specificity: The less specific the phenotype associated with a disease is, the more likely it is to be caused by variants in a large number of individual genes [adapted with permission from (Wright, FitzPatrick, and Firth 2018)]:

Idiopathic generalized epilepsies (IGEs), also called genetic generalized epilepsies (GGEs), constitute one of the most common group of epilepsies, accounting for approximately 1/3 of all epilepsies (Helbig et al. 2009). These epilepsies are characterized by generalized spike-wave activity on EEG and combinations of typical absence seizures, myoclonic seizures, and generalized tonic-clonic seizures (Helbig et al. 2008), without an attributable medical or traumatic cause. Among them there are several syndromes, including CAE, Juvenile Absence Epilepsy (JAE), Juvenile Myoclonic Epilepsy (JME) and Generalized Tonic–Clonic Seizures Alone (GTCSA) (Hildebrand et al. 2013; Falco-Walter, Scheffer, and Fisher 2018). IGEs present a complex pattern of inheritance and, to date, only a small fraction of the susceptibility genes has been identified (Hildebrand et al. 2013; Sirven 2015). Table 1.2 summarizes some genes associated with IGEs (data retrieved from (Helbig et al. 2008; Spillane, Kullmann, and Hanna 2016), and form ILAE database).

Table 1.2 – Genes associated with generalized epilepsies

Phenotypes	Protein family	Gene
IGE	Na+/K+ -ATPase	*ATP1A2*
Episodic ataxia and CAE	Calcium channels	*CACNA1A*
CAE	Calcium channels	*CACNA1H*
JME	EF-hand calcium-binding protein	*EFHC1*
CAE/IGE/JME/Infantile spasms, Lennox-Gastaut	GABA receptors	*GABRA1*
Infantile spasms, Lennox-Gastaut	GABA receptors	*GABRB2*
AE/Infantile spasms, Lennox-Gastaut	GABA receptors	*GABRB3*
GEFS+/JME	GABA receptors	*GABRD*
GEFS+/CAE/SMEI/IGE	GABA receptors	*GABRG2*
Infantile spasms	Glutamate receptors	*GRIN1*
JME/IGE/CAE	Glutamate receptors	*GRM4*
IGE	Glutamate receptors	*GRM7*
IGE	Glutamate receptors	*GRM8*
Autosomal dominant familial nocturnal frontal lobe epilepsy	Nicotinic acetylcholine receptors	*CHRNB2*
Autosomal dominant familial nocturnal frontal lobe epilepsy	Nicotinic acetylcholine receptors	*CHRNA2*
Autosomal dominant familial nocturnal frontal lobe epilepsy	Nicotinic acetylcholine receptors	*CHRNA4*
Epilepsy with episodic ataxia	Potassium channels	*KCNA1*
Myoclonic epilepsy and ataxia	Potassium channels	*KCNA2*
Generalised epilepsy with paroxysmal movement disorder	Potassium channels	*KCNMA1*
MPSI	Potassium channels	*KCNT1*
Developmental delay, epilepsy and neonatal diabetes mellitus (DEND syndrome)	Potassium channels	*KCNJI*
IGE/BFNS/IE	Potassium channels	*KCNQ2*
IGE/BFNS	Potassium channels	*KCNQ3*
IAE	Potassium channels	*KCNK9 (TASK-3)*
IGE/SMEI/IEGTC/GEFS+/MPSI	Sodium channels	*SCN1A*
IGE/SMEI/GEFS+	Sodium channels	*SCN1B*
IGE/IGE with GTCS/SMEI/West/GEFS+/Infantile spasms/Ohtahara/BFNIS	Sodium channels	*SCN2A*
IGE	Sodium channels	*SCN2B*
Partial epilepsy	Sodium channels	*SCN3A*
IGE/IEE	Sodium channels	*SCN8A*
IGE	Dopamine transporter	*SLC6A3*

Legend: IGE – idiopathic generalized epilepsy; CAE – child absence epilepsy; JME – juvenile myoclonic epilepsy; AE – absence epilepsy; GEFS+ – generalized epilepsy with febrile seizures plus; SMEI – severe myoclonic epilepsy of infancy; MPSI – migrating partial seizures of infancy; BFNS - benign familial neonatal seizures; IAE – idiopathic absence epilepsy; IEGTC – intractable epilepsy with generalized tonic-clonic seizures; GTCS – generalized tonic-clonic seizures; BFNIS – benign familial neonatal-infantile seizures.

The currently known genetic causes of NDs implicate numerous cell biological pathways, critical for normal brain development (Hu, Chahrour, and Walsh 2014), which can be grouped according to their function in several cellular aspects:

1. Transcriptional factors

Transcriptional regulation is a key point of neuronal differentiation programs; therefore, it is not surprising that genes involved in transcription regulation, such as *TCF4,* have been implicated in well-known NDs. The *TCF4* gene encodes a basic helix-loop-helix transcription factor that, when mutated, cause Pitt-Hopkins syndrome, characterized by significant global DD with moderate to severe ID, in which seizures are also a common feature (Zweier et al. 2007).

2. Chromatin modifiers/chromatin remodeling proteins

Chromatin regulation helps to orchestrate transcriptional programs underlying the maturation of developing neurons and the plasticity of adult neurons (Gallegos et al. 2018). Mutations in genes involved in chromatin regulation have hence been described in association with NDs and, in particular, with ID (Kleefstra et al. 2014). The chromodomain helicase DNA-binding (CHD) family of proteins are ATP-dependent chromatin remodelers that contribute to the reorganization of chromatin structure and deposition of histone variants necessary to regulate gene expression. There are nine CHD proteins identified in humans (CHD1–CHD9), the majority being ubiquitously expressed in human tissues, with only CHD5 expression being largely confined to neurons (Lamar and Carvill 2018). In mice, altered chromatin due to loss of the chromatin remodeler Chd5 causes a premature activation of neural stem cell, which is accompanied by transcriptional de-repression of ribosomal subunits, enhanced ribosome biogenesis, and increased translation and ultimately deregulate cell fate decisions, culminating in the generation of excessive numbers of astrocytes at the expense of neurons (Hwang et al. 2018). In man, pathogenic variants in *CHD1*, *CHD2*, *CHD4*, *CHD7* and *CHD8* genes have been associated with a range of neurological phenotypes, including ASD, ID and epilepsy [reviewed in (Lamar and Carvill 2018)]. ID and epilepsy are also present in the syndromes caused by pathogenic variants affecting the *EHMT1*, *CREBBP* and *MECP2* genes. *EHMT1* encodes a histone methyltransferase and when mutated cause Kleefstra syndrome (Kleefstra et al. 2006), while *CREBBP* encodes a histone acetyltransferase and cause Rubinstein-

Taybi syndrome when mutated (Petrij et al. 1995). As for the *MECP2* gene, it encodes a chromatin-associated protein that can both activate and repress transcription and is required for neuron maturation; mutations in *MECP2* gene cause Rett syndrome, characterized, among other features, by arrested development and regression of acquired skills between 6 and 18 months of age, seizures and ID (Swanberg et al. 2009).

3. Cell cycle regulators

The *NDE1* gene encodes a multidomain protein (nuclear distribution E (NudE) neurodevelopment protein 1) required for centrosome duplication and formation and function of the mitotic spindle, being essential for the development of the cerebral cortex (Bakircioglu et al. 2011). In fact, *NDE1* deficiency causes both a severe failure of neurogenesis and a deficiency in cortical lamination, as illustrated by patients with a severe microlissencephaly syndrome caused by homozygous *NDE1* frameshift mutations that truncate the C-terminal domain of the NDE1 protein, preventing cells to progress through the G2/M phase of mitosis (Alkuraya et al. 2011). Moreover, *NDE1* gene has previously been implicated in SZ through both genetic evidence [rare heterozygous missense *NDE1* variants have been described in association with SZ susceptibility (Kimura et al. 2015)] and through interaction of the NDE1 protein with the known SZ risk factor Disrupted in Schizophrenia 1 (DISC1) scaffold protein, which in turn modulates the function of NDE1 protein [reviewed by (Bradshaw 2016)]. Moreover, this gene lies within the consensus region of the 16p13.11 rearrangements, where deletions have been associated with ID, microcephaly and epilepsy, and particularly GGE, duplications with behavioural problems in addition to ID and/or congenital anomalies, and both deletions and duplications with SZ and ADHD [reviewed by (Torres, Barbosa, and Maciel 2016)].

4. Ubiquitin signaling

Ubiquitin-mediated degradation of proteins is a crucial mechanism for cell maintenance and viability (Clague, Coulson, and Urbé 2012). The ubiquitin signaling system is crucial for neuronal biology, as ubiquitin controls diverse cellular processes including cell fate determination, cell survival, neurite outgrowth and morphogenesis, synapse development, and synaptic function (Tai and Schuman 2008; Kowalski and Juo 2012). Several genes belonging to this pathway are also described to be associated with NDs, as is the case of *CUL4B* and

HUWE1, both encoding E3 ubiquitin ligases and both associated with syndromic X-linked ID (Tarpey et al. 2007; Froyen et al. 2008), or the *UBE3A* gene, encoding also a E3A ubiquitin ligase, that causes Angelman syndrome when mutated (Kishino, Lalande, and Wagstaff 1997).

5. Cytoskeleton regulation and organization, cell shape and motility

Cytoskeletal rearrangements are essential for every aspect of neurodevelopment, from the regulation of cell division and migration, to axon/dendrite formation, axonal pathfinding and the transport of cargo along those fibers (Hu, Chahrour, and Walsh 2014). In neurons, small guanosine triphosphatases (GTPases) of the RHO family are powerful initiators and modulators of structural changes. These members of the Ras superfamily function as molecular switches that are implicated in basic cellular processes at nearly all brain developmental steps, from neurogenesis and migration to axon guidance, linking extracellular cues to the neuronal responses required for the construction of neuronal networks, as well as for synaptic function and plasticity (Zamboni et al. 2018). The mis-regulation of the activity of RHO GTPases, such as RHOA, RAC1/RAC3 and CDC42, as well as of other proteins that interact with RHO GTPases, such as the p21-activated kinase, encoded by the *PAK1* gene, has been linked with ID and other neurodevelopmental conditions that comprise ID (Harms et al. 2018). In fact, a wide spectrum of structural brain abnormalities are caused by mutations in genes associated with cytoskeleton regulation and organization and frequently manifest with ID, epilepsy, and/or ASD (Stouffer, Golden, and Francis 2016), being another example the mutations affecting the beta-actin coding gene, *ACTB* (Cuvertino et al. 2017).

6. Intracellular vesicular trafficking and exocytosis

The proper function of the exocytotic machinery is crucial for synaptic membrane fusion and neurotransmitter release (Quick 2006). Syntaxin-binding protein 1 (STXBP1) participates in this process and is essential for synaptic vesicle exocytosis, most probably functioning as a chaperone to syntaxin-1, enabling it to stabilize normally and assisting in trafficking to the plasma membrane (Yamashita et al. 2016). *De novo* mutations of its encoding gene, *STXBP1*, are among the most frequent causes of epilepsy and encephalopathy, most patients presenting also severe to profound ID (Stamberger et al. 2016).

7. Signaling mediators/ transducers/ receptor activity/ transmembrane proteins

Signaling processes are essential for proper cellular function and usually implicate enzymes, transmembrane proteins and voltage ion-channels whose disruption may be associated with disease (van Bokhoven 2011). Ion-channel proteins, for example, have a crucial contribution for proper cellular signaling and synapse development and function: inherited disorders affecting ion channel function, generally called genetic channelopathies, can cause many different neurological diseases, from epileptic encephalopathies, such as Dravet syndrome, caused by mutations of *SCN1A* gene, to different types of ataxic syndromes (Spillane, Kullmann, and Hanna 2016).

8. Solute carriers/transporters

Thyroid hormones [thyroxine (T4), and 3,5,3'-triiodothyronine (T3)] play an essential role in the development of mammalian brain, by acting on migration and differentiation of neural cells, synaptogenesis, and myelination (Bernal 2007), and later on in the regulation of neuropsychological function in children and adults (Williams 2008). Their action is mediated through nuclear thyroid hormone (TH) transporters and regulation of gene expression (Bernal 2007). The key physiological role for these transporters was confirmed in patients with mutations in *SLC16A2* gene; this gene, located on chromosome Xq13.2, encodes the monocarboxylate transporter 8 (MCT8) involved in both TH uptake and efflux across the cell membrane, being thus important for the regulation of local TH activity in the brain and for brain development. Affected boys with hemizygous mutations in *SLC16A2* gene (and thus with a defective MCT8 protein) have severe intellectual and motor disability, as well as altered concentrations of thyroid hormones (low serum T4 and high T3 levels) (Bernal, Guadaño-Ferraz, and Morte 2015; Groeneweg, Visser, and Visser 2017). Another example concerns the gene that encodes the major glucose transporter in the mammalian blood-brain barrier, the solute carrier family 2 member 1 (*SLC2A1*). Pathogenic variants in this gene cause GLUT1 deficiency, a neurologic disorder showing wide phenotypic variability. The disorder, is part of a spectrum of neurologic phenotypes: the most severe 'classic' phenotype comprises infantile-onset epileptic encephalopathy associated with delayed development, acquired microcephaly, motor incoordination, and spasticity (De Giorgis and Veggiotti 2013); the less severe

phenotype is associated with paroxysmal exercise-induced dystonia with or without seizures (Brockmann 2009). A correct diagnosis of GLUT1 deficiency is of utmost importance since a ketogenic diet often results in marked clinical improvement of the motor and seizure symptoms (Leen et al. 2010; De Giorgis and Veggiotti 2013).

9. Local translation at synapses

Synaptic plasticity, learning, and memory require high temporal and spatial control of gene expression (Hutten, Sharangdhar, and Kiebler 2014). The production of proteins from mRNAs localized at synapses ultimately controls the strength of synaptic transmission (Thomas et al. 2014). Therefore, the disruption of local translation at synapses result in aberrant signaling, physiology and morphology of neurons, and a shift in the balance of synaptic plasticity, being associated with various neuropsychiatric disorders characterized by (and leading to) changes in behavior, cognitive abilities and memory (Khlebodarova et al. 2018). The synthesis of many synaptic proteins is, thus, under local control, being mRNA binding proteins (RBPs), such as FMRP, key regulators of local RNA translation (Liu-Yesucevitz et al. 2011). FMRP is a translational repressor codified by the FMR1 gene, located at Xq27.3, essential for proper synaptic function. In fact, loss of FMR1 expression due to the expansion of a CGG trinucleotide repeat in the 50 upstream region of the gene causes fragile-X syndrome, the most prevalent genetic cause of ID and autism in humans. The phenotype can be explained, at least in part, by the excess of translation of transcripts normally regulated by FMRP and consequently by an overabundance of certain proteins, originating reduced synaptic strength due to AMPA receptor trafficking abnormalities (Garber, Visootsak, and Warren 2008). Another example comes from the mammalian target of rapamycin (mTOR) pathway, which regulates a variety of neuronal functions, including cell proliferation, survival, growth, and plasticity (Han and Sahin 2011). Several genes have been linked to this pathway, such as the tuberous sclerosis complexes 1 and 2 (*TSC1*, *TSC2*) genes, that codify the proteins hamartin and tuberin, respectively. These proteins form a complex that acts as negative regulator of mTOR signaling; loss of function of either *TSC1* or *TSC2* genes lead to hyperactivation of the mTOR pathway and to tuberous sclerosis, manifesting not only by the widespread development of non-malignant tumors in multiple organs, but also by neurological features, such as ID, epilepsy and autism (LaSarge and

Danzer 2014). Epilepsy is the most common neurological symptom, occurring in 80% to 90% of affected individuals over the course of their lifetimes and causing significant morbidity and mortality (Curatolo 2015).

Both epilepsy and ID, as well as autism, represent a major burden not only for the individual, but also his/her immediate family and their community, being frequently associated with stigma and discrimination (Maulik et al. 2011; Sirven 2015). Being so, the clarification of etiology may help to manage and to guide expectations, while providing information regarding prognosis or expected clinical course (treatment, symptom management, or surveillance for known complications) (Makela et al. 2009). In fact, an accurate diagnosis is fundamental to guide an adequate intervention, the main goal of medical practice. For individuals with a genetic disorder, a robust genetic diagnosis may unlock access to a wealth of information in the literature, providing advice on management and therapy, enabling accurate determination of risk to existing and future family members and enabling also access to disorder-specific support groups, thus reducing isolation for families affected by a rare genetic disorder (Wright, FitzPatrick, and Firth 2018). Moreover, it avoids unnecessary and redundant diagnostic tests, providing, eventually, some hope for treatment or cure in future (Makela et al. 2009). Nevertheless, the determination of the precise molecular cause (genotype) that explains the clinical features of a particular disease (phenotype) is not always straightforward, and may, in fact, be a considerable challenge, either due to an incomplete knowledge about a particular disease or because of the variability (both genetic and phenotypic) associated with certain conditions (Wright, FitzPatrick, and Firth 2018). Causative genetic variants can range from single base pair alterations (substitution, deletion or duplication), causing dysfunction of single genes, to structural variants and to altered copy numbers of an entire chromosome (aneuploidy) or genome (as in diploid/triploid mosaicism). Moreover, not only some of these variants are individually rare, but they are also associated to clinical variability in the penetrance of the disorder and in the expressivity of individual features within individuals with the same variant (Wright, FitzPatrick, and Firth 2018; Spielmann, Lupiáñez, and Mundlos 2018). Whichever the genetic lesion present, a proper evaluation of the patients, identification of etiology and establishment of prognosis should contribute to improvements in the well-being of

those affected by these conditions (Moeschler and Shevell 2014; Allen et al. 2013).

Types of genetic lesions associated with NDs and their detection

As referred previously, causative genetic variants can vary in size from a single base pair, i.e. single nucleotide variants (SNVs), to hundreds or thousands of base pairs, as observed in the rearrangements affecting chromosomal regions of even entire chromosomes. Structural variants (SVs) collectively designate deletions, duplications, inversions, insertions and translocations and constitute the majority of varying nucleotides among human genomes (Weischenfeldt et al. 2013). SVs comprise unbalanced rearrangements, also known as copy number variations (CNVs), that alter the diploid status of DNA by changing the copy number of chromosomes or chromosomal region, and balanced rearrangements, such as inversions, reciprocal translocations or copy-number-neutral insertions that do not result in loss or gain of genetic material (Weischenfeldt et al. 2013; Spielmann, Lupiáñez, and Mundlos 2018).

Traditionally, the detection of such chromosomal rearrangements has relied upon genome-wide cytogenetic tests, such as the G-banded microscopy based karyotype (typical resolution ~5Mb), used to detect chromosomal abnormalities such as common aneuploidies, like trisomy 13, 18 or 21, and other structural chromosomal abnormalities, providing that are microscopically visible (Strassberg, Fruhman, and Van Den Veyver 2011). Nevertheless, the diagnostic yield of G-banded karyotyping is low: although genomic imbalances as small as 3Mb can sometimes be detected, others in the 5–10Mb range are often missed, depending on the genomic region involved and/or conditions of the assay [reviewed by (Miller et al. 2010)].

On the other hand, the diagnosis of clinical conditions that are caused by just one or a few genes relied upon highly focused single-gene molecular tests, in which a particular gene is studied by Sanger sequencing or genotyping and/or by dosage assessment to detect exon-level deletions and duplications. However, for conditions presenting phenotypic and/or genetic heterogeneity, many hundreds of genes may need to be tested, rendering single-gene approaches impracticable (Wright, FitzPatrick, and Firth 2018). In this context, the development of both chromosomal microarray technology and, particularly in the past few years, of

Next-Generation Sequencing (NGS) technology, have rendered genomic-wide tests accessible options that have revolutionized clinical genetics (Wright, FitzPatrick, and Firth 2018). A summary of the genomic testing strategies is briefly presented in Figure 1.4.

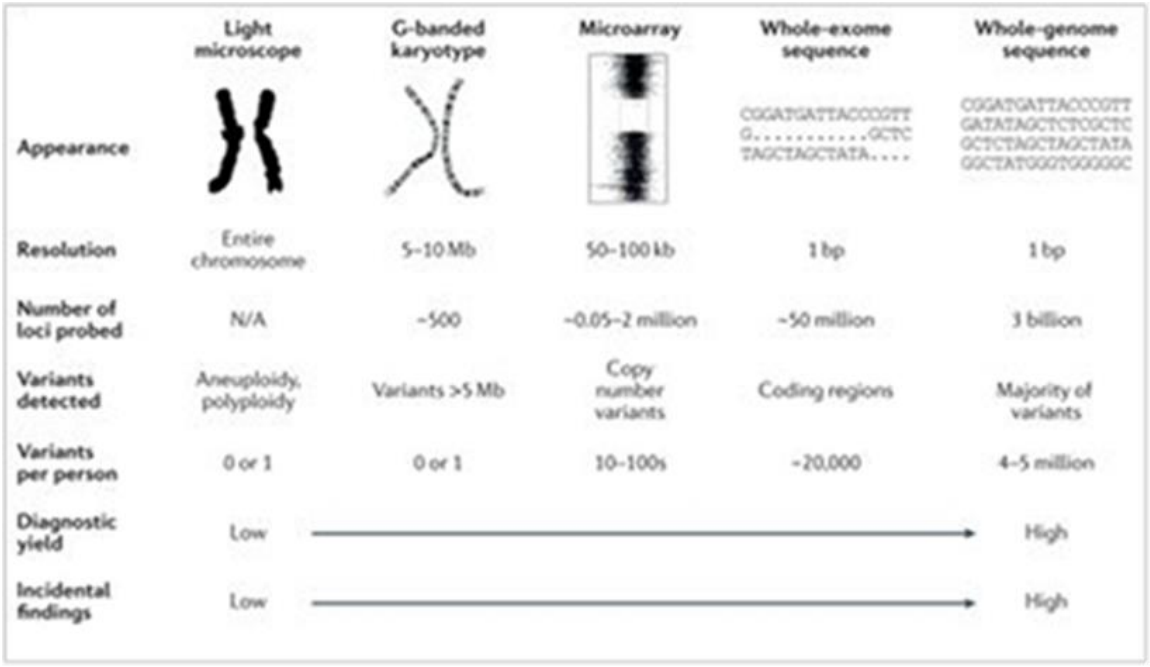


Figure1.4: Genomic testing strategies [adapted with permission from (Wright, FitzPatrick, and Firth 2018)].

Chromosomal abnormalities

Chromosome abnormalities contribute significantly to genetic disease resulting in reproductive loss, infertility, stillbirths, congenital anomalies, abnormal sexual development, ID and pathogenesis of malignancy. They are present in at least 50% of spontaneous abortions, in 6% of stillbirths, in about 5% of couples with two or more miscarriages and approximately in 0.5% of newborns (Luthardt and Keitges 2001). Common aneuploidies of chromosomes 13, 18 and 21 are among the most common chromosomal abnormalities. In fact, Down syndrome is the most common genetic cause of ID, with a prevalence of approximately 1 in 800 births, being the majority of cases the result of complete trisomy 21. Trisomy 13 and trisomy 18 are rarer, but together with trisomy 21, account for 89% of chromosome abnormalities with a severe phenotype identified in prenatal samples (Adams and Clark 2015; Mann, Petek, and Pertl 2015). Table 1.3 lists describes

the incidence and clinical characteristics ot the most common chromosomal syndromes.

Table 1.3 – Common autosomal and sex chromosome syndromes.

Syndrome	**Karyotype**	**Incidence**	**Clinical features**
Down	Trisomy 21	1/800 newborns	ID, growth retardation, characteristic facies and other abnormalities
Edwards	Trisomy 18	1/8000 live births	Failure to thrive, cardiac and kidney problems and other abnormalities, more than 90% die within the first 6 months
Patau	Trisomy 13	1/20000 live births	Central nervous system malformations, heart defects, growth retardation and numerous congenital anomalies, rarely survive the newborn period
Turner	45,X	1/8000 newborn females	Girls are typically short, sexually immature and infertile
Klinefelter	47,XXY	1/1000 newborn males	Generally normal in appearance before puberty; frequently ascertained in infertility clinics in adulthood, usually with tall stature; have a higher incidence of educational and emotional problems
Triple X	47,XXX	1/1000 newborn females	Clinically normal with normal gonadal function and fertility; show an increased risk for learning disabilities, reduction in performance IQ, menstrual problems and early menopause
47,XYY	47,XYY	1/1000 newborn males	Without discernible clinical features at birth or in infancy; tend to be taller than normal and have an increased tendency for behavioural and learning problems as children and young adults, usually with normal fertility

Copy number variations

Genomic rearrangements leading to the appearance of CNVs are frequent and widespread events, mostly as a consequence of the inherent repeat architecture of the human genome (Shaw and Lupski 2004; Malhotra and Sebat 2012). A CNV is a segment of DNA of 1 kb or larger that is present at a variable number of copies in different individuals from a population in comparison with a reference genome (Feuk, Carson, and Scherer 2006).

CNVs can be categorized into two major groups based on the breakpoint pattern analysis: (1) recurrent rearrangements, occurring in multiple unrelated individuals with clustering of breakpoints and sharing a rearrangement with a minimal overlapping region and (2) nonrecurrent rearrangements, with variable breakpoints (F Zhang et al. 2009). Rearrangements between different chromosomes or between different regions of the same chromosome are facilitated by the presence of region-specific low-copy repeats (LCRs) (Stankiewicz and Lupski 2002). LCRs

usually span approximately 10–400 kb of genomic DNA; may contain genes, gene fragments, pseudogenes, endogenous retroviral sequences or other paralogous fragments and provide the perfect substrates for homologous recombination, as they share more than 97% of sequence identity (Stankiewicz and Lupski 2002). There are five major mechanisms leading to CNV formation: (1) nonallelic homologous recombination (NAHR); (2) nonhomologous end joining (NHEJ); (3) fork stalling and template switching (FoSTeS); (4) microhomology-mediated break-induced replication (MMBIR) and (5) L1-mediated retrotransposition (Malhotra and Sebat 2012; Hastings, Ira, and Lupski 2009). Recurrent interstitial rearrangements are often flanked by LCRs, indicating that NAHR is a major causative mechanism for these structural variations. A summary of the five mechanisms accounting for CNV formation is presented in Figure 1.5.

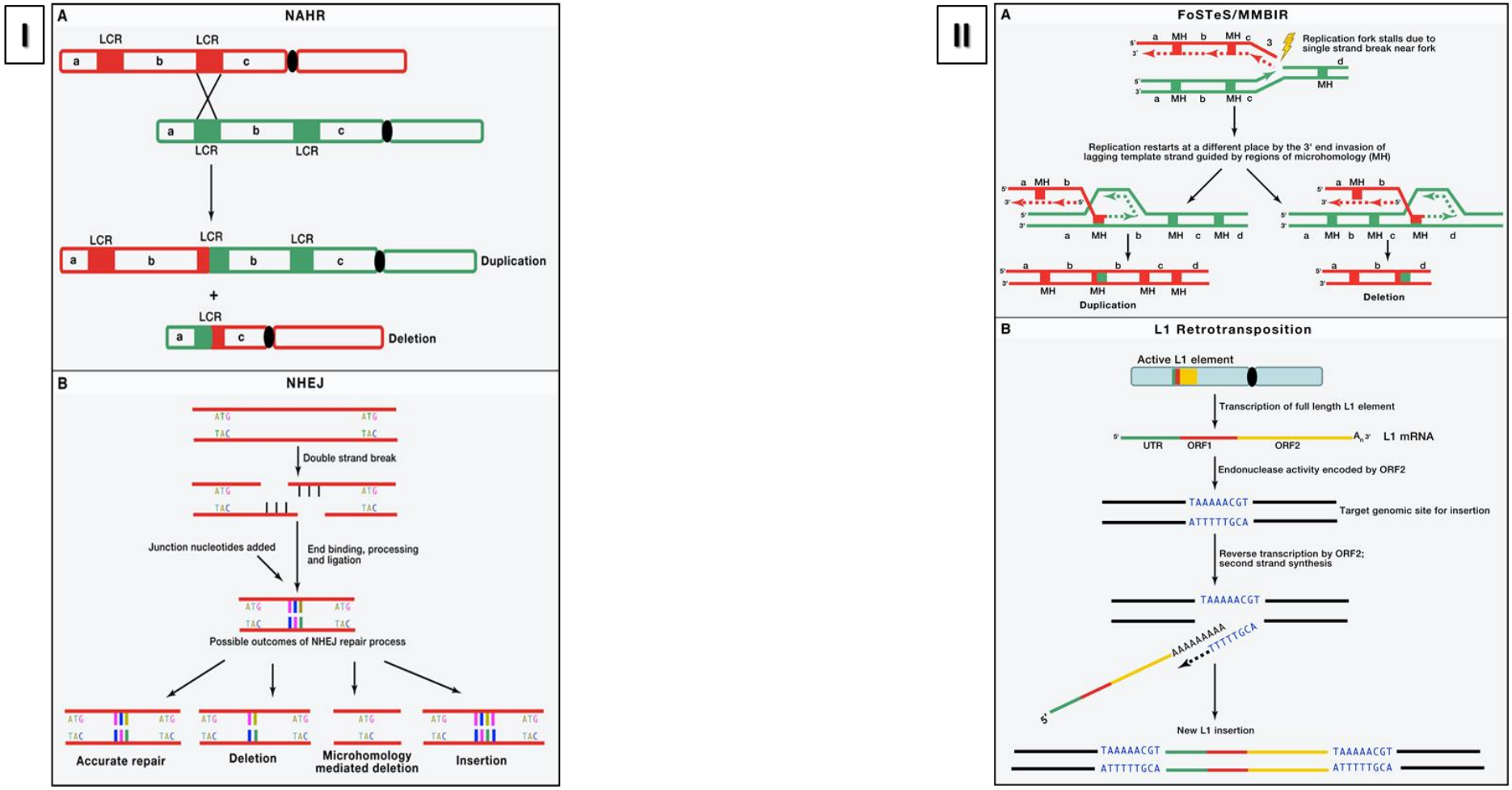


Figure 1.5: Major mechanisms of CNV formation: [adapted with permission from (Malhotra and Sebat 2012)].

Section I: Nonallelic homologous recombination (NAHR) and nonhomologous end-joining (NHEJ). (A) NAHR occurs by unequal crossing over between flanking segmental duplications (SDs, represented by two red and two green bars on respective homologous chromosomes), which results in reciprocal deletion and duplication of intervening sequence (b). These homologous chromosomes segregate from each other at the next cell division, thus leading to a

change in copy number in both daughter cells. (B) In classical NHEJ double-strand break repair pathways, the ends of DNA double-strand breaks are repaired through many rounds of enzymatic activity (including tethering of DNA ends by the Ku protein, followed by recruitment of DNA-dependent protein kinase DNA-PKcs by Ku and DNA PKcs-mediated activation of the Artemis nuclease, which trims back overhangs in preparation for ligation). The different types of DNA double-strand breaks fixed by NHEJ combined with other alternate repair mechanisms, including microhomology mediated end-joining (MMEJ), leads to diverse repaired products. Although limited base pairing can guide accurate repair, deletions of variable size and, to a lesser extent, insertions are often formed.

Section II: Fork Stalling and Template switching (FoSTes)/Microhomology Mediated Break-Induced Replication (MMBIR) and Retrotransposition.

(A) A simple model of FoSTeS/MMBIR is described. When a replication fork encounters a nick (striking arrowhead) in a template strand, one arm of the fork breaks off, producing a collapsed fork. At the single double-strand end, the 5' end of the lagging strand (dashed red lines) is resected, giving a 3' overhang. The 3' single-strand end of lagging-strand template (solid red lines) invades the sister leading-strand DNA (green lines) guided by regions of microhomology (MH, red and green boxes), forming a new low-processivity replication fork. The extended end dissociates repeatedly (due to migration of holiday junction or some other helicase activity) with 5' ends resected and reforms the fork. Whether the template switch occurs in front of or behind the position of the original collapse determines whether there is a deletion or duplication respectively. The 3' end invasion of lagging-strand template can reform replication forks on different genomic templates (>100 kb apart) before returning to the original sister chromatid and forming a processive replication fork that completes replication. Thus, the final product usually contains sequence from different genomic regions (not shown). Each line represents a DNA nucleotide strand. Polarity is indicated by arrows on the 3' end. New DNA synthesis is shown by dashed lines.

(B) LINE-1 retrotransposition. A full-length L1 (red, green, and orange bar on gray chromosome) is transcribed, and translation of ORF1 (red) and ORF2 (orange) protein encoded by the L1 messenger RNA (mRNA) leads to ribonucleoprotein (RNP) formation. L1 RNP is transported to the nucleus, and retrotransposition occurs by target site primed reverse transcription (TPRT). During TPRT, the L1 endonuclease (EN) activity of ORF2 nicks target genomic DNA (black lines), exposing a free 3'-OH that serves as a primer for reverse transcription of the L1 RNA. The mechanistic details of target site second-strand cleavage, second-strand complementary DNA (cDNA) synthesis, and completion of L1 integration require further elucidation. TPRT results in the insertion of a new, often 5'-truncated L1 copy at a newgenomic location that generally is flanked by target site duplications. Alu, SINE-R/VNTR/Alu (SVA), and cellular mRNAs can also hijack the L1-encoded protein(s) in the cytoplasm to mediate their transmobilization.

ORF: open reading frame;

CNVs as risk factors for neurodevelopmental disorders

Recurrent CNVs with common breakpoints may define clinically consistent and recognizable genomic disorders, due to the fact that a particular group of genes is consistently deleted or duplicated and is associated with a particular phenotype (Watson et al. 2014). Several syndromes are currently known to be caused either by deletion or duplication of specific chromosomal regions, many of which are associated with NDs. Chromosomal regions 1q21.1, 3q29, 15q11.2, 15q13.3, 16p11.2, 16p13.1 and 22q11 harbour rare but recurrent CNVs that have been uncovered as being important risk factors for several of these disorders. A significant proportion of risk for ID, ASD, SZ, epilepsy, bipolar disease (BD) and ADHD can be explained by these rare variants, which have been associated with odds ratio (OR) of disease greater than 2, and in some cases considerably larger [reviewed in (Torres, Barbosa, and Maciel 2016)].

As for epilepsy, microdeletions at 15q13.3 were the first to be described in association with common epilepsies, being now recognized as a major risk factor (OR >50) for IGE (Helbig et al. 2009). After this first publication, other groups have confirmed not only this association, but also the involvement of other CNVs in epileptogenesis, namely microdeletions at 15q11.2 and 16p13.11, strengthening the evidence that recurrent microdeletions at 15q11.2, 15q13.3 and 16p13.11 confer a pleiotropic susceptibility effect to a broad range of NDs, including IGE (De Kovel et al. 2010).

CNVs can lead to disease by means of one of the following mechanisms: gene dosage effects (either haploinsufficiency or triplosensitivity), gene disruption, gene fusion and other effects on gene function, including the disruption of regulatory elements such as enhancers and boundaries of topologically associated domains (TADs) and/or unmasking of recessive mutations in the unaffected allele (Feng Zhang and Lupski 2015; Middelkamp et al. 2017), reviewed in (Torres, Lopes, and Maciel 2018). Evidence obtained during the past few years supports the concept that CNVs in noncoding regions can also be pathogenic by a position effect mechanism (Feng Zhang and Lupski 2015). In fact, it was recently demonstrated that CNVs affecting noncoding regulatory elements are a major cause of congenital limb malformations. For example, CNVs at 7q36 (dup), 2q35-q36.1 (del), 7q21.3 (del) and 10q24.31-q24.32 (dup) which do not include any limb development – related genes were shown to affect known limb genes by a position

effect: this mutational mechanism implies the disruption of a TAD boundary, thereby allowing enhancers from neighboring domains to ectopically activate genes to cause misexpression and consequently disease (Flöttmann et al. 2017). Figure 1.6 schematizes the functional consequences of some SVs, including CNVs.

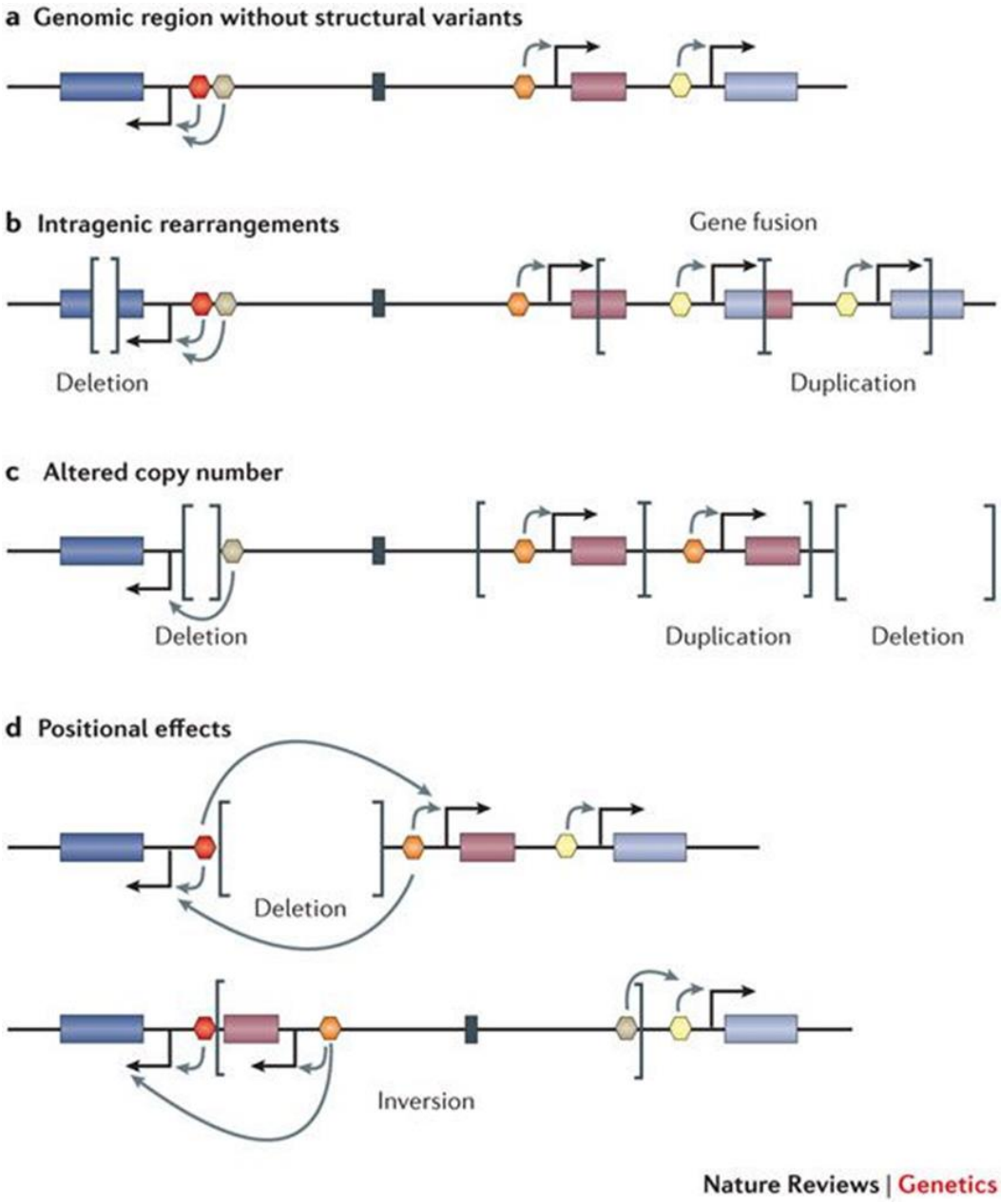


Figure 1.6: Schematic representation of the functional consequences of some structural variants [adapted with permission from (Weischenfeldt et al. 2013)]: **a.** Genes (represented by boxes) are regulated by the collective and combinatorial input of regulatory elements, including tissue-specific enhancers (hexagons, with different colors according tissue-specificity, and arrows pointing to the target gene) and insulators (enhancer–blockers, represented by black rectangles), which block the activity of regulatory elements; **b, c, d.** Structural variants (SVs),

represented by square brackets can have phenotypic consequences by altering coding regions; **b.** SVs can remove part of a coding region or fuse different coding regions after a duplication, resulting in aberrant transcripts; **c.** SVs such as deletions or duplications can lead to altered doses of otherwise functionally intact elements, resulting in altered regulatory input (left) or altered gene copy number (right) leading to either haploinsufficiency or triplosensitivity; **d.** SVs can also affect the expression of genes outside of the variants (through by a positional effect) thus resulting in a gain or loss of regulatory inputs.

The recurrent CNVs often exhibit variable expressivity and different penetrance between individuals with the same CNV, and normally underlie different phenotypes, ranging from normal development to disease (Watson et al. 2014; Torres, Barbosa, and Maciel 2016). Important sources of variability include genetic variants at one or more other *loci* (modifiers) and environmental factors, which are currently difficult to identify and quantify (Wright, FitzPatrick, and Firth 2018). This implies that: (1) the very same damaging CNV might impact differently in different individuals, depending on their genetic background; (2) one may be able to predict the phenotypical severity of an individual with a deleterious CNV by using parents as proxies. This knowledge could be useful in postnatal diagnosis, for tailored follow-up and early intervention, and in prenatal care. Complementary, detailed study of carriers of neuropsychiatric CNVs revealed that they show cognitive abilities and brain structure changes situated between those of controls (i.e., without a neuropsychiatric CNV) and patients with NDs, even though they often did not fulfill criteria for ASD, ID or SZ. One could speculate that what is sometimes perceived as lack of penetrance is actually variable expressivity, which could be detectable provided more granular phenotyping had been performed (Torres, Barbosa, and Maciel 2016).

Methodology overview

The rapid development in genetic tools and sequencing technology in the past few years had a particular impact in the understanding of the genetics of NDs (Hu, Chahrour, and Walsh 2014).

Advances in chromosomal microarray technology have allowed for the analysis of CNVs in very large case–control cohorts and highlighted the biological relevance of these structural variants (Malhotra and Sebat 2012). Array comparative genomic hybridization (aCGH) was the method of choice for many years in the study of CNVs associated with disease and genetic variability. In fact, the 15–20% diagnostic yield of chromosomal microarray led multiple medical entities, such as the American College of Medical Genetics and Genomics (ACMG), the American Academy of Neurology and the American Academy of Pediatrics, to recommend microarray as a first-tier clinical diagnostic test for individuals with ASD, ID/DD or multiple congenital anomalies (Miller et al. 2010). This contributed to: (A) a better understanding of the entire spectrum of fully penetrant genes and regions that cause syndromic NDs, the current understanding being that the spectrum fades into non-syndromic mild ID and ASD; (B) detection of CNVs that are significantly enriched in cases but also present in controls. In fact, the highly increased risk of developing neurodevelopmental phenotypes associated with some of these CNVs makes them an unavoidable element in the clinical context in pediatrics, neurology and psychiatry and should be addressed by a multidisciplinary clinical team, ideally including a geneticist (Torres, Barbosa, and Maciel 2016).

However, in the past decade, laboratories throughout the world started the transition from aCGH to massive parallel sequencing (MPS), as it allows the detection of a variety of genomic alterations at the same time (very small CNVs, structural anomalies such as inversions and translocations, etc.) as well as the estimation of breakpoints in a precise manner (Valsesia et al. 2013). This transition is the consequence of the great advances in sequencing technology that allowed an exponential drop in costs and increased the ability to capture high read depth to detect low-level mutations, hence expanding the sequencing coverage of the genome and enabling the capacity to identify variants that explain many Mendelian diseases in both known and new disease genes (Hu, Chahrour, and Walsh 2014).

Chromosomal microarray technology

Single nucleotide polymorphism (SNP) arrays

Single nucleotide polymorphisms (SNPs) comprise a major part of DNA variants and occur in a large proportion of the human population (>1%). Each individual inherits one allele copy from each parent, so that the individual genotype at an SNP site is AA, BB, or AB (Iacobucci et al. 2013). The Human Genome Project (Lander et al. 2001), the SNP Consortium (Sachidanandam et al. 2001), the International HapMap Project ("The International HapMap Project" 2003; Belmont et al. 2005), the 1000 Genomes Project (Abecasis et al. 2010) and the integrated map of genetic variation from 1092 human genomes (Abecasis et al. 2012) have collectively identified and catalogued approximately 15 million common single nucleotide variants, mostly SNPs, providing a collection of the levels of genetic variation in humans at both the inter-individual and inter-population levels (reviewed in (Iacobucci et al. 2013)). These projects contributed to (and had been driven by) an enormous development of SNP-arrays methodology that allowed the genotyping of thousands of SNPs across the genome simultaneously (LaFramboise 2009). Figure 1.7 provides an overview of the process used in two of the more common SNP-array platforms: Affymetrix and Illumina.

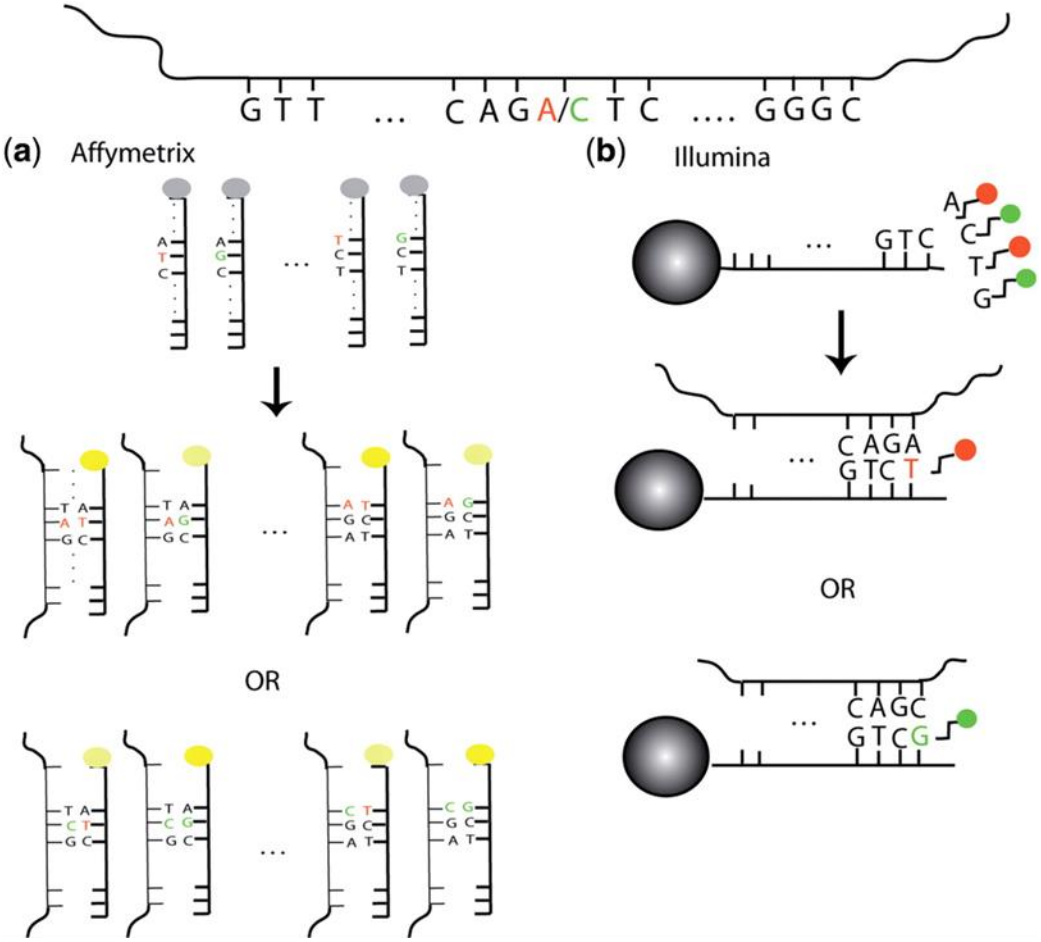


Figure 1.7: Overview of SNP array technology [reprinted with permission from (LaFramboise 2009)]. At the top is the fragment of DNA harboring an A/C SNP to be interrogated by the probes. (a) Affymetrix assay: there are 25-mer probes for both alleles, and the location of the SNP *locus* varies from probe to probe. The DNA binds to both probes regardless of the allele it carries, but it does so more efficiently when it is complementary to all 25 bases (bright yellow) rather than mismatching the SNP site (dimmer yellow). This impeded binding manifests itself in a dimmer signal. (b) Illumina assay: Attached to each Illumina bead is a 50-mer sequence complementary to the sequence adjacent to the SNP site. The single-base extension (T or G) that is complementary to the allele carried by the DNA (A or C, respectively) then binds and results in the appropriately-colored signal (red or green, respectively). For both platforms, the computational algorithms convert the raw signals into inferences regarding the presence or absence of each of the two alleles.

SNP-arrays were primarily developed with the aim of performing genome-wide linkage and association studies (GWAS), i.e., association studies between SNP genotype and disease in order to identify the variants that are statistically more prevalent in individuals with a disease than in individuals free of the disease, in

order to pinpoint the predisposition allele, as this can lead to genetic tests, unravel the disease biology and allow treatment options (van der Sijde, Ng, and Fu 2014). The platform's applications have then expanded from its original goal to include the detection and characterization of CNVs (whether somatic, inherited, or *de novo*), as well as of loss-of heterozygosity (LOH) in cancer cells. LOH, also known as uniparental disomy (UPD) or gene conversion, is quite common in some cancers (Simons et al. 2012) and, by definition, implies a change from a heterozygous state to a homozygous state. Since the genotype information obtained with SNP array analysis enables the detection of stretches of homozygosity, SNP-arrays are natural tools for LOH detection, and may allow the possible identification of recessive disease genes, mosaic aneuploidy, or UPD, when a patient-parents trio analysis is performed (LaFramboise 2009; De Leeuw et al. 2011).

Array comparative genomic hybridization (aCGH)

aCGH enables high-resolution, genome-wide screening of segmental CNVs, due to its higher detection rate (ranging from 12% to 19%) when compared with classic cytogenetic procedures (Riggs et al. 2014), among patients with NDs. The analysis is based in the comparison of equal amounts of DNA from a patient and from a healthy control (or frequently a pool of controls): both samples are labelled with different fluorochromes, usually Cyanine 3 (Cy3) and Cyanine 5 (Cy5), so that they can be distinguished, and then competitively co-hybridized in a slide containing the immobilized reference DNA fragments (probes) that represent the genome (Shinawi and Cheung 2008) (see figure 1.8).

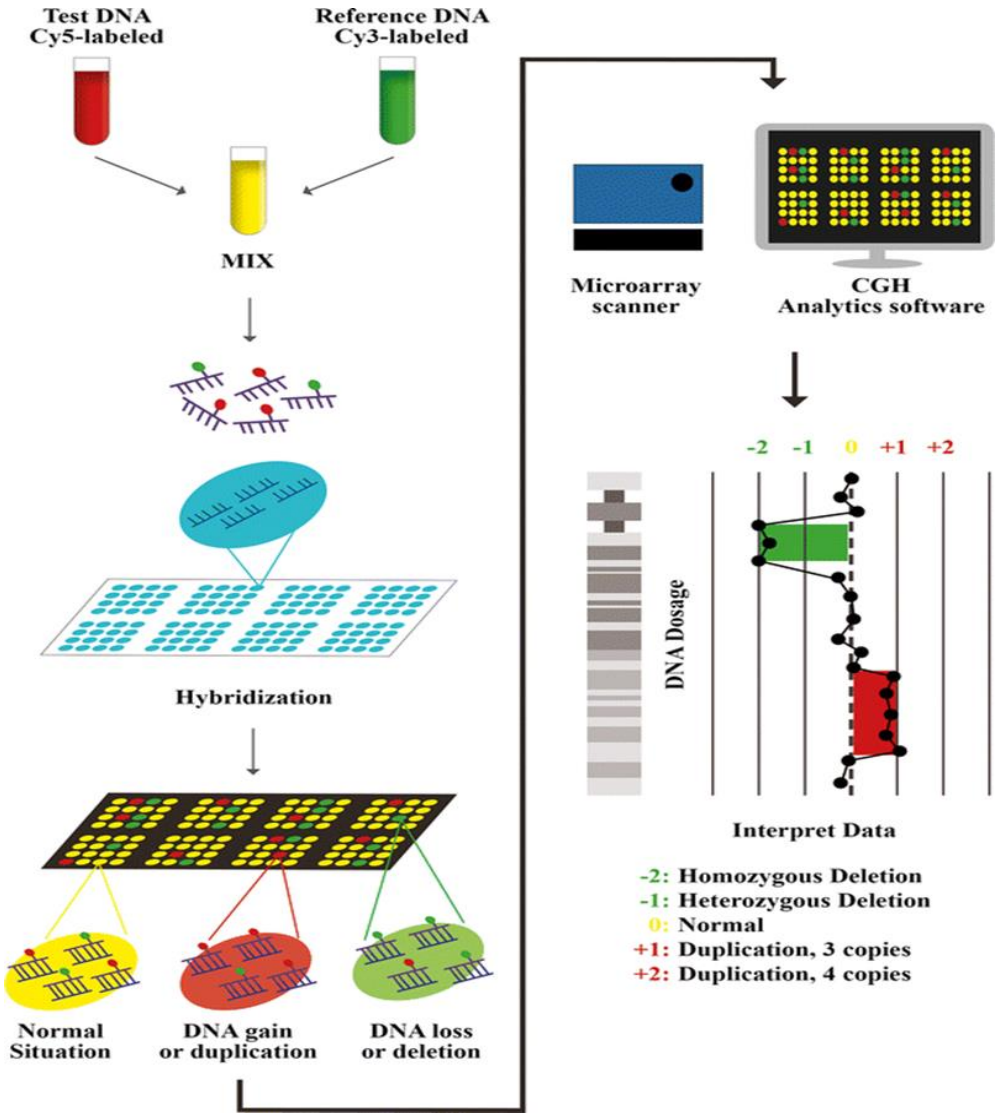


Figure 1.8: Schematic representation of aCGH technology [adapted with permission from (Colaianni, Mazzei, and Cavallaro 2016), according the terms of Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/)]. DNA samples from a patient/test and from a control/reference, are independently labeled with two different fluorophores of different colors (usually red- cyanine 5 and green-cyanine 3) and hybridized on an array containing thousands of known probes, arranged in a precise grid on chip. The microarray scanner detects the fluorescent signals on each probe and the array analytical software calculates the log2 ratio of fluorescence (Cy5/Cy3), allowing the identification of deletions or duplications in DNA. A higher intensity of the test sample color in a specific region of a chromosome versus the control indicates the gain of DNA of that region, while a higher intensity of the control sample color versus the test sample indicates the loss of material in that specific region. A neutral color (yellow when are used red and green fluorophores) indicates no difference between the two samples in that location (normal dose).

The sensitivity of aCGH is mainly influenced by probe coverage, resolution, and genomic spacing of probes selected for the array. While some clinical arrays in

use detect imbalances of 20–50 kb in targeted regions (for example, within known Mendelian genes) and imbalances of 100–250 kb in nontargeted (backbone) regions of the genome (Miller et al. 2010), whole-genome arrays provide an equally high resolution coverage of the entire genome (Vermeesch et al. 2012) allowing the detection of rare variants associated with rare syndromes (Riggs et al. 2014).

Nevertheless, some difficulties remain in the analysis and interpretation of the clinical significance of CNVs: (a) many of the CNVs detected by aCGH are rare and restricted to one patient/family; (b) many CNVs are associated with incomplete penetrance and/or expressivity; (c) due to the previous points, great challenges are faced when counselling for a recurrent CNV with known incomplete penetrance and/or variable expressivity [reviewed in (Torres, Barbosa, and Maciel 2016)], or when trying to address the clinical significance of a particular "private" familial CNV.

In order to facilitate the classification and thus the clinical interpretation of CNVs, several recommendations for systematic evaluation have been published (Miller et al. 2010; Leung, Pooh, and Wang 2010; Kearney et al. 2011).

Briefly, CNVs can be classified as follows:

- Pathogenic: when a high degree of certainty of their clinically significance is present in the literature. This group includes large CNVs that overlap significantly with a region with an established pathogenic effect (Kearney et al. 2011), and susceptibility *loci* of variable expressivity and incomplete penetrance [reviewed in (Torres, Barbosa, and Maciel 2016)].
- Likely pathogenic: include newly described gene-rich and usually large CNVs that comprise good candidates for disease association, or CNVs for which speculation of pathogenicity is supported in the literature (for example, there is another similar patient described or it includes gene(s) with a compelling function) (Kearney et al. 2011). However, the uncertainty of this claim stills remains, being possible that variants in this category will later be classified as pathogenic (with the report of other similar CNVs in patients with overlapping phenotype) or as benign (for example, if the variant is described later on in several unaffected individuals).
- Variants of unknown clinical significance (VOUS): include variants whose clinical significance is not yet possible to establish, due to the fact that: 1)

there is a lack of overlapping CNVs reported in the literature and/or databases; 2) the CNV contains genes but it is not yet known whether they are dosage sensitive; 3) the CNV is described in multiple contradictory publications and/or databases, and firm conclusions regarding clinical significance are not yet established (Kearney et al. 2011).

- Benign: includes CNVs that have been reported in multiple peer-reviewed publications or curated databases as a benign variant, and/or common polymorphisms (present in >1% of the population), except for established risk *loci* (Kearney et al. 2011; Leung, Pooh, and Wang 2010).

Despite its higher sensitivity, there are limitations in aCGH technology, namely its incapacity to identify balanced rearrangements such as translocations and inversions (Shinawi and Cheung 2008), to allow the differentiation of free trisomies from unbalanced Robertsonian translocations, or the detection of some marker chromosomes, which will depend on the size, marker composition, and array coverage of the specific chromosomal region present on the marker (Manning and Hudgins 2010). According to a retrospective study performed by Hochstenbach and colleagues in 36,325 patients with idiopathic ID/DD, in approximately 0.78% of the cases a balanced chromosomal rearrangement would have remained undetected by array-based investigation. These include familial rearrangements (0.48% of all referrals), *de novo* reciprocal translocations and inversions (0.23% of all referrals), *de novo* Robertsonian translocations (0.04% of all referrals), and 69,XXX triploidy (0.03% of all referrals) (Hochstenbach et al. 2009). In this case, the detection would only be possible by classical karyotyping.

Massive parallel sequencing (MPS)

MPS methods combines the technique developed in Sanger sequencing with array-based technologies to process millions of reactions in parallel (hence the designation massive parallel sequencing), resulting in high-throughput sequencing methodology (Xuan et al. 2013). Briefly, in MPS technologies, the DNA is fragmented into small pieces and then used to construct a library of fragments that have synthetic DNAs (adapters) added covalently to each fragment end by a DNA ligase. These adapters are universal sequences, specific to each platform, used to polymerase-amplify the library fragments; this amplification occurs *in situ* on a

solid surface, either a bead or a flat glass microfluidic channel, that is covalently derivatized with adapter sequences that are complementary to those on the library fragments (Mardis 2013). This allows the downstream sequencing reaction to operate as millions of micro-reactions carried out in parallel on each spatially distinct template (Rizzo and Buck 2012). Most sequencing platforms cannot monitor single-molecule reactions and template amplification is necessary to produce sufficient signal for detection from each of the DNA sequencing reaction steps in order to determine the sequencing data for that library fragment (Mardis 2013; Rizzo and Buck 2012) (figure 1.9).

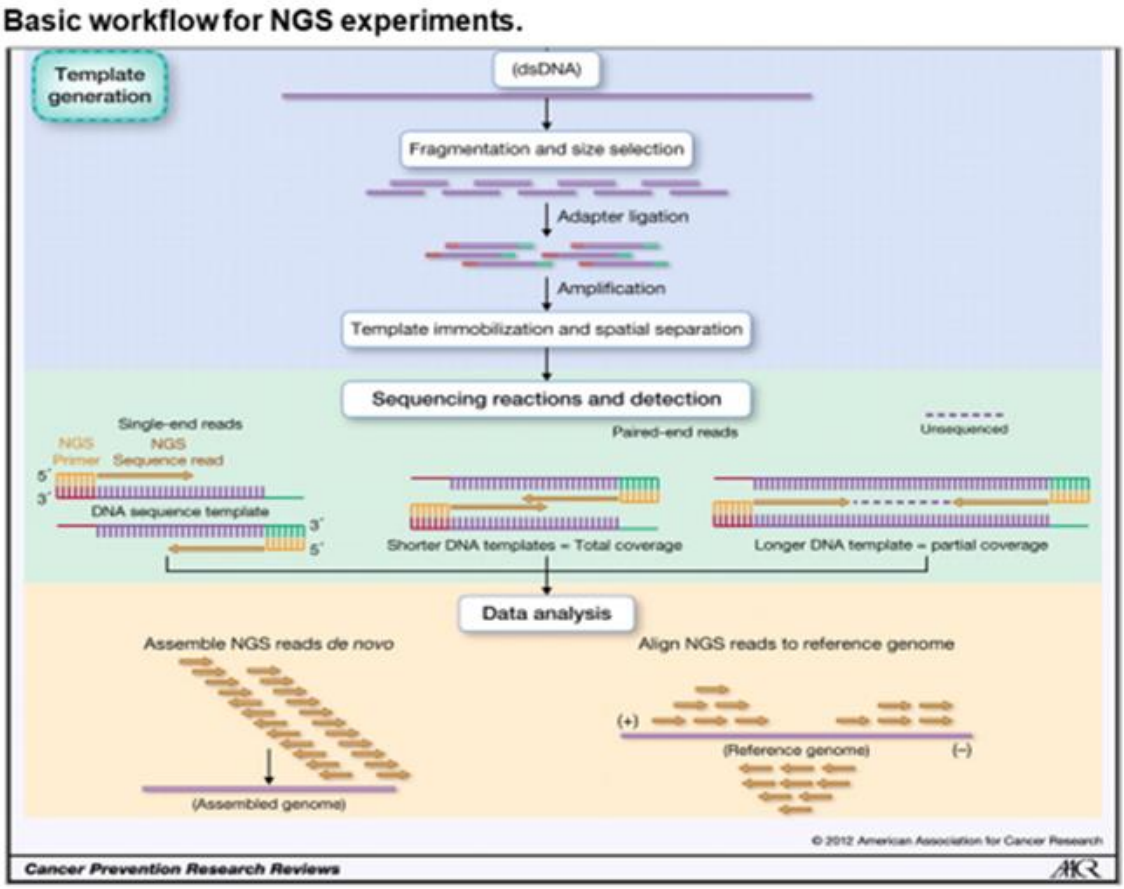


Figure 1.9: Schematic representation of a basic workflow for NGS experiments [adapted with permission from (Rizzo and Buck 2012)]. The main steps consist in template generation, sequencing reactions and data analysis.

The rapid evolution of this field in the last decade, in terms of speed, read length, and throughput, and namely throughout the development of several benchtop systems, has significantly decreased the cost of MPS compared to Sanger

sequencing; consequently, its applications have been democratized and, nowadays, MPS is currently used not only in the context of clinical genetics diagnosis and research (Gullapalli et al. 2012), but also in forensic sciences, microbiology or agricultural genomics (van Dijk et al. 2014)

MPS technologies are having a major impact in medical practice. In fact, most molecular diagnostic laboratories now offer NGS-based gene panels in which specific gene sets are sequenced (the particular set of genes being associated with a particular disorder or group of disorders), at a much lower cost than would be possible with Sanger sequencing; moreover, the popularization of the whole exome sequencing (WES) analysis has rendered it a tool currently used in clinical genetics practice, especially when searching for genetic variants responsible for clinical phenotypes previously without a known cause (Rauch et al. 2012; de Ligt et al. 2014).

MPS technologies generate an enormous amount of raw data requiring complex bioinformatic analyses to extract useful information. The initial analysis or base calling is typically conducted by proprietary software on the sequencing platform. After base calling, the sequencing data are aligned to a reference genome if available or a *de novo* assembly is conducted. After the alignment to a reference genome, the data is then analyzed in an experiment-specific fashion (Rizzo and Buck 2012). Data analysis is a critical feature and will depend on the goal and type of project: the analysis may focus on the entire genome (whole-genome sequencing analysis, WGS), on the whole exome (WES), on a selection of genes, transcriptomes or other selected parameters (Van El et al. 2013).

Similar to the interpretation of CNV significance, single nucleotide variants (SNVs) generated by MPS can also raise doubts in the interpretation and can be associated with variable expressivity. Therefore, the ACMG published a set of standards and guidelines for the interpretation of SNVs (detected by MPS or otherwise), which provide a pragmatic series of criteria and combination rules in order to classify a variant in one of the following categories: pathogenic, likely pathogenic, unknown significance, likely benign and benign (Richards et al. 2015). More recently, a joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists led to the publication of updated standards and guidelines for the interpretation and reporting of SNVs in Cancer (Li et al. 2017).

Moreover, one must keep in mind that such amount of data generated by MPS may reveal some incidental findings (IFs) that need to be handled properly. Incidental or secondary findings, relate to sequence variants not related to the indication for ordering the sequencing but that may be of medical value or utility to the ordering physician and the patient. The term incidental findings (IFs) is used "*to indicate the results of a deliberate search for pathogenic or likely pathogenic alterations in genes that are not apparently relevant to a diagnostic indication for which the sequencing test was ordered*" (Green et al. 2013). Being so, it is currently recommended by this college to report constitutional relevant variants detected in any of the 56 ACMG gene list, irrespective of the indication for which the clinical sequencing was ordered (Green et al. 2013). Recently, there was an update of the secondary findings minimum list in order to include 59 medically actionable genes (Kalia et al. 2017). Nevertheless, this policy for reporting of IFs solely taking into account the medical point of view is generally perceived as paternalistic (Ploug and Holm 2017). In fact, the European Society of Human Genetics (ESHG), despite its recommendation to report serious and actionable IFs, has stressed that an adequate policy should taking into account individuals' interests and, in general, patients should be able, like in every pre-symptomatic genetic test setting, to apply and change their preference regarding the disclosure of results (Van El et al. 2013; Saelaert et al. 2018).

In recent years, several methods have been developed for the identification of CNVs using exome-derived data. WES data are potentially useful for CNV detection because coding regions are enriched for causal genetic variants and the data may be easily analysed for this secondary purpose at a relatively low cost. Such tools used to identify CNVs using clinical grade WES data, are mainly based on detection of changes at the level of read-depth. Programs such as the eXome-Hidden Markov Model (XHMM) and the Copy Number Inference From Exome Reads (CoNIFER) detect rare CNVs based on a batched-comparison principle, while CNVnator detects CNVs based on changes in normalized read depth (Yao et al. 2017). However, comparisons made between these methods have revealed a lower sensitivity and uncertain specificity of WES-based CNV detection in comparison with aCGH-based CNV detection (Yao et al. 2017), indicating that methods for identifying exome-derived CNVs still need to be rigorously validated against independent standards (Kadalayil et al. 2014; Yao et al. 2017).

It is, however, expected that the development of higher accuracy methods will permit that a broad range of variants, from SNVs to large structural variants, could be assessed in a single genomic analysis, WES and/or WGS. In fact, a novel pipeline was recently developed that allowed the identification of deletion-CNVs from WES and WGS trio data, based on the clustering of Mendelian errors (Manheimer et al. 2018). The Mendelian Error Method (MEM) is based on the principle that the presence of a heterozygous deletion reduces the underlying genotype to a hemizygous state (McCarroll et al. 2006; Conrad et al. 2006) and, therefore, variants within heterozygous deletions frequently display Mendelian errors in the context of a trio design, as a result of a genotype mis-assignment (Manheimer et al. 2018). MEM could identify deletions with a positive predictive value (PPV) exceeding 90%, for both WES and WGS, identifying additional *de novo* deletions usually missed by other structural variants callers used for identify CNVs from WGS data, such as Lumpy and Structural variation and indel Analysis By Assembly (SvABA) (Manheimer et al. 2018).

Such genomic analyses could provide the basis for unique individual-directed healthcare and will be the basis of Precision Medicine, namely in the field of NDDs.

Aims of the work

The aims of this work were:

1) To unravel the genetic etiology of neurodevelopmental disease in a large series of patients of Portuguese origin;
2) To identify novel genes/*loci* involved in neurodevelopmental disorders;
3) To establish phenotype-genotype relationships and contribute to define novel genetic syndromes.

In order to achieve this purpose, aCGH was performed in a large group of patients with idiopathic ID and in a large family with epilepsy. WES analysis was also performed in selected patients from this group. Whenever necessary, quantitative real-time polymerase chain reaction (qPCR) and Sanger sequencing analysis were also used, either to confirm aCGH and WES findings, respectively, or to perform segregation analysis in family members, and qRT-PCR or RT-PCR plus sequencing were used to assess the functional impact of the chromosomal rearrangements.

Bibliography

Abecasis, Gonçalo R., David L. Altshuler, Adam Auton, Lisa D Brooks, Richard M. Durbin, Richard A. Gibbs, Matthew E. Hurles, and Gil A. McVean. 2010. “A Map of Human Genome Variation from Population-Scale Sequencing.” *Nature* 467 (7319):1061–73. https://doi.org/10.1038/nature09534.

Abecasis, Gonçalo R., Adam Auton, Lisa D Brooks, Mark A. DePristo, Richard M. Durbin, Robert E. Handsaker, Hyun Min Kang, Gabor T. Marth, and Gil A. McVean. 2012. “An Integrated Map of Genetic Variation from 1,092 Human Genomes.” *Nature* 491 (7422):56–65. https://doi.org/10.1038/nature11632.

Adams, Darius J., and David A. Clark. 2015. “Common Genetic and Epigenetic Syndromes.” *Pediatric Clinics of North America* 62 (2). Elsevier Inc:411–26. https://doi.org/10.1016/j.pcl.2014.11.005.

Alkuraya, Fowzan S., Xuyu Cai, Carina Emery, Ganeshwaran H. Mochida, Mohammed S. Al-Dosari, Jillian M. Felie, R. Sean Hill, et al. 2011. “Human Mutations in NDE1 Cause Extreme Microcephaly with Lissencephaly.” *American Journal of Human Genetics* 88:536–47. https://doi.org/10.1016/j.ajhg.2011.04.003.

Allen, Andrew S, Samuel F Berkovic, Patrick Cossette, Norman Delanty, Dennis Dlugos, Evan E Eichler, Michael P Epstein, et al. 2013. “De Novo Mutations in Epileptic Encephalopathies.” *Nature* 501 (7466). Nature Publishing Group:217–21. https://doi.org/10.1038/nature12439.

American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5 (5th Ed.).* Arlington, VA: American Psychiatric Publishing.

Baglio, Francesca, Monia Cabinio, Cristian Ricci, Gisella Baglio, Susanna Lipari, Ludovica Griffanti, Maria G. Preti, et al. 2014. “Abnormal Development of Sensory-Motor, Visual Temporal and Parahippocampal Cortex in Children with Learning Disabilities and Borderline Intellectual Functioning.” *Frontiers in Human Neuroscience* 8 (October):1–11. https://doi.org/10.3389/fnhum.2014.00806.

Bakircioglu, Mehmet, Ofélia P. Carvalho, Maryam Khurshid, James J. Cox, Beyhan Tuysuz, Tanyeri Barak, Saliha Yilmaz, et al. 2011. “The Essential Role of Centrosomal NDE1 in Human Cerebral Cortex Neurogenesis.” *American Journal of Human Genetics* 88 (5):523–35. https://doi.org/10.1016/j.ajhg.2011.03.019.

Belmont, John W., Andrew Boudreau, Suzanne M. Leal, Paul Hardenbol, Shiran Pasternak, David A. Wheeler, Thomas D. Willis, et al. 2005. “A Haplotype Map of the Human Genome.” *Nature* 437 (7063):1299–1320. https://doi.org/10.1038/nature04226.

Berg, Anne T., Samuel F. Berkovic, Martin J. Brodie, Jeffrey Buchhalter, J. Helen Cross, Walter Van Emde Boas, Jerome Engel, et al. 2010. “Revised Terminology and

Concepts for Organization of Seizures and Epilepsies: Report of the ILAE Commission on Classification and Terminology, 2005-2009." *Epilepsia* 51 (4):676–85. https://doi.org/10.1111/j.1528-1167.2010.02522.x.

Bernal, Juan. 2007. "Thyroid Hormone Receptors in Brain Development and Function." *Nature Clinical Practice Endocrinology & Metabolism* 3 (3):249–59. https://doi.org/10.1038/ncpendmet0424.

Bernal, Juan, Ana Guadaño-Ferraz, and Beatriz Morte. 2015. "Thyroid Hormone Transporters-Functions and Clinical Implications." *Nature Reviews Endocrinology* 11 (7):406–17. https://doi.org/10.1038/nrendo.2015.66.

Bokhoven, Hans van. 2011. "Genetic and Epigenetic Networks in Intellectual Disabilities." *Annual Review of Genetics* 45 (1):81–104. https://doi.org/10.1146/annurev-genet-110410-132512.

Braat, Sien, and R. Frank Kooy. 2015. "The GABA<inf>A</Inf> Receptor as a Therapeutic Target for Neurodevelopmental Disorders." *Neuron* 86 (5). Elsevier Inc.:1119–30. https://doi.org/10.1016/j.neuron.2015.03.042.

Bradshaw, N. J. 2016. "Cloning of the Promoter of NDE1, a Gene Implicated in Psychiatric and Neurodevelopmental Disorders through Copy Number Variation." *Neuroscience* 324. IBRO:262–70. https://doi.org/10.1016/j.neuroscience.2016.03.018.

Brockmann, Knut. 2009. "The Expanding Phenotype of GLUT1-Deficiency Syndrome." *Brain and Development* 31 (7). Elsevier B.V.:545–52. https://doi.org/10.1016/j.braindev.2009.02.008.

Clague, Michael J., Judy M. Coulson, and Sylvie Urbé. 2012. "Cellular Functions of the DUBs." *Journal of Cell Science* 1251 Clagu (Pt 2):277–86. https://doi.org/10.1242/jcs.090985.

Colaianni, Valeria, Rosalucia Mazzei, and Sebastiano Cavallaro. 2016. "Copy Number Variations and Stroke." *Neurological Sciences* 37 (12):1895–1904. https://doi.org/10.1007/s10072-016-2658-y.

Conrad, Donald F., T. Daniel Andrews, Nigel P. Carter, Matthew E. Hurles, and Jonathan K. Pritchard. 2006. "A High-Resolution Survey of Deletion Polymorphism in the Human Genome." *Nature Genetics* 38 (1):75–81. https://doi.org/10.1038/ng1697.

Curatolo, Paolo. 2015. "Mechanistic Target of Rapamycin (MTOR) in Tuberous Sclerosis Complex-Associated Epilepsy." *Pediatric Neurology* 52 (3). Elsevier Inc:281–89. https://doi.org/10.1016/j.pediatrneurol.2014.10.028.

Cuvertino, Sara, Helen M. Stuart, Kate E. Chandler, Neil A. Roberts, Ruth Armstrong, Laura Bernardini, Sanjeev Bhaskar, et al. 2017. "ACTB Loss-of-Function Mutations Result in a Pleiotropic Developmental Disorder." *American Journal of Human Genetics* 101 (6):1021–33. https://doi.org/10.1016/j.ajhg.2017.11.006.

Devinsky, Orrin, Annamaria Vezzani, Terence J O, Nathalie Jette, Ingrid E Scheffer, Marco de Curtis, and Piero Perucca. 2018. “Epilepsy.” *Nature Publishing Group* 4 (May). https://doi.org/10.1038/nrdp.2018.24.

Dijk, Erwin L. van, Hélène Auger, Yan Jaszczyszyn, and Claude Thermes. 2014. “Ten Years of Next-Generation Sequencing Technology.” *Trends in Genetics : TIG* 30 (9):418–26. https://doi.org/10.1016/j.tig.2014.07.001.

El, Carla G. Van, Martina C. Cornel, Pascal Borry, Ros J. Hastings, Florence Fellmann, Shirley V. Hodgson, Heidi C. Howard, et al. 2013. “Whole-Genome Sequencing in Health Care.” *European Journal of Human Genetics* 21 (6). Nature Publishing Group:580–84. https://doi.org/10.1038/ejhg.2013.46.

Falco-Walter, Jessica J., Ingrid E. Scheffer, and Robert S. Fisher. 2018. “The New Definition and Classification of Seizures and Epilepsy.” *Epilepsy Research* 139 (July 2017):73–79. https://doi.org/10.1016/j.eplepsyres.2017.11.015.

Feuk, Lars, Andrew R Carson, and Stephen W Scherer. 2006. “Structural Variation in the Human Genome.” *Nature Reviews Genetics* 7 (February):85–97. https://doi.org/10.1038/nrg1767.

Fisher, Robert S., Carlos Acevedo, Alexis Arzimanoglou, Alicia Bogacz, J. Helen Cross, Christian E. Elger, Jerome Engel, et al. 2014. “ILAE Official Report: A Practical Clinical Definition of Epilepsy.” *Epilepsia* 55 (4):475–82. https://doi.org/10.1111/epi.12550.

Fisher, Robert S., J. Helen Cross, Jacqueline A. French, Norimichi Higurashi, Edouard Hirsch, Floor E. Jansen, Lieven Lagae, et al. 2017. “Operational Classification of Seizure Types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology.” *Epilepsia* 58 (4):522–30. https://doi.org/10.1111/epi.13670.

Flöttmann, Ricarda, Bjørt K Kragesteen, Sinje Geuer, Magdalena Socha, Lila Allou, Anna Sowińska-Seidler, Laure Bosquillon de Jarcy, et al. 2017. “Noncoding Copy-Number Variations Are Associated with Congenital Limb Malformation.” *Genet Med* 00 (April):1–9. https://doi.org/10.1038/gim.2017.154.

Forsgren, Lars, E. Beghi, A. Õun, and M. Sillanpää. 2005. “The Epidemiology of Epilepsy in Europe - A Systematic Review.” *European Journal of Neurology* 12 (4):245–53. https://doi.org/10.1111/j.1468-1331.2004.00992.x.

Froyen, Guy, Mark Corbett, Joke Vandewalle, Irma Jarvela, Owen Lawrence, Cliff Meldrum, Marijke Bauters, et al. 2008. “Submicroscopic Duplications of the Hydroxysteroid Dehydrogenase HSD17B10 and the E3 Ubiquitin Ligase HUWE1 Are Associated with Mental Retardation.” *American Journal of Human Genetics* 82 (2):432–43. https://doi.org/10.1016/j.ajhg.2007.11.002.

Gallegos, David A., Urann Chan, Liang Fu Chen, and Anne E. West. 2018. “Chromatin Regulation of Neuronal Maturation and Plasticity.” *Trends in Neurosciences* 41 (5). Elsevier Ltd:311–24. https://doi.org/10.1016/j.tins.2018.02.009.

Garber, Kathryn B., Jeannie Visootsak, and Stephen T. Warren. 2008. “Fragile X Syndrome.” *European Journal of Human Genetics* 16 (6):666–72. https://doi.org/10.1038/ejhg.2008.61.

Giorgis, Valentina De, and Pierangelo Veggiotti. 2013. “GLUT1 Deficiency Syndrome 2013: Current State of the Art.” *Seizure* 22 (10). BEA Trading Ltd:803–11. https://doi.org/10.1016/j.seizure.2013.07.003.

Green, Robert C, Jonathan S Berg, Wayne W Grody, Robert L Nussbaum, Julianne M O Daniel, and E Kelly. 2013. “ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing.” *Genetics in Medicine* 15 (7):565–74. https://doi.org/10.1038/gim.2013.73.ACMG.

Groeneweg, Stefan, W. Edward Visser, and Theo J. Visser. 2017. “Disorder of Thyroid Hormone Transport into the Tissues.” *Best Practice and Research: Clinical Endocrinology and Metabolism* 31 (2). Elsevier Ltd:241–53. https://doi.org/10.1016/j.beem.2017.05.001.

Gullapalli, Rama R, Ketaki V Desai, Lucas Santana-Santos, Jeffrey A Kant, and Michael J Becich. 2012. “Next Generation Sequencing in Clinical Medicine: Challenges and Lessons for Pathology and Biomedical Informatics.” *Journal of Pathology Informatics* 3 (1):40. https://doi.org/10.4103/2153-3539.103013.

Han, Juliette M., and Mustafa Sahin. 2011. “TSC1/TSC2 Signaling in the CNS.” *FEBS Letters* 585 (7). Federation of European Biochemical Societies:973–80. https://doi.org/10.1016/j.febslet.2011.02.001.

Harms, Frederike L., Katja Kloth, Annette Bley, Jonas Denecke, René Santer, Davor Lessel, Maja Hempel, and Kerstin Kutsche. 2018. “Activating Mutations in PAK1, Encoding P21-Activated Kinase 1, Cause a Neurodevelopmental Disorder.” *American Journal of Human Genetics* 103 (4):579–91. https://doi.org/10.1016/j.ajhg.2018.09.005.

Hastings, P. J., Grzegorz Ira, and James R. Lupski. 2009. “A Microhomology-Mediated Break-Induced Replication Model for the Origin of Human Copy Number Variation.” *PLoS Genetics* 5 (1). https://doi.org/10.1371/journal.pgen.1000327.

Helbig, Ingo, Heather C Mefford, Andrew J Sharp, Michel Guipponi, Marco Fichera, Andre Franke, Hiltrud Muhle, et al. 2009. “15q13.3 Microdeletions Increase Risk of Idiopathic Generalized Epilepsy.” *Nat Genet* 41 (2):160–62. https://doi.org/10.1038/ng.292.15q13.3.

Helbig, Ingo, Ingrid E. Scheffer, John C. Mulley, and Samuel F. Berkovic. 2008.

"Navigating the Channels and beyond: Unravelling the Genetics of the Epilepsies." *The Lancet Neurology* 7 (3):231–45. https://doi.org/10.1016/S1474-4422(08)70039-5.

Heyne, Henrike O., Tarjinder Singh, Hannah Stamberger, Rami Abou Jamra, Hande Caglayan, Dana Craiu, Peter De Jonghe, et al. 2018. "De Novo Variants in Neurodevelopmental Disorders with Epilepsy." *Nature Genetics*, 1. https://doi.org/10.1038/s41588-018-0143-7.

Hildebrand, Michael S, Hans-Henrik M Dahl, John Anthony Damiano, Richard J H Smith, Ingrid E Scheffer, and Samuel F Berkovic. 2013. "Recent Advances in the Molecular Genetics of Epilepsy." *Journal of Medical Genetics* 50 (5):271–79. https://doi.org/10.1136/jmedgenet-2012-101448.

Hochstenbach, Ron, Ellen van Binsbergen, John Engelen, Aggie Nieuwint, Abeltje Polstra, Pino Poddighe, Claudia Ruivenkamp, Birgit Sikkema-Raddatz, Dominique Smeets, and Martin Poot. 2009. "Array Analysis and Karyotyping: Workflow Consequences Based on a Retrospective Study of 36,325 Patients with Idiopathic Developmental Delay in the Netherlands." *European Journal of Medical Genetics* 52 (4). Elsevier Masson SAS:161–69. https://doi.org/10.1016/j.ejmg.2009.03.015.

Hu, Wen F., Maria H. Chahrour, and Christopher A. Walsh. 2014. "The Diverse Genetic Landscape of Neurodevelopmental Disorders." *Annual Review of Genomics and Human Genetics* 15 (1):195–213. https://doi.org/10.1146/annurev-genom-090413-025600.

Huang, Jichong, Tingting Zhu, Yi Qu, and Dezhi Mu. 2016. "Prenatal, Perinatal and Neonatal Risk Factors for Intellectual Disability: A Systemic Review and Meta-Analysis." *Plos One* 11 (4):e0153655. https://doi.org/10.1371/journal.pone.0153655.

Hutten, Saskia, Tejaswini Sharangdhar, and Michael Kiebler. 2014. "Unmasking the Messenger." *RNA Biology* 11 (8):992–97. https://doi.org/10.4161/rna.32091.

Hwang, Dong-Woo, Anbalagan Jaganathan, Padmina Shrestha, Ying Jin, Nour El-Amine, Sidney H Wang, Molly Hammell, and Alea A Mills. 2018. "Chromatin-Mediated Translational Control Is Essential for Neural Cell Fate Specification." *Life Science Alliance* 1 (4):e201700016. https://doi.org/10.26508/lsa.201700016.

Iacobucci, Ilaria, Annalisa Lonetti, Cristina Papayannidis, and Giovanni Martinelli. 2013. "Use of Single Nucleotide Polymorphism Array Technology to Improve the Identification of Chromosomal Lesions in Leukemia." *Current Cancer Drug Targets* 13 (7):791–810. https://doi.org/55011 [pii].

Kadalayil, Latha, Sajjad Rafiq, Matthew J.J. Rose-Zerilli, Reuben J. Pengelly, Helen Parker, David Oscier, Jonathan C. Strefford, et al. 2014. "Exome Sequence Read Depth Methods for Identifying Copy Number Changes." *Briefings in Bioinformatics* 16 (3):380–92. https://doi.org/10.1093/bib/bbu027.

Kalia, Sarah S., Kathy Adelman, Sherri J. Bale, Wendy K. Chung, Christine Eng, James P. Evans, Gail E. Herman, et al. 2017. “Recommendations for Reporting of Secondary Findings in Clinical Exome and Genome Sequencing, 2016 Update (ACMG SF v2.0): A Policy Statement of the American College of Medical Genetics and Genomics.” *Genetics in Medicine* 19 (2). IOP Publishing:249–55. https://doi.org/10.1038/gim.2016.190.

Karnebeek, Clara D.M. Van, and Sylvia Stockler. 2012. “Treatable Inborn Errors of Metabolism Causing Intellectual Disability: A Systematic Literature Review.” *Molecular Genetics and Metabolism* 105 (3). Elsevier Inc.:368–81. https://doi.org/10.1016/j.ymgme.2011.11.191.

Kaufman, Liana, Muhammad Ayub, and John B. Vincent. 2010. “The Genetic Basis of Non-Syndromic Intellectual Disability: A Review.” *Journal of Neurodevelopmental Disorders* 2 (4):182–209. https://doi.org/10.1007/s11689-010-9055-2.

Kearney, Hutton M, Erik C Thorland, Kerry K Brown, Fabiola Quintero-Rivera, and Sarah T South. 2011. “American College of Medical Genetics Standards and Guidelines for Interpretation and Reporting of Postnatal Constitutional Copy Number Variants EVALUATION AND CLINICAL INTERPRETATION.” *Genetics in Medicine* 13 (7):680–85. https://doi.org/10.1097/GIM.0b013e3182217a3a.

Khlebodarova, T. M., V. V. Kogai, E. A. Trifonova, and V. A. Likhoshvai. 2018. “Dynamic Landscape of the Local Translation at Activated Synapses.” *Molecular Psychiatry* 23 (1). Nature Publishing Group:107–14. https://doi.org/10.1038/mp.2017.245.

Kimura, Hiroki, Daisuke Tsuboi, Chenyao Wang, Itaru Kushima, Takayoshi Koide, Masashi Ikeda, Yoshimi Iwayama, et al. 2015. “Identification of Rare, Single-Nucleotide Mutations in NDE1 and Their Contributions to Schizophrenia Susceptibility.” *Schizophrenia Bulletin* 41 (3):744–53. https://doi.org/10.1093/schbul/sbu147.

Kishino, Tatsuya, Marc Lalande, and Joseph Wagstaff. 1997. “UBE3A/E6-AP Mutations Cause Angelman Syndrome.” *Nature Genetics* 15 (1):70–73. https://doi.org/10.1038/ng0197-70.

Kleefstra, Tjitske, Han G Brunner, Jeanne Amiel, Astrid R Oudakker, Willy M Nillesen, Alex Magee, David Geneviève, et al. 2006. “Loss-of-Function Mutations in Euchromatin Histone Methyl Transferase 1 (EHMT1) Cause the 9q34 Subtelomeric Deletion Syndrome.” *American Journal of Human Genetics* 79 (79):370–77.

Kleefstra, Tjitske, Annette Schenck, Jamie M. Kramer, and Hans Van Bokhoven. 2014. “The Genetics of Cognitive Epigenetics.” *Neuropharmacology* 80 (January). Elsevier Ltd:83–94. https://doi.org/10.1016/j.neuropharm.2013.12.025.

Kovel, Carolien G F De, Holger Trucks, Ingo Helbig, Heather C Mefford, Carl Baker,

Costin Leu, Christian Kluck, et al. 2010. “Recurrent Microdeletions at 15q11.2 and 16p13.11 Predispose to Idiopathic Generalized Epilepsies.” *Brain* 133 (1):23–32. https://doi.org/10.1093/brain/awp262.

Kowalski, Jennifer R., and Peter Juo. 2012. “The Role of Deubiquitinating Enzymes in Synaptic Function and Nervous System Diseases.” *Neural Plasticity* 2012 (Figure 1). https://doi.org/10.1155/2012/892749.

LaFramboise, Thomas. 2009. “Single Nucleotide Polymorphism Arrays: A Decade of Biological, Computational and Technological Advances.” *Nucleic Acids Research* 37 (13):4181–93. https://doi.org/10.1093/nar/gkp552.

Lamar, Kay-Marie J., and Gemma L. Carvill. 2018. “Chromatin Remodeling Proteins in Epilepsy: Lessons From CHD2-Associated Epilepsy.” *Frontiers in Molecular Neuroscience* 11 (June):1–12. https://doi.org/10.3389/fnmol.2018.00208.

Lander, E S, L M Linton, B Birren, C Nusbaum, M C Zody, J Baldwin, K Devon, et al. 2001. “Initial Sequencing and Analysis of the Human Genome.” *Nature* 409 (6822):860–921. https://doi.org/10.1038/35057062.

LaSarge, Candi L., and Steve C. Danzer. 2014. “Mechanisms Regulating Neuronal Excitability and Seizure Development Following MTOR Pathway Hyperactivation.” *Frontiers in Molecular Neuroscience* 7 (March):1–15. https://doi.org/10.3389/fnmol.2014.00018.

Leen, Wilhelmina G., Joerg Klepper, Marcel M. Verbeek, Maike Leferink, Tom Hofste, Baziel G. Van Engelen, Ron A. Wevers, et al. 2010. “Glucose Transporter-1 Deficiency Syndrome: The Expanding Clinical and Genetic Spectrum of a Treatable Disorder.” *Brain* 133 (3):655–70. https://doi.org/10.1093/brain/awp336.

Leeuw, N. De, J. Y. Hehir-Kwa, A. Simons, A. Geurts Van Kessel, D. F. Smeets, B. H.W. Faas, and R. Pfundt. 2011. “SNP Array Analysis in Constitutional and Cancer Genome Diagnostics - Copy Number Variants, Genotyping and Quality Control.” *Cytogenetic and Genome Research* 135 (3–4):212–21. https://doi.org/10.1159/000331273.

Leung, Tak Yeung, Ritsuko K Pooh, and Chi Chiu Wang. 2010. “Classification of Pathogenic or Benign Status of CNVs Detected by Microarray Analysis.” *Expert Rev Mol Diagn* 10 (6):717–21.

Li, Marilyn M., Michael Datto, Eric J. Duncavage, Shashikant Kulkarni, Neal I. Lindeman, Somak Roy, Apostolia M. Tsimberidou, et al. 2017. “Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists.” *Journal of Molecular Diagnostics* 19 (1). American Society for Investigative Pathology and the Association

for Molecular Pathology:4–23. https://doi.org/10.1016/j.jmoldx.2016.10.002.

Ligt, Joep de, Philip M. Boone, Rolph Pfundt, Lisenka E L M Vissers, Nicole de Leeuw, Christine Shaw, Han G. Brunner, James R. Lupski, Joris A. Veltman, and Jayne Y. Hehir-Kwa. 2014. “Platform Comparison of Detecting Copy Number Variants with Microarrays and Whole-Exome Sequencing.” *Genomics Data* 2. Elsevier B.V.:144–46. https://doi.org/10.1016/j.gdata.2014.06.009.

Liu-Yesucevitz, L., G. J. Bassell, A. D. Gitler, A. C. Hart, E. Klann, J. D. Richter, S. T. Warren, and B. Wolozin. 2011. “Local RNA Translation at the Synapse and in Disease.” *Journal of Neuroscience* 31 (45):16086–93. https://doi.org/10.1523/JNEUROSCI.4105-11.2011.

Luthardt, Frederick W, and Elisabeth Keitges. 2001. “Chromosomal Syndromes and Genetic Disease.” *Encyclopedia of Life Sciences* 21 (6):625–730. https://doi.org/10.1038/npg.els.0001446.

Makela, Nancy L., Patricia H. Birch, Jan M. Friedman, and Carlo A. Marra. 2009. “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.” *American Journal of Medical Genetics, Part A* 149 (11):2393–2402. https://doi.org/10.1002/ajmg.a.33050.

Malhotra, Dheeraj, and Jonathan Sebat. 2012. “CNVs: Harbingers of a Rare Variant Revolution in Psychiatric Genetics.” *Cell* 148 (6). Elsevier Inc.:1223–41. https://doi.org/10.1016/j.cell.2012.02.039.

Manheimer, Kathryn B., Nihir Patel, Felix Richter, Joshua Gorham, Angela C. Tai, Jason Homsy, Marko T. Boskovski, et al. 2018. “Robust Identification of Deletions in Exome and Genome Sequence Data Based on Clustering of Mendelian Errors.” *Human Mutation*, 1–38. https://doi.org/10.1002/humu.23419.

Mann, Kathy, Erwin Petek, and Barbara Pertl. 2015. “Prenatal Diagnosis : Perspective” 1885:2015. https://doi.org/10.1007/978-1-4939-8889-1.

Manning, Melanie, and Louanne Hudgins. 2010. “Array-Based Technology and Recommendations for Utilization in Medical Genetics Practice for Detection of Chromosomal Abnormalities.” *Genetics in Medicine* 12 (11):742–45. https://doi.org/10.1097/GIM.0b013e3181f8baad.

Mardis, Elaine R. 2013. “Next-Generation Sequencing Platforms.” *Annual Review of Analytical Chemistry* 6 (1):287–303. https://doi.org/10.1146/annurev-anchem-062012-092628.

Maulik, Pallab K, Maya N Mascarenhas, Colin D Mathers, Tarun Dua, and Shekhar Saxena. 2011. “Prevalence of Intellectual Disability: A Meta-Analysis of Population-Based Studies.” *Research in Developmental Disabilities* 32:419–36.

https://doi.org/10.1016/j.ridd.2010.12.018.

McCarroll, Steven A., Tracy N. Hadnott, George H. Perry, Pardis C. Sabeti, Michael C. Zody, Jeffrey C. Barrett, Stephanie Dallaire, et al. 2006. “Common Deletion Polymorphisms in the Human Genome.” *Nature Genetics* 38 (1):86–92. https://doi.org/10.1038/ng1696.

Middelkamp, Sjors, Sebastiaan van Heesch, A. Koen Braat, Joep de Ligt, Maarten van Iterson, Marieke Simonis, Markus J. van Roosmalen, et al. 2017. “Molecular Dissection of Germline Chromothripsis in a Developmental Context Using Patient-Derived IPS Cells.” *Genome Medicine* 9 (1). Genome Medicine:1–14. https://doi.org/10.1186/s13073-017-0399-z.

Miller, David T., Margaret P. Adam, Swaroop Aradhya, Leslie G. Biesecker, Arthur R. Brothman, Nigel P. Carter, Deanna M. Church, et al. 2010. “Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies.” *American Journal of Human Genetics* 86 (5). The American Society of Human Genetics:749–64. https://doi.org/10.1016/j.ajhg.2010.04.006.

Moeschler, J. B. 2006. “Clinical Genetic Evaluation of the Child With Mental Retardation or Developmental Delays.” *Pediatrics* 117 (6):2304–16. https://doi.org/10.1542/peds.2006-1006.

Moeschler, J. B., and M. Shevell. 2014. “Comprehensive Evaluation of the Child With Intellectual Disability or Global Developmental Delays.” *Pediatrics* 134 (3):e903–18. https://doi.org/10.1542/peds.2014-1839.

Pearce, Elizabeth N., John H. Lazarus, Rodrigo Moreno-Reyes, and Michael B. Zimmermann. 2016. “Consequences of Iodine Deficiency and Excess in Pregnant Women: An Overview of Current Knowns and Unknowns.” *The American Journal of Clinical Nutrition* 104:918S–923S. https://doi.org/10.3945/ajcn.115.110429.

Petrij, Fred, Rachel H. Giles, Hans G. Dauwerse, Jasper J. Saris, Raoul C.M. Hennekam, Mitsuo Masuno, Niels Tommerup, et al. 1995. “Rubinstein-Taybi Syndrome Caused by Mutations in the Transcriptional Co-Activator CBP.” *Nature* 376 (6538):348–51. https://doi.org/10.1038/376348a0.

Ploug, Thomas, and Søren Holm. 2017. “Clinical Genome Sequencing and Population Preferences for Information about ‘Incidental’ Findings-From Medically Actionable Genes (MAGs) to Patient Actionable Genes (PAGs).” *PLoS ONE* 12 (7):1–13. https://doi.org/10.1371/journal.pone.0179935.

Quick, M. W. 2006. “The Role of SNARE Proteins in Trafficking and Function of Neurotransmitter Transporters.” *Handbook of Experimental Pharmacology* 175:181–96. https://doi.org/10.1007/3-540-29784-7-9.

Rauch, Anita, Dagmar Wieczorek, Elisabeth Graf, Thomas Wieland, Sabine Endele, Thomas Schwarzmayr, Beate Albrecht, et al. 2012. “Range of Genetic Mutations Associated with Severe Non-Syndromic Sporadic Intellectual Disability: An Exome Sequencing Study.” *The Lancet* 380 (9854):1674–82. https://doi.org/10.1016/S0140-6736(12)61480-9.

Reichenberg, Abraham, Martin Cederlöf, Andrew McMillan, Maciej Trzaskowski, Ori Kapara, Eyal Fruchter, Karen Ginat, et al. 2016. “Discontinuity in the Genetic and Environmental Causes of the Intellectual Disability Spectrum.” *Proceedings of the National Academy of Sciences* 113 (4):1098–1103. https://doi.org/10.1073/pnas.1508093112.

Richards, Sue, Nazneen Aziz, Sherri Bale, David Bick, Soma Das, Julie Gastier-Foster, Wayne W. Grody, et al. 2015. “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.” *Genetics in Medicine* 17 (5):405–24. https://doi.org/10.1038/gim.2015.30.

Riggs, E. R., K. E. Wain, D. Riethmaier, B. Smith-Packard, W. A. Faucett, N. Hoppman, E. C. Thorland, V. C. Patel, and D. T. Miller. 2014. “Chromosomal Microarray Impacts Clinical Management.” *Clinical Genetics* 85 (2):147–53. https://doi.org/10.1111/cge.12107.

Rizzo, Jason M., and Michael J. Buck. 2012. “Key Principles and Clinical Applications of ‘next-Generation’ DNA Sequencing.” *Cancer Prevention Research* 5 (7):887–900. https://doi.org/10.1158/1940-6207.CAPR-11-0432.

Sachidanandam, Ravi, David Weissman, Steven C. Schmidt, Jerzy M. Kakol, Lincoln D. Stein, Gabor Marth, Steve Sherry, et al. 2001. “A Map of Human Genome Sequence Variation Containing 1.42 Million Single Nucleotide Polymorphisms.” *Nature* 409 (6822):928–33. https://doi.org/10.1038/35057149.

Saelaert, Marlies, Heidi Mertes, Elfride de Baere, and Ignaas Devisch. 2018. “Incidental or Secondary Findings: An Integrative and Patient-Inclusive Approach to the Current Debate.” *European Journal of Human Genetics*. Springer US, 1–8. https://doi.org/10.1038/s41431-018-0200-9.

Scheffer, Ingrid E., Samuel Berkovic, Giuseppe Capovilla, Mary B. Connolly, Jacqueline French, Laura Guilhoto, Edouard Hirsch, et al. 2017. “ILAE Classification of the Epilepsies: Position Paper of the ILAE Commission for Classification and Terminology.” *Epilepsia* 58 (4):512–21. https://doi.org/10.1111/epi.13709.

Shaw, Christine J, and James R Lupski. 2004. “Implications of Human Genome Architecture for Rearrangement-Based Disorders : The Genomic Basis of Disease.” *Human Molecular Genetics* 13 (1):57–64. https://doi.org/10.1093/hmg/ddh073.

Shevell, M, S Ashwal, and D Donley. 2003. “Practice Parameter : Evaluation of the Child with Global Developmental Delay : Report of the Quality Standards Subcommittee of the American Academy of Neurology and The Practice Committee of the Child Neurology Society Practice Parameter : Evaluation of T.” *American Academy of Neurology* 60:367–80. https://doi.org/10.1212/01.WNL.0000031431.81555.16.

Shinawi, Marwan, and Sau Wai Cheung. 2008. “The Array CGH and Its Clinical Applications.” *Drug Discovery Today* 13 (17–18):760–70. https://doi.org/10.1016/j.drudis.2008.06.007.

Sijde, Marijke R. van der, Aylwin Ng, and Jingyuan Fu. 2014. “Systems Genetics: From GWAS to Disease Pathways.” *Biochimica et Biophysica Acta - Molecular Basis of Disease* 1842 (10). Elsevier B.V.:1903–9. https://doi.org/10.1016/j.bbadis.2014.04.025.

Simons, Annet, Birgit Sikkema-Raddatz, Nicole de Leeuw, Nicole Claudia Konrad, Rosalind J. Hastings, and Jacqueline Schoumans. 2012. “Genome-Wide Arrays in Routine Diagnostics of Hematological Malignancies.” *Human Mutation* 33 (6):941–48. https://doi.org/10.1002/humu.22057.

Sirven, Joseph I. 2015. “Epilepsy: A Spectrum Disorder.” *Cold Spring Harbor Perspectives in Biology* 7 (9):1–16. https://doi.org/10.1101/cshperspect.a022848.

Spielmann, Malte, Darío G. Lupiáñez, and Stefan Mundlos. 2018. “Structural Variation in the 3D Genome.” *Nature Reviews Genetics* 19 (7):453–67. https://doi.org/10.1038/s41576-018-0007-0.

Spillane, J., D. M. Kullmann, and M. G. Hanna. 2016. “Genetic Neurological Channelopathies: Molecular Genetics and Clinical Phenotypes.” *Journal of Neurology, Neurosurgery and Psychiatry* 87 (1):37–48. https://doi.org/10.1136/jnnp-2015-311233.

Stamberger, Hannah, Marina Nikanorova, Marjolein H. Willemsen, Patrizia Accorsi, Marco Angriman, Hartmut Baier, Ira Benkel-Herrenbrueck, et al. 2016. “STXBP1 Encephalopathy.” *Neurology* 86 (10):954–62. https://doi.org/10.1212/WNL.0000000000002457.

Stankiewicz, Pawel, and James R Lupski. 2002. “Genome Architecture , Rearrangements and Genomic Disorders.” *Trends in Genetics* 18 (2):74–82.

Stouffer, Melissa A, Jeffrey A Golden, and Fiona Francis. 2016. “Neurobiology of Disease Neuronal Migration Disorders: Focus on the Cytoskeleton and Epilepsy.” *Neurobiology of Disease* 92 (Pt A). Elsevier Inc.:18–45. https://doi.org/10.1016/j.nbd.2015.08.003.

Strassberg, Melissa, Gary Fruhman, and Ignatia B. Van Den Veyver. 2011. “Copy-Number Changes in Prenatal Diagnosis.” *Expert Review of Molecular Diagnostics* 11

(6):579–92. https://doi.org/10.1586/erm.11.43.

Swanberg, Susan E., Raman P. Nagarajan, Sailaja Peddada, Dag H. Yasui, and Janine M. Lasalle. 2009. “Reciprocal Co-Regulation of EGR2 and MECP2 Is Disrupted in Rett Syndrome and Autism.” *Human Molecular Genetics* 18 (3):525–34. https://doi.org/10.1093/hmg/ddn380.

Tai, Hwan-Ching, and Erin M. Schuman. 2008. “Ubiquitin, the Proteasome and Protein Degradation in Neuronal Function and Dysfunction.” *Nature Reviews Neuroscience* 9 (11):826–38. https://doi.org/10.1038/nrn2499.

Tarpey, Patrick S, F Lucy Raymond, Sarah O Meara, Sarah Edkins, Jon Teague, Adam Butler, Ed Dicks, et al. 2007. “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.” *American Journal of Human Genetics* 80:345–52.

“The International HapMap Project.” 2003. *Nature*, no. 426 (December):789–96. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3584704&tool=pmcentrez&rendertype=abstract.

Thomas, María Gabriela, Malena Lucía Pascual, Darío Maschi, Luciana Luchelli, and Graciela Lidia Boccaccio. 2014. “Synaptic Control of Local Translation: The Plot Thickens with New Characters.” *Cellular and Molecular Life Sciences* 71 (12):2219–39. https://doi.org/10.1007/s00018-013-1506-y.

Torres, Fátima, Mafalda Barbosa, and Patrícia Maciel. 2016. “Recurrent Copy Number Variations as Risk Factors for Neurodevelopmental Disorders: Critical Overview and Analysis of Clinical Implications.” *Journal of Medical Genetics* 53 (2):73–90. https://doi.org/10.1136/jmedgenet-2015-103366.

Torres, Fátima, Fátima Lopes, and Patrícia Maciel. 2018. “Relevance of Copy Number Variation to Human Genetic Disease.” *Encyclopedia of Life Sciences. In: ELS. John Wiley & Sons, Ltd: Chichester.*, DOI: 10.1002/9780470015902.a0020226.pub2. https://doi.org/10.1002/9780470015902.a0020226.

Valsesia, Armand, Aurélien Macé, Sébastien Jacquemont, Jacques S. Beckmann, and Zoltán Kutalik. 2013. “The Growing Importance of CNVs: New Insights for Detection and Clinical Interpretation.” *Frontiers in Genetics* 4 (MAY):1–19. https://doi.org/10.3389/fgene.2013.00092.

Vermeesch, Joris R., Paul D. Brady, Damien Sanlaville, Klaas Kok, and Rosalind J. Hastings. 2012. “Genome-Wide Arrays: Quality Criteria and Platforms to Be Used in Routine Diagnostics.” *Human Mutation* 33 (6):906–15. https://doi.org/10.1002/humu.22076.

Watson, Corey T, Tomas Marques-Bonet, Andrew J Sharp, and Heather C Mefford. 2014. "The Genetics of Microdeletion and Microduplication Syndromes: An Update." *Annu Rev Genomics Hum Genet* 15:215–44. https://doi.org/10.1002/bdra.23483.Autoantibodies.

Weischenfeldt, Joachim, Orsolya Symmons, François Spitz, and Jan O. Korbel. 2013. "Phenotypic Impact of Genomic Structural Variation: Insights from and for Human Disease." *Nature Reviews Genetics* 14 (2). Nature Publishing Group:125–38. https://doi.org/10.1038/nrg3373.

Werling, D M Donna M, and Daniel H D H Geschwind. 2013. "Sex Differences in Autism Spectrum Disorders." *Current Opinion in Neurology* 26 (2):146–53. https://doi.org/10.1097/WCO.0b013e32835ee548.Sex.

Wieland, J., and F. G. Zitman. 2016. "It Is Time to Bring Borderline Intellectual Functioning Back into the Main Fold of Classification Systems." *BJPsych Bulletin* 40 (4):204–6. https://doi.org/10.1192/pb.bp.115.051490.

Williams, Graham R. 2008. "Neurodevelopmental and Neurophysiological Actions of Thyroid Hormone." *Journal of Neuroendocrinology* 20 (6):784–94. https://doi.org/10.1111/j.1365-2826.2008.01733.x.

Wright, Caroline F., David R. FitzPatrick, and Helen V. Firth. 2018. "Paediatric Genomics: Diagnosing Rare Disease in Children." *Nature Reviews Genetics* 19 (5). Nature Publishing Group:253–68. https://doi.org/10.1038/nrg.2017.116.

Xuan, Jiekun, Ying Yu, Tao Qing, Lei Guo, and Leming Shi. 2013. "Next-Generation Sequencing in the Clinic: Promises and Challenges." *Cancer Lett* 340 (2):284–95. https://doi.org/10.5588/ijtld.16.0716.Isoniazid.

Yamashita, Satoshi, Tomohiro Chiyonobu, Michiko Yoshida, Hiroshi Maeda, Masashi Zuiki, Satoshi Kidowaki, Kenichi Isoda, et al. 2016. "Mislocalization of Syntaxin-1 and Impaired Neurite Growth Observed in a Human IPSC Model for STXBP1-Related Epileptic Encephalopathy." *Epilepsia* 57 (4):e81–86. https://doi.org/10.1111/epi.13338.

Yao, Ruen, Cheng Zhang, Tingting Yu, Niu Li, Xuyun Hu, Xiumin Wang, Jian Wang, and Yiping Shen. 2017. "Evaluation of Three Read-Depth Based CNV Detection Tools Using Whole-Exome Sequencing Data." *Molecular Cytogenetics* 10 (1). Molecular Cytogenetics:1–7. https://doi.org/10.1186/s13039-017-0333-5.

Zamboni, Valentina, Rebecca Jones, Alessandro Umbach, Alessandra Ammoni, Maria Passafaro, Emilio Hirsch, and Giorgio R. Merlo. 2018. *Rho GTPases in Intellectual Disability: From Genetics to Therapeutic Opportunities*. *International Journal of Molecular Sciences*. Vol. 19. https://doi.org/10.3390/ijms19061821.

Zhang, F, W Gu, ME Hurles, and JR Lupski. 2009. "Copy Number Variation in Human

Health, Disease, and Evolution." *Annu Rev Genomics Hum Genet*, 451–81. https://doi.org/10.1146/annurev.genom.9.081307.164217.Copy.

Zhang, Feng, and James R. Lupski. 2015. "Non-Coding Genetic Variants in Human Disease." *Human Molecular Genetics* 24 (R1):R102–10. https://doi.org/10.1093/hmg/ddv259.

Zweier, Christiane, Maarit M. Peippo, Juliane Hoyer, Sérgio Sousa, Armand Bottani, Jill Clayton-Smith, William Reardon, et al. 2007. "Haploinsufficiency of TCF4 Causes Syndromal Mental Retardation with Intermittent Hyperventilation (Pitt-Hopkins Syndrome)." *The American Journal of Human Genetics* 80 (5):994–1001. https://doi.org/10.1086/515583.

CHAPTER 2

Copy number variations as the cause of intellectual disability

Sub-Chapter 2.1 – Genomic imbalances in known and novel candidate *loci* in a group of patients with idiopathic intellectual disability

Disclaimer:

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). This work was accepted for publication in the international peer-reviewed journal ***Orphanet Journal of Rare Diseases*** under the following title and authorship:

Genomic imbalances in known and novel candidate *loci* in a group of 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, Adriana Sampaio, Erik A Sistermans, Bauke Ylstra, Paula Rendeiro, Purificação Tavares, Margarida Reis-Lima, Jorge Pinto-Basto, Ana Maria Fortuna, Patrícia Maciel

*both authors contributed equally for this work

The author of this thesis contributed for all the laboratory work, genomic data acquisition, analysis and interpretation of results for the CC, as well as for the manuscript preparation and discussion of all the cases. For the RC, the laboratory work and genomic data acquisition was performed with particular contribution of Lopes F (co-first author). Torres F also participated in the interpretation of results for the RC.

Orphanet Journal of Rare Diseases

RESEARCH Open Access

Genomic imbalances defining novel intellectual disability associated *loci*

Fátima Lopes[1,2†], Fátima Torres[3,4†], Gabriela Soares[5], Mafalda Barbosa[5,6,7,8,9], João Silva[5,10,11], Frederico Duque[12,13], Miguel Rocha[5,14], Joaquim Sá[3,15], Guiomar Oliveira[12,13], Maria João Sá[5,6], Teresa Temudo[16], Susana Sousa[1,2,10,11], Carla Marques[12], Sofia Lopes[1,2], Catarina Gomes[1,2], Gisela Barros[1,2], Arminda Jorge[17,18], Felisbela Rocha[19], Cecília Martins[19], Sandra Mesquita[20], Susana Loureiro[21], Elisa Maria Cardoso[21], Maria José Cálix[21], Andreia Dias[21], Cristina Martins[22], Céu R. Mota[23], Diana Antunes[24], Juliette Dupont[25], Sara Figueiredo[26], Sónia Figueiroa[27], Susana Gama-de-Sousa[19], Sara Cruz[28], Adriana Sampaio[28], Paul Eijk[29], Marjan M. Weiss[30], Bauke Ylstra[29], Paula Rendeiro[3], Purificação Tavares[3], Margarida Reis-Lima[5,31], Jorge Pinto-Basto[3], Ana Maria Fortuna[5] and Patrícia Maciel[1,2*]

Abstract

Background: High resolution genome-wide copy number analysis, routinely used in clinical diagnosis for several years, retrieves new and extremely rare copy number variations (CNVs) that provide novel candidate genes contributing to disease etiology. The aim of this work was to identify novel genetic causes of neurodevelopmental disease, inferred from CNVs detected by array comparative hybridization (aCGH), in a cohort of 325 Portuguese patients with intellectual disability (ID).

Results: We have detected CNVs in 30.1% of the patients, of which 5.2% corresponded to novel likely pathogenic CNVs. For these 11 rare CNVs (which encompass novel ID candidate genes), we identified those most likely to be relevant, and established genotype-phenotype correlations based on detailed clinical assessment. In the case of duplications, we performed expression analysis to assess the impact of the rearrangement. Interestingly, these novel candidate genes belong to known ID-related pathways. Within the 8% of patients with CNVs in known pathogenic *loci,* the majority had a clinical presentation fitting the phenotype(s) described in the literature, with a few interesting exceptions that are discussed.

Conclusions: Identification of such rare CNVs (some of which reported for the first time in ID patients/families) contributes to our understanding of the etiology of ID and for the ever-improving diagnosis of this group of patients.

Keywords: CNVs, Neurodevelopment, Genotype-phenotype correlation, *CUL4B* overexpression

Background

Intellectual disability (ID) is one of the most common neurodevelopmental disorders (NDDs), affecting nearly 3% of the population worldwide. ID has a complex etiology resulting from the combination of environmental and genetic factors [1]. Relatively recent approaches to the identification of copy number variations (CNVs), have highlighted the relevance of rare de novo, and essentially private mutations that contribute to a significant proportion of the risk of NDDs, being presently an unavoidable element of diagnosis in the field of Neuropsychiatry, Neuropediatrics and Neurodevelopmental Pediatrics.

A substantial number of ID patients have CNVs resulting from deletions or duplications [2, 3]. The frequency of detection of chromosome abnormalities and/or genomic rearrangements in patients with NDDs by array comparative genomic hybridization (aCGH) depends mainly on the patient inclusion clinical criteria and on the microarray design; nevertheless, detection rates are usually higher in patients with ID/developmental delay (DD) that also

* Correspondence: pmaciel@med.uminho.pt
†Fátima Lopes and Fátima Torres contributed equally to this work.
[1]Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, 4710-057 Braga, Portugal
[2]ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal
Full list of author information is available at the end of the article

present malformations or dysmorphic features and more severe cognitive impairment [2]. The characterization of these CNVs in different patient cohorts as well as in the general population is necessary to clarify their clinical relevance and establish adequate genotype-phenotype correlations [4].

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

Results

Global data

From the 325 patients, 30.1% had at least one non-polymorphic CNV detected by aCGH (Part 1 of Additional file 1: Table S1): 8% had pathogenic CNVs, 5.2% had likely pathogenic CNVs and 16.9% had genomic variants of unknown significance (VOUS). The remaining 69.9% patients had only known polymorphic CNVs.

Pathogenic CNVs

The pathogenic CNVs detected were mainly de novo CNVs, including deletions at 1p36.23-p36.21, 2p13.1–13.3, 3q22.1-q23, 5p15.33-p15.32, 6q25.3, 7q11.23, 8p23.1, 11q24.2-q25, 12q24.21-q24.22, 16p11.2, 17q21.31, 22q11.21 and 22q13.3, as well as duplications at 1q21.1, 12q24.21, 9q34.13–34.3, 13q12.12-q34, 14q32.31-q32.33, 14q32.33, 15q11.2-q13.1, 16p13.11, 21q11.2-q22.11, Xp11.22 and Xq28 (see Table 1 for the list of all patients and findings). For most of these CNVs there are reports in the literature describing the phenotypic and genetic findings for similar patients, therefore only some particular cases are described in detail and discussed in Part 1 of Additional file 1, namely: (a) the interstitial deletion at 1p36.23-p36.21 found de novo in patient R1, of interest since interstitial deletions in this region are rarely described in association with NDDs; (b) the deletion at 3q22.1-q23 found de novo in patient R3, which reinforces the association of deletions affecting *FOXL2* gene with blepharophimosis syndrome; (c) 7q11.23 deletions, detected in two non-related patients (C2 and R29), neither of whom presents the classical Williams-Beuren syndrome phenotype; (d) the 22q13.3 deletion found in patient C7, due to the incomplete overlap of the patient's phenotype with that previously described for Phelan-McDermid syndrome; (e) the 9q34 duplications, detected in two non-related patients (C19 and R14): patient C19 has an intragenic *EHMT1* duplication and a clinical presentation that overlaps the core phenotype of Kleefstra syndrome, commonly caused by deletions or point mutations affecting the *EHMT1* gene; patient R14 has three de novo duplications at 9q34.13-q34.3 (affecting the whole *EHMT1* gene), at 14q32.31-q32.33 and at 14q32.33, illustrating the difficulty to ascertain the specific role of each imbalance in complex rearrangements. We also included in this category CNVs occurring in risk-associated *loci*.

Likely pathogenic CNVs

Likely pathogenic CNVs were detected in 5.2% of patients in this study (Table 2; Figs. 1 and 2). They comprise candidate ID-causative *loci* located in 1q43-q44, 2q11.2-q12.2, 7q33, 10q26.3, 17p11.2 and 20q13.12-q13.13 (losses); 1p22.1-p21.3, 7q33, 9q33.2-q33.3, 9q34.3, Xq24 and Xq26.3 (gains) (Table 2). Patients with 1q43-q44, 7q33 and 10q26.3 CNVs have been described elsewhere in detail [5–7]; the patient with a 9q34.3 gain is described together with patient R14 in Part 1 of Additional file 1; therefore, we focus next on the remaining candidate *loci*.

2q11.2-q12.2 deletion

Patient R16 is a 17 year old girl with syndromic ID, cerebral ventricular enlargement, dysmorphic features and hirsutism. She carries a de novo 4.5 Mb deletion at 2q11.2-q12.2 affecting 26 genes, of which *MAP4K4, FHL2, POU3F3* and *CNOT11* have the highest haploinsufficiency score in DECIPHERr [8]. *POU Class 3 Homeobox 3* (*POU3F3*) was previously reported deleted in a boy with ID and dysmorphic features (such as flat nose, prominent

Table 1 Clinical overview of RC patients for whom non-polymorphic CNVs vs likely benign and polymorphic CNVs were detected in the aCGH

Pathogenic + Likely pathogenic (*n* = 23)	Polymorphic CNVs (*n* = 134)
Gender	Gender
Males 15 (65%)	Males 84 (63%)
Females 8 (35%)	Females 50 (37%)
ID	ID
Syndromic 19 (83%)	Syndromic 74 (55%)
Non-syndromic 4 (17%)	Non-syndromic 60 (45%)
Borderline 1 (4%)	Borderline 8 (6%)
Mild 15 (65%)	Mild 75 (56%)
Moderate 6 (26%)	Moderate 30 (22%)
Severe 0 (0%)	Severe 15 (11%)
Profound 1 (4%)	Profound 6 (4%)
History	History
Sporadic 11 (48%)	Sporadic 54 (40%)
Family history of ID 15 (65%)	Family history of ID 80 (60%)
Co-morbidities	Co-morbidities
Congenital anomalies 11 (48%)	Congenital anomalies 64 (48%)
Epilepsy 2 (9%)	Epilepsy 19 (14%)
Microcephaly 4 (17%)	Microcephaly 23 (17%)
Macrocephaly 1 (4%)	Macrocephaly 13 (10%)

ears, large eyebrows and low hairline) [9], 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 [10], thus standing out as a good candidate for the DD/ID in the patient.

17p11.2 deletions

Patient C15 is a 10 year old boy referred for consultation for DD, namely language and motor impairment, ataxia and some dysmorphic features, including hypertelorism, strabismus and low-set ears. It was not possible to re-evaluate for IQ testing, but at the time of first evaluation he had no formal cognitive deficit (according with the GMDS score when he was 5 years old) and cerebral magnetic resonance imaging (MRI) showed no alterations. He has what appear to be two consecutive deletions at 17p11.2: a 420.6Kb deletion, that encompasses 5 genes, and a 2.77 Mb deletion that encompasses 36 genes. 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); however, the phenotype of the patient and mother is not similar to that of SMS, and the deletion does not affect the retinoic acid induced 1 (*RAI1*) gene, thought to cause most of the SMS core phenotype [11]. Among the genes affected by patient C15's deletions, there are several others whose function could potentially contribute for his phenotype (detailed in Part 1 of Additional file 1).

20q13.12-q13.13 deletions

Patient R20 is a 16 year old girl with mild ID (IQ = 56), speech delay, MIC and facial dysmorphisms. Brain imaging studies revealed no structural alterations. She also has astigmatism and attention deficit hyperactivity disorder (ADHD). She carries a de novo 5.5 Mb deletion at 20q13 .12-q13.13 encompassing 123 genes. Among these, the genes *KCNB1, PIGT, CTSA, SLC2A10* and *ARFGEF2* were associated with human disease (detailed in Part 1 of Additional file 1).

1p22.1p21.3 duplications

Patient C16 is a 7 year old girl with motor and speech delay, with a global DQ of 56.3 (GMDS). She carries a maternal 1p22.1p21.3 duplication of 6.461 Mb that affects 44 genes. 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 a 10 year old brother suspected of having cognitive deficit: he was not evaluated yet, but he is attending the 2nd grade and does not yet know how to read. There is also a positive history of learning difficulties on the maternal grandfather's family side. The duplication affects several genes (Fig. 2a), including the *FAM69A* gene, which encodes a member of the FAM69 family of cysteine-rich type II transmembrane proteins. FAM69 proteins are thought to play a fundamental role in the endoplasmic reticulum, in addition to specialized roles in the vertebrate nervous system, according to a brain-specific or brain-including expression pattern [12]. Consistently, several *FAM69* genes have been linked to neuropsychiatric disorders: *C3ORF58* (*DIA1*) with autism [13]; *CXORF36* (*DIA1R*) with X-linked ID [14] and *FAM69A* with schizophrenia and bipolar disease [15]. Even though the contribution of the excess of dosage for NDDS is still unknown, this gene can be considered a good candidate to explain the disease in the patient.

9q33.2-q33.3 triplication

Patient R21 is a 17 year old girl with mild ID (IQ = 53) and familial history of ID. During the neonatal period she presented seizures (flexion spasms and later generalized tonic-clonic), controlled with Phenobarbital, which was discontinued at 23 months; EEG initially showed lateral paroxystic activity, bilaterally, and a normal result at 6 months; brain MRI was normal. Additionally, she presented dysmorphic facial features (Fig. 2), a muscular ventricular septal defect that closed spontaneously, hypothyroidism, hypotonia, global DD, growth deceleration (height and weight around the 3rd centile after 12 months) with normal head size, around the 75th centile, delayed bone maturation (~ 3 years), growth hormone deficiency and short neck. 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. Moreover, this triplication apparently disrupts the *FBXW2* gene that encodes for an F-box protein. F-box proteins are one of the four subunits of ubiquitin protein ligases, called 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 [16]. Also included in the CNV are the *LHX2* and *LHX6* genes, both encoding transcription factors described to play roles in brain development [17, 18]. Additionally, *LHX2* was also described to be involved in osteoclast differentiation and its overexpression inhibits skeletal muscle differentiation [19]. *LHX6* is also known to play a role in cranial and tooth development [20], hence these genes could be of relevance to the cranioskeletal phenotype of the patient.

Based on the location within the triplication region and the expression levels described we selected the *FBXW2*, *NEK6* and *PSMB7* genes (detailed in Part 1 of Additional file 1) to study at the mRNA level in peripheral blood in the patient. The three genes had an increased expression when compared to controls (Fig. 2b). For *NEK6* these findings are in accordance with the fact it is included inside

Table 2 List of pathogenic CNVs

Patients	Gender	Alteration (Hg19)	Type	Size (Mb)	Genes	Key gene(s) involved	Associated syndrome	Phenotype overlap	Inheritance	Confirmation	Array platform	Ref
R1	Male	arr 1p36.23-p36.21(8,593,674-15,396,672)x1dn	del	6.7	86	*ANGPTL7, CASZ1, MAD2L2, RERE*	–	–	de novo	NP	1	–
R2	Male	arr 2p13.1-p13.3(70,894,906-74,986,518)x1dn[c]	del	4	62	*CYP26B1, EXOC6B*	–	–	de novo	NP	1	Wen J, 2013
R3	Male	arr 3q22.1-q23(131,415,639-141,618,552)x1dn	del	1.020	65	*FOXL2*	BPES	Yes (eye features)	de novo	NP	1	–
C1	Male	arr 5p15.33-p15.32(204,849-5,014,883)x1	del	4.81	30	*TERT [CTNND2 not involved]*	–	–	ND	NP	2	–
R4	Male	arr 6q25.3(156,012,754-158,804,494)x1dn[c]	del	2.6	14	*ARID1B*	Coffin-Siris syndrome	Yes	de novo	NP	1	Santen GW, 2013
C2	Male	arr 7q11.23(72,721,760-74,140,846)x1	del	1.419	28	*BAZ1B, STX1A, WBSCR22, ELN*	Williams-Beuren syndrome	Partially	ND	NP	2	–
R5	Female	arr 8p23.1(7,039,276-12,485,558)x1dn	del	5.5	70	SOX7, *GATA4*	8p23.1 deletion syndrome	Yes (cardiac)	de novo	NP	1	–
C3	Male	arr 11q24.2-q25(125,232,584-134,446,160)x1dn	del	9.214	54	*KIRREL3, ETS1, FLI1, KCNJ1, KCNJ5, RICS*	–	Partially	de novo	qPCR	2	–
R6	Female	arr 12q24.21-q24.22(115,505,500-117,441,683)x1dn[c]	del	0.2	10	*MED13L*	–	Yes	de novo	qPCR	1	Adegbola A, 2015
C4	Male	arr 16p11.2(29,674,336-30,198,123)x1dn	del	0.524	29	*KCTD13*	16p11.2 deletion syndrome	–	de novo	NP	2	–
C5	Male	arr 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	NP	3	–
C6	Male	arr 22q11.21(18,894,835-21,505,417)x1	del	2.611	59	*TBX1*	22q11 deletion syndrome	–	ND	NP	2	–
C7	Male	arr 22q13.3(49,513,903-51,178,264)x1	del	1.664	39	*SHANK3*	22q13.3 deletion syndrome (Phelan-McDermid syndrome)	Partially	ND	NP	2	–
C8	Male	arr 1q21.1q21.2(146,106,723-147,830,830)x3dn	dup	1.7	17	*HYDIN2, PRKAB2*	1q21.1 duplication syndrome [e]	Partially	de novo	qPCR	4	–
R7	Male	arr 1q21.1(145,883,119-148,828,690)x3pat	dup	2.5	23	*HYDIN2, PRKAB2, GJA5*	1q21.1 duplication syndrome [e]	Yes	paternal	NP	1	–
R8	Male	arr 12q24.21(116,408,736-116,704,303)x3dn[c]	dup	0.3	2	*MED13L*	–	Yes	de novo	qPCR	1	Adegbola A, 2015
C9	Male	arr 13q12.12-q34(23,749,431-115,083,342)x2.15[a]	dup	91.33	##	–	Trisomy 13 (mosaicism)	Yes	ND	Karyotype[d]	2	–

Table 2 List of pathogenic CNVs *(Continued)*

Patients	Gender	Alteration (Hg19)	Type	Size (Mb)	Genes	Key gene(s) involved	Associated syndrome	Phenotype overlap	Inheritance	Confirmation	Array platform	Ref
C10	Female	arr 15q11.2-q13.1(22880274–29,331,964)x3mat	dup	6.45	111	*CYF1P1, NIPA2, NIPA1, MKRN3, NDN, MAGEL2, SNURF/SNRPN, UBE3A GABRB3*	15q11-q13 duplication syndrome[b]	Yes	maternal	NP	2	–
C11	Female	arr 16p13.11(15,034,010-16,199,882)x3	dup	1.166	11	*NDE1*	16p13.11 duplication syndrome [e]	–	ND	NP	5	–
R9	Male	arr 16p13.11(15,421,671-16,443,968)x3mat	dup	1	19	*NDE1*	16p13.11 duplication syndrome [e]	Yes	maternal	NP	1	–
R10	Male	arr 16p13.11(15,484,180-16,308,344)x3mat	dup	0.8	9	*NDE1*	16p13.11 duplication syndrome [e]	Yes	maternal	NP	1	–
C12	Male	arr 21q11.2-q22.11(14,417,523-34 ,894,625)x3	dup	20.47	110	*DSCR1, DSCR2, DSCR3, DSCR4, APP*	–	No	ND	NP	2	–
R11	Male	arr Xp11.22(53,569,653-53,769,748)x2mat	dup	0.2	3	*HUWE1*	–	Yes	maternal	qPCR	1	–
R12	Male	arr Xq28(152,348,378-155,228,013)x2dn	dup	2.8	78	*MECP2*	MECP2 duplication syndrome	Yes	de novo	NP	1	–
R13	Male	arr Xq28(153,130,545-153,602,293)x2mat	dup	0.5	16	*MECP2*	MECP2 duplication syndrome	Yes	maternal	NP	1	–
R14	Male	arr 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	NP	1	–
		arr 14q32.31-q32.33(102,959,110-104,578,612)x3dn	dup	1.620	22	*MARK3, KLC1, EIF5*	–	–	de novo	NP	1	–
		arr 14q32.33(105,104,831-106,531,339)x3dn	dup	1.427	24		–	–	de novo	NP	1	–

Patients R1 to R14: from research cohort; Patients C1 to C12: from clinical cohort; *NP* Not performed, *ND* Not determined; ([a]): mosaicism; ([b]) methylation status for *SNRPN* is normal (studied by MLPA); ([c]): Published in detail elsewhere; ([d]): karyotype revealed a balanced translocation between chromosomes 13 and 14, resulting in mosaic trisomy 13; ([e]): Other causes of disease were not excluded therefore the variant might not explain the total phenotypic presentation. Array platform 1: Agilent 180 K; 2: KaryoArray®v3.0 (Agilent 8x60k); 3: Affymetrix CytoScan HD array; 4: Affymetrix CytoScan 750 K; 5: Agilent Whole Genome 244 K

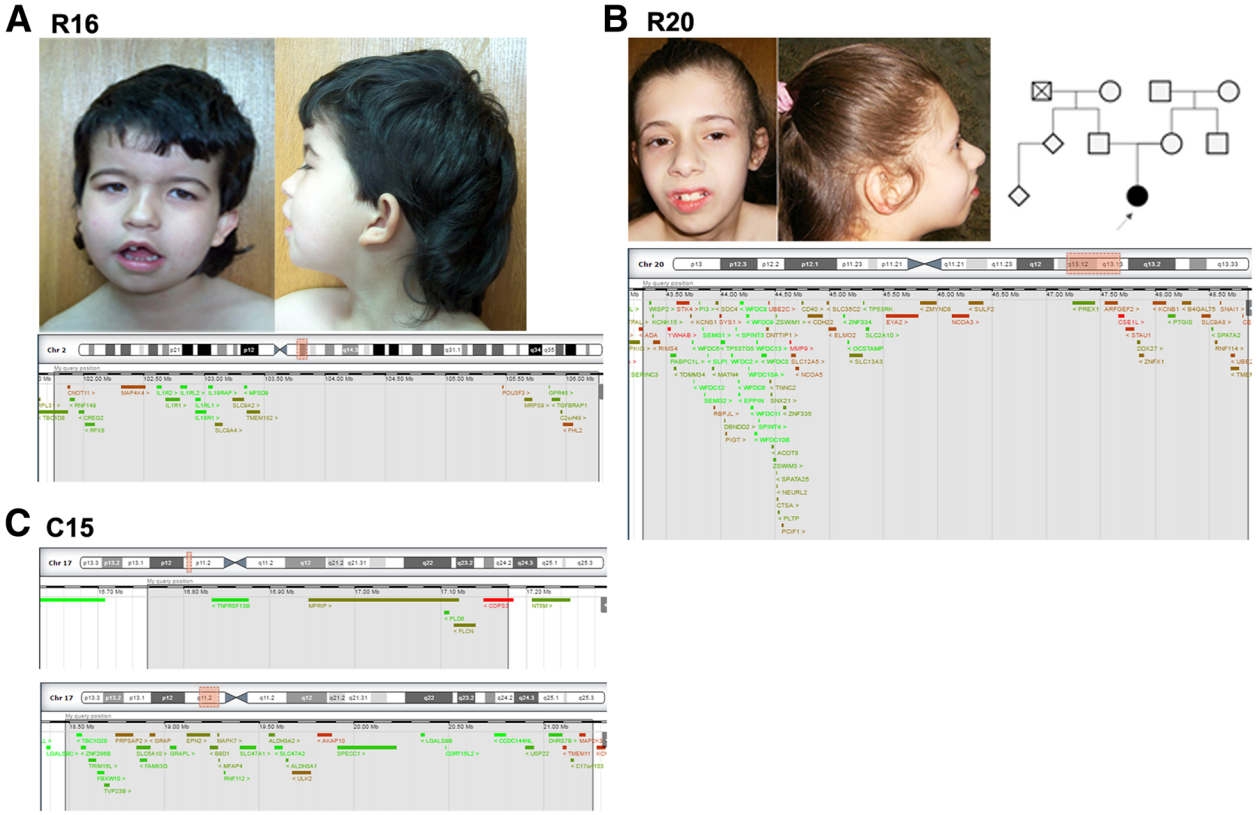


Fig. 1 Facial appearance of patients and schematic representation of the deletions. **a** Patient R16 facies, with low set posteriorly rotated ears, anteverted ears with simplified helix, temporal narrowing with prominent metopic suture, arched eyebrows, synophrys, bilateral epicanthal folds, bulbous nasal tip, thin upper lip, open mouth with downturned corners, micrognathia; pedigree and deleted region of chromosome 2 (highlighted in red in the chromosome scheme (above) and in grey in the genes' portion (below), adapted from DECIPHER). **b** Patient R20 facies, with wide forehead, strabismus, high nasal bridge, wide base of nose, bulbous nasal tip, short and smooth philtrum, thin upper lip with effaced cupid's bow, prominent central incisors and micrognathia; pedigree and deleted region of chromosome 20. **c** Patient C15: deleted region in chromosome 17

the triplicated region. Regarding *FBXW2* and *PSMB7*, we had hypothesized that their expression could be diminished since they are located at the breakpoints, which we concluded not to be the case. To the best of our knowledge no mutations in any of these three genes were reported in human NDDs, making their involvement in our patient's symptomatology difficult to confirm at this stage.

Xq24 duplication

Patient R22 is a 14 year old boy with borderline IQ (IQ = 80) and a familial history of ID (two brothers and cousins with ID), an apparently benign cardiac arrhythmia, overweight (BMI 23.6 Kg/m^2 P90), stereotypies and ADHD. He carries a 0.3 Mb maternally inherited duplication at Xq24 affecting four genes (*CUL4B, LAMP2, C1GALT1C1, MCTS1*), his mother being asymptomatic. Both point mutations and large deletions in the *CUL4B* gene are described as causative of X-linked ID and cerebral malformations [21, 22]. *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 [23]. Additionally, *CUL4B* also targets *WDR5* for ubiquitylation leading to its degradation in neurons nucleus, which causes impaired neurite outgrowth [24]. However, to our knowledge, there is only one 47.2 Mb duplication encompassing *CUL4B* (and other genes) described in a patient with ID [25], 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 disorder in the patient is in fact driven by a dosage increase in *CUL4B*. The *LAMP2* gene, located in the duplication breakpoint and encoding a protein with roles in autophagy/lysosomal function, does not present altered expression in the patient, suggesting that it may not be contributing to this phenotype (Fig. 2c).

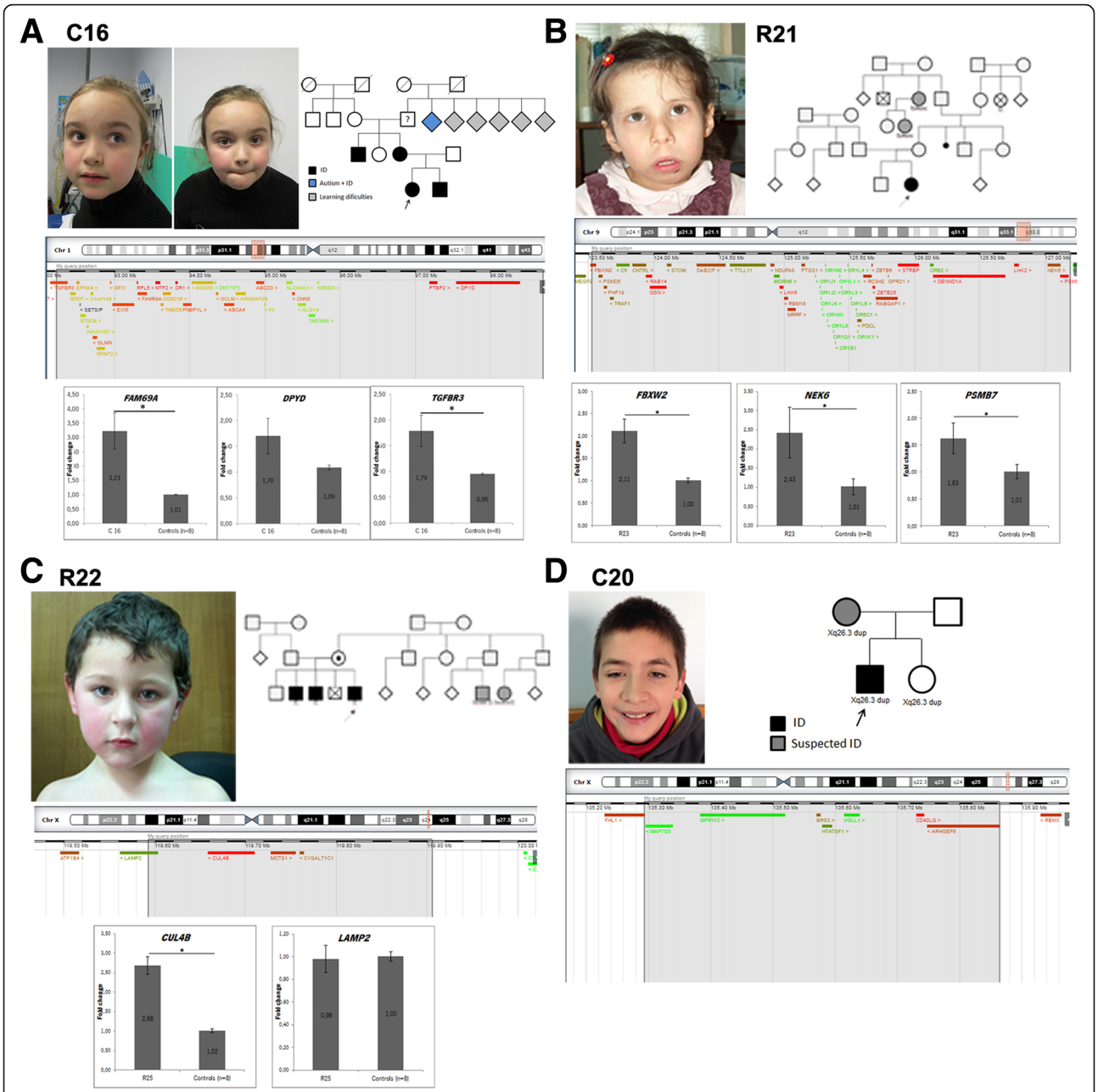


Fig. 2 Overview of some patients with likely pathogenic duplications. **a** Patient C16 - facial appearance: mildly dysmorphic, with high forehead and frontal bossing, thick eyebrows and mildly anteverted nares; pedigree, schematic representation of the duplicated 1p region and expression pattern for genes *FAM69A*, *DPYD* and *TGFBR3*. **b** Patient R21 - facial appearance: large forehead, sparse lateral eyebrows, epicanthal folds, large nose, anteverted nares, long smooth philtrum, downturned corners of mouth and micrognathia; pedigree, schematic representation of the triplicated 9q region and expression pattern for genes *FBXW2*, *NEK6* and *PSMB7*. **c** Patient R22 - facial appearance: mildly dysmorphic with large forehead and frontal central hair whorl; pedigree, schematic representation of the duplicated Xq region and expression pattern for *CUL4B* and *LAMP2* genes. **d** Patient C20 - facial appearance: mildly dysmorphic patient with thick eyebrows, wide palpebral fissures and thin upper lip; pedigree and schematic representation of the duplicated Xq region. *B2M* and *PPIB* were used as housekeeping genes; * $p < 0.05$ (Student t-test)

Xq26.3 duplication

Patient C20 is a 17 year old boy referred to the consultation due to general DD. He carries a 570.1Kb duplication at Xq26.3 inherited from his mother, who has a suspicion of some cognitive impairment but for whom no formal intellectual assessment was possible. He has a global DQ of 57.1 (evaluated at the age of 10 years), scoring below the average in all GMDS sub-scales, namely on language and eye-hand co-ordination, and is described as a friendly boy. He has speech delay, dolichocephaly and several

dysmorphisms, including micrognatia, syndactyly and clinodactyly. His younger sister (8 years old) also carries the duplication but has no ID and has a normal development for her age which, this being an X-linked gene, is not incompatible with the causality of disease. The duplication encompasses the several genes (Fig. 2d) including the *ARHGEF6* gene. *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 [26]. *ARHGEF6* specifically has been implicated in the regulation of spine morphogenesis and loss of function (LoF) mutations have been found in patients with X-linked ID [27]. A 2.8 Mb duplication in Xq26.2-Xq26.3 has also been described in two brothers with ID and the *ARHGEF6*, *PHF6*, *HPRT1* and *SLC9A6* genes have been identified as potential contributors to those patients' phenotype [28]. 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 ID, dolichocephaly and facial dysmorphisms) suggest a determinant role for *ARHGEF6* gene in phenotypes associated with Xq26 microduplications [28]. Expression data in the periphery for some of the genes involved in the duplication didn't retrieve results that we could interpret.

CNVs of unknown significance

In the VOUS group, we included CNVs which did not encompass a known NDDs-related CNV region and for which (i) pathogenicity was not sufficiently supported by biological data, and/or (ii) similar copy number changes were described in control databases, and/or (iii) were inherited from a parent for whom the clinical presentation was not known. For 50% of these cases, inheritance from parents was not possible to determine due to parental sample unavailability, thus reducing our ability to interpret their clinical significance. A summary of the VOUS identified in this study is presented in Part 1 of Additional file 1: Table S2).

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 8% of patients and to propose novel candidate ID *loci* in 5.2%. 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 [29–31]. The CNVs classified as pathogenic often appear de novo and affect (in general) dozens of genes. Some difficulties arose when classifying several of these CNVs as, in some cases, although they occurred in known syndrome regions not all the patients carrying them presented the major clinical features established for that particular syndrome. In fact, even these well-established pathogenic CNVs can be associated with a broad and distinctive phenotypic presentation, as observed in patients C2 and R29, both with WBS associated deletions but not presenting the full-blown phenotype of this syndrome. In this perspective, we believe that the main contributions of this work are: (I) the reporting of new patients with CNVs in regions associated with identified syndromes but with different clinical presentations; (II) the reporting of novel candidate ID-causative *loci* at 2q11.2-q12.2 (del), 7q33 (del and dup), 10q26.3 (del), 17p11.2 (del), 20q13.12-q13.13 (del), 1p22.1-p21.3 (dup), 9q33.2-q33.3 (tri), 9q34.3 (dup), Xq24 (dup) and Xq26.3 (dup); (III) the study in patients with copy number gains of the mRNA expression in peripheral blood 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. Many of the CNVs here detected by aCGH were rare and restricted to one patient/family, which made their contribution to the patient's phenotype difficult to assess. Several of these have been therefore classified as VOUS and their clinical significance needs to be carefully addressed in future studies. Individually rare intermediate-size CNVs (frequency, ≤0.05%; ≥250 kb), and not necessarily assigned a priori as pathogenic, appear to be collectively common in unselected populations (10.5%), and have been associated with ID and negatively with educational attainment [4]; being so, even these should not be excluded as cause of disease but rather re-assessed in the face of accumulating information, in order to establish useful genotype-phenotype correlations. Nevertheless, one cannot exclude the possibility that some of these CNVs are unrelated to pathogenesis, namely in patients where no other genomic testing (such as whole-exome or whole-genome sequencing) was performed to rule out other causes, this being a potential limitation of this work.

NDDs associated pathways: old and new genes

The likely pathogenic CNVs here proposed as novel candidate *loci* for ID encompass several genes that either were already associated with NDDs (like *CUL4B*) or are now 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

Transcriptional regulation is an essential component of the neuronal differentiation programs and of the response to stimulation patterns underlying neuronal plasticity;

genes involved in these pathways have been implicated in well-known NDDs, as is the case of *FOXL2* [32], *BAZ1B* [33], and *EBF3* [7]. This work revealed genes that appear to be good candidate *loci* for ID; of those, *POU3F3*, already described deleted in a patient with ID [9], stands as a strong candidate.

Chromatin modifiers/chromatin remodeling proteins

An excess of mutation genes encoding proteins involved in chromatin regulation have been described in NDDs [34]. *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. *ARID5A* encodes for a protein belonging to the ARID family of proteins with important roles in development, tissue-specific gene expression and proliferation control [35].

Ubiquitin signaling

Ubiquitin-mediated degradation of proteins is a crucial mechanism for cell maintenance and viability [36]. Several genes belonging to this pathway are described to be associated with NDDs, as is the case of *CUL4B* [21], shown here to be duplicated in two patients. *UBE2C* encodes a key component of the ubiquitin proteasome system (UPS) that participates in cell cycle progression and checkpoint control [37]. The *NEURL3* and *CNOT4* genes also encode for proteins with E3 ubiquitin-protein ligase activity; as for *FBXW2*, it encodes for one of the four types of subunits of SCF ubiquitin-protein ligases. Neither of these genes has been linked, until now, with NDDs, but our findings reinforce the idea that genes encoding for proteins belonging to the UPS are possible new candidate genes for NDD phenotypes.

Cytoskeleton regulation and organization, cell shape and motility

Several NDDs are caused by mutations in genes regulating neuronal migration, which often encode for proteins involved in the function of the cytoskeleton [38]. *TSC1*, involved in microtubule-mediated protein transport due to unregulated mTOR signaling [39], and *ARHGEF6*, here described in different CNVs, have been previously associated with NDDs [39, 40]. *B9D1* has been confirmed as a novel Meckel syndrome gene [41].

Intracellular vesicular trafficking and exocytosis

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) [42, 43]. The collection of patients presented herein also allowed the first description of *EXOC6B* gene haploinsufficiency in association with DD/ID (reported in detail in a dedicated publication) [44].

Signaling mediators/transducers/ receptor activity/ transmembrane proteins

Disruption of synaptogenesis has been associated with ID and NDDs [45] and in this work we could identify CNVs in several genes associated with this pathway. *SEMA4C* gene encodes a transmembrane semaphorin which regulates axonal guidance in the developing nervous system [46]. Syntaxins, such as Syntaxin 1A, encoded by *STX1A* gene, are key molecules implicated in the docking of synaptic vesicles with the presynaptic plasma membrane [47]. Signaling processes are essential for proper cellular function and usually implicate enzymes, transmembrane proteins and voltage ion-channels whose disruption may be associated with disease [48]. Many of the genes described herein, including *CACNA1C, GPR45, TNFRSF13B, FAM69A, AKT3* and *CSE1L,* are associated with these pathways, highlighting once again the crucial contribution of proper cellular signaling and synapse development and function for ID/DD.

Of notice, and although our attempts of establishing genotype-phenotype correlations was mostly focused on dosage impact of individual genes (e.g. haploinsufficiency/ overexpression), CNVs may also lead to disease through other mechanisms, namely gene fusion generation [49] and impact on genome architecture, for example Topological Associated Domain disruption, with impact on the expression of genes located outside the affected regions [50].

Conclusion

The 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 a decrease in the rate at which new syndromes were described has been observed, most likely because the most frequent/recurrent CNVs were described in the early days of aCGH [51]. For the remaining and rarer (often "private") forms, it is still important, however, to make an effort to share their clinical and genetic features as well as the CNV data, to support future diagnosis and establishment of genotype-phenotype correlations, as well as the identification of novel candidate genes for disease, as those advanced here.

Subjects and methods

Subjects

This work included the analysis of 325 ID patients (full IQ (FIQ) below 70 and borderline FIQ 70–80) of Portuguese origin (36.9% females, 63.1% males), of which 188 (mostly trios) were included in a research cohort (RC) and 137 were studied in the context of routine clinical genetics diagnostics (clinical cohort, CC), all being referenced as having NDDs (detailed description of inclusion

Table 3 List of likely pathogenic CNVs

Patients	Gender	Alteration (Hg19)	Type	Size (Kb)	Genes	Relevant genes involved	Confirmation	Inheritance	DGV controls	DECIPHER	Array platform	Ref
C13	Male	arr 1q43-q44(240,043,427-249,233,096)x1dn [f]	del	3.7	18	*AKT3*	qPCR	de novo[d]	No	250,152, 250,915 (smaller)	1	Lopes F, et al., 2019
R15	Female	arr 1q43-q44(243,552,007-243,738,675)x1dn [f]	del	0.19	2	*AKT3*	qPCR	de novo[d]	*No*	252,432 (smaller)	2	Lopes F, et al., 2019
C14	Male	arr 1q43q44(243,592, 147-243,749,968)x1pat [f]	del	0.16	2	*AKT3*	qPCR	paternal	*No*	252,432 (smaller)	1	Lopes F, et al., 2019
R16	Female	arr 2q11.2-q12.2(101,756,265-106,265,018)x1dn	del	4500	24	*MAP4K4, FHL2, POU3F3, CNOT11*	qPCR	de novo	No	251,756	2	–
R17, R18[e]	Male, Female	arr 7q33(133,176,651-135,252,871)x1mat [f]	del	2076	23	*AGBL3, CNOT4, CALD1, EXOC4*	qPCR	maternal[a]	No	256,036	2	Lopes F, et al., 2018
R19	Female	arr 10q26.3(131,374,701-132,030,468)x1dn	del	600	3	*EBF3*	qPCR	de novo	3/6564[b]	No	2	Lopes F et al., 2017
C15	Male	arr 17p11.2(16,757,564-17,178,161)x1mat	del	420	5	*COPS3*	NP	maternal[a]	No	No	3	–
		arr 17p11.2(18,478,816-21,255,056)x1mat	del	2770	36	*EPN2, RNF112, ULK2, ALDH3A2, AKAP10, B9D1*	NP	maternal[a]	No	340,692 (smaller)		–
R20	Female	arr 20q13.12-q13.13(43,283,820-48,850,844)x1dn	del	5500	88	*KCNB1, PIGT, CTSA, SLC2A10, ARFGEF2*	NP	de novo	No	309	2	–
C16	Female	arr 1p22.1p21.3(92,227,986-98,689,243)x3mat	dup	6461	44	*FAM69A, TGFBR3, GLMN, EVI5, RPL5, MTF2, DR1, ABCA4, ABCD3, CNN3, PTBP2, DPYD*	qPCR	maternal[a]	No	318,358	1	–
C17,C18[e]	Male, Male	arr 7q33(134,598,205-134,815,177)x3mat [f]	dup	216	2	*CALD1, AGBL3*	qPCR	maternal[a]	No	No	1	Lopes F, et al., 2018
R21	Female	arr 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	2	–
C19	Female	arr 9q34.3(140540819–140,659,057)×3mat	dup	0.118	2	*EHMT1*	NP	maternal	1/2504 (smaller)	No	1	–
R22, R23[e]	Male, Male	arr Xq24(119,592,606-119,904,981)x2mat	dup[c]	300	4	*CUL4B, LAMP2, C1GALT1C1, MCTS1*	qPCR	maternal	No	No	2	–
C20	Male	arr Xq26.3(135,293,144-135,863,290)x2mat	dup	570	9	*ARHGEF6, CD40LG, BRS3, MAP7D3*	qPCR	maternal	No	No	3	–

Patients R15 to R23: from research cohort; Patients C13 to C20: from clinical cohort; *NP* Not performed; ([a]): inherited from an affected parent; ([b]): doubt regarding the quality of the call in these controls; ([c]) duplication may disrupt gene if located in tandem; ([d]) paternity and maternity confirmed; ([e]): siblings; ([f]): family described elsewhere. Array platform 1: Affymetrix Cytoscan 750 K; 2: Agilent 180 K; 3: KaryoArray®v3.0 (Agilent 8x60k)

criteria and clinical characterization provided in Part 1 of Additional file 1). For the RC we were able to obtain DNA for all the parents as well as a more extensive clinical description (see Table 3).

Methods

Genomic DNA was extracted from peripheral blood using either the Citogene® DNA isolation kit (Citomed, Portugal) manually or the QIAsymphony SP kit and apparatus. aCGH was performed using the following platforms Agilent 180 K (GPL15397); KaryoArray®v3.0 (Agilent 8x60k); Agilent Whole Genome 244 K (GPL10118); Affymetrix CytoScan HD (GPL1613) or CytoScan 750 K (GPL18637) (detailed description provided in Additional file 1).

Data analysis

CNVs detected were classified using criteria adapted from those previously described elsewhere [3, 52] as: pathogenic, likely pathogenic, CNVs of unknown clinical significance (VOUS) (detailed description in Part 2 of Additional file 1). For simplification of terminology throughout the text and in the tables, the term CNV is used for pathogenic and likely pathogenic variants, as well as VOUS. Polymorphic CNVs were not further considered in our analysis, except where specifically indicated (e.g. known risk *loci,* although relatively frequent, were considered pathogenic). All alteration are described in the tables as in the Decipher database (for example 12q24.21-q24). For CNV confirmation we performed qRT-PCR (7500-FAST Real Time PCR, Thermo Fisher Scientific, Waltham, MA, USA), using *SDC4* and *ZNF80* as reference genes (detailed description in Part 2 of Additional file 1; primers in Table S3). Total RNA was isolated from leukocytes using the QIAsymphony RNA Kit (QIAGEN GmbH, Germany), according to the manufacturer's protocol. First-strand cDNA synthesized using SuperScript® III Reverse Transcriptase (RT) (Thermo Fisher Scientific, Waltham, MA, USA). Expression analysis was performed by quantitative real-time reverse transcription PCR (qRT-PCR) using Power SYBR Green® (Thermo Fisher Scientific, Waltham, MA, USA) (detailed description in Part 2 of Additional file 1; genes and primers listed in Table S4).

Additional file

Additional file 1: Figure S1. Facial appearance of some patients carrying pathogenic variants. **Figure S2.** Clinical features of patients R14 and C19 and images of their CNVs. **Table S1.** Patients with altered aCGH results (i.e. with CNVs classified as non-polymorphic). **Table S2.** List of variants of unknown clinical significance (VOUS). **Table S3.** Primers used for quantitative PCR confirmation. **Table S4.** Primers used for expression studies. **Table S5.** OMIM entrance, haploinsufficiency score and constrain metrics for the selected genes in patient R16. **Table S6.** OMIM entrance, haploinsufficiency score and constrain metrics for the selected genes in patient C15. **Table S7.** OMIM entrance, haploinsufficiency score and constrain metrics for the selected genes in patient R20. **Table S8.** OMIM entrance, haploinsufficiency score and constrain metrics for the selected genes in patient C16. **Table S9.** OMIM entrance, haploinsufficiency score and constrain metrics for the selected genes in patient R21. **Table S10.** OMIM entrance, haploinsufficiency score and constrain metrics for the selected genes in patient C19. **Table S11.** OMIM entrance, haploinsufficiency score and constrain metrics for the selected genes in patients R22 and R23. **Table S12.** OMIM entrance, haploinsufficiency score and constrain metrics for the selected genes in patient C20. (DOC 11550 kb)

Acknowledgements

We would like to thank all the patients and their families for participation in this study and for allowing this publication. We would also like to acknowledge the DECIPHER Consortium, Database of Genomic Variants and OMIM since this study makes use of data generated and managed by these platforms.

Financial disclosure

FT, PR, PT and JPB authors are employed by company CGC Genetics. All other authors declare no financial competing interests.

Authors' contributions

FL, FT, SS, SL and PR performed the molecular studies and analysed the molecular data. PE, JW and BY contributed to the molecular studies and to the analysis of molecular data. GS, MB, JS, FD, MR, JS, GO, MJS, TT, CM, CG, GB, AJ, FR, CM, SM, SL, EMC, MJC, AD, CN, CRM, DA, JD, SF, SF, SGS, SC, AS, MRL, JPB and AMF collected and analysed clinical data. FL, FT and PM drafted the paper. PM and PT obtained funding for this study. The study was performed under the direction of PM. All authors have agreed with and approved the final version.

Funding

This work has been funded by FEDER funds, through the Competitiveness Factors Operational Programme (COMPETE), and by National funds, through the Foundation for Science and Technology (FCT), under the scope of the projects: PIC/IC/83026/2007, PIC/IC/83013/2007 and POCI-01-0145-FEDER-007038. This work has also been funded by the project NORTE-01-0145-FEDER-000013, supported by the Northern Portugal Regional Operational Programme (NORTE 2020), under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (FEDER). FL was supported by Foundation for Science and Technology (FCT) through the fellowship SFRH/BD/90167/2012.

Availability of data and materials

All data generated or analysed during this study are included in this published article and in its supplementary information files.

Ethics approval and consent to participate

The enrollment of the patients and families was done by the referring doctor, clinical information was gathered in an anonymized database and written informed consent was obtained for all participants and/or their legal guardians for both study participation and publication of identifying information/images according to the Portuguese Data Protection Authority (CNPD). This study was approved by the ethics committee of Center for Medical Genetics Dr. Jacinto Magalhães, Porto Hospital Center and all research was performed in accordance with relevant guidelines/regulations.

Consent for publication

Informed consent was obtained for all participants and/or their legal guardians for publication of identifying information/images according to the Portuguese Data Protection Authority (CNPD).

Competing interests

FT, PR, PT and JPB authors are employed by company CGC Genetics. All other authors declare no competing interests.

Author details

[1]Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, 4710-057 Braga, Portugal. [2]ICVS/3B's - PT Government

Associate Laboratory, Braga/Guimarães, Portugal. [3]CGC Genetics, Porto, Portugal. [4]Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, Porto, Portugal. [5]Center for Medical Genetics Dr. Jacinto Magalhães, Porto Hospital Center, Praça Pedro Nunes, Porto, Portugal. [6]Unit for Multidisciplinary Research in Biomedicine, Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, Porto, Portugal. [7]The Mindich Child Health & Development Institute and the Department of Genetics & Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [8]The Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [9]Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [10]Centro de Genética Preditiva e Preventiva - CGPP, Instituto de Biologia Molecular e Celular - IBMC, Universidade do Porto, Porto, Portugal. [11]Instituto de Investigação e Inovação em Saúde - i3S, Universidade do Porto, Porto, Portugal. [12]Unidade de Neurodesenvolvimento e Autismo do Serviço do Centro de Desenvolvimento da Criança and Centro de Investigação e Formação Clínica, Pediatric Hospital, Centro Hospitalar e Universitário de Coimbra, 3041-80 Coimbra, Portugal. [13]University Clinic of Pediatrics and Institute for Biomedical Imaging and Life Science, Faculty of Medicine, University of Coimbra, Coimbra, Portugal. [14]Medical Genetics Unit, Hospital de Braga, Braga, Portugal. [15]Department of Medical Genetics, Hospital de Faro, Faro, Portugal. [16]Pediatric Neurology Department, Centro Materno-Infantil Centro Hospitalar do Porto, Porto, Portugal. [17]Development Unit, Pediatrics Service, Hospital Centre of Cova da Beira, Covilhã, Portugal. [18]CICS - Health Sciences Research Centre, University of Beira Interior, Covilhã, Portugal. [19]Department of Pediatrics, Médio Ave Hospital Center, Vila Nova de Famalicão, Portugal. [20]Development Unit, Pediatrics Service, Hospital Centre of Cova da Beira, Covilhã, Portugal. [21]Department of Pediatrics, Hospital S. Teotónio, Tondela/Viseu Hospital Center, Viseu, Portugal. [22]Neuropaediatric Unit – Garcia de Orta Hospital, Almada, Portugal. [23]Pediatric and Neonatal Intensive Care, Department of Pediatrics, Porto Hospital Center, Porto, Portugal. [24]Department of Genetics, Hospital D. Estefânia, Lisboa-Norte Hospital Center, Lisbon, Portugal. [25]Genetics Service, Paediatric Department, University Hospital Santa Maria, Lisbon, Portugal. [26]Department of Pediatrics, Médio Ave Hospital Center, Santo Tirso, Portugal. [27]Division of Pediatric Neurology, Department of Child and Adolescent, Centro Hospitalar do Porto e Hospital de Santo António, Porto, Portugal. [28]Neuropsychophysiology Lab, CIPsi, School of Psychology, University of Minho, Braga, Portugal. [29]Department of Pathology, VU University Medical Center, Amsterdam 1007, MB, The Netherlands. [30]Department of Clinical Genetics, VU University Medical Center, Amsterdam 1007, MB, The Netherlands. [31]GDPN- SYNLAB, Porto, Portugal.

Received: 16 January 2019 Accepted: 12 June 2019
Published online: 05 July 2019

References

1. Reichenberg A, Cederlöf M, McMillan A, Trzaskowski M, Kapra O, Fruchter E, et al. Discontinuity in the genetic and environmental causes of the intellectual disability spectrum. Proc Natl Acad Sci. 2015;113(4):1098–103.
2. Coe BP, Girirajan S, Eichler EE. The genetic variability and commonality of neurodevelopmental disease. Am J Med Genet Part C Semin Med Genet. 2012;160 C(2):118–29.
3. 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(5):749–64.
4. 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(20):2044–54.
5. Lopes F, Torres F, Soares G, van Karnebeek CD, Martins C, Antunes D, et al. The role of AKT3 copy number changes in brain abnormalities and neurodevelopmental disorders: four new cases and literature review. Front Genet. 2019;10(February):1–8.
6. Lopes F, Torres F, Lynch SA, Jorge A, Sousa S, Silva J, et al. The contribution of 7q33 copy number variations for intellectual disability. Neurogenetics. 2018;19(1):27–40.
7. Lopes F, Soares G, Gonçalves-Rocha M, Pinto-Basto J, Maciel P. Whole gene deletion of EBF3 supporting Haploinsufficiency of this gene as a mechanism of neurodevelopmental disease. Front Genet. 2017;8(143):1–7.
8. Firth HV et al. Am J Hum Genet. 2009;84(4):524–33.
9. Dheedene A, Maes M, Vergult S, Menten B. A de novo POU3F3 deletion in a boy with intellectual disability and dysmorphic features. Mol Syndromol. 2014;5:32–5.
10. 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;23:2632–43.
11. 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(April):633–49.
12. Dudkiewicz M, Lenart A, Pawłowski K. A novel predicted calcium-regulated kinase family implicated in neurological disorders. PLoS One. 2013;8(6):1–10.
13. Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, Hill RS, et al. Identifying autism loci and genes by tracing recent shared ancestry. Science (80-). 2008;321(5886):218–23.
14. 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(4):560–72.
15. Wang KS, Liu XF, Aragam N. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. Schizophr Res. 2010; 124(1–3):192–9.
16. Silfhout ATV, Nakagawa T, Bahi-Buisson N, Haas SA, Hu H, Bienek M, et al. Variants in CUL4B are associated with cerebral malformations. Hum Mutat. 2014;36(Umr 8104):106–17.
17. Roy A, Gonzalez-Gomez M, Pierani A, Meyer G, Tole S. Lhx2 regulates the development of the forebrain hem system. Cereb Cortex. 2014;24:1361–72.
18. Flandin P, Zhao Y, Vogt D, Jeong J, Long J, Potter G, et al. Lhx6 and Lhx8 coordinately induce neuronal expression of Shh that controls the generation of interneuron progenitors. Neuron. 2011;70(5):939–50.
19. 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;21:1613–21.
20. 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(4):2485–500.
21. 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;80:345–52.
22. 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. Am J Med Genet Part A. 2009;152A:175–80.
23. 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(4):585–93.
24. 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(3):381–91.
25. 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(1):98–102.
26. Murali A, Rajalingam K. Small rho GTPases in the control of cell shape and mobility. Cell Mol Life Sci. 2014;71(9):1703–21.
27. 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; 26:247–50.
28. 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(12):822–6.
29. 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(2):180–6.
30. 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(3):e327.
31. 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(3):139–46.
32. Verdin H, De Baere E. FOXL2 impairment in human disease. Horm Res Paediatr. 2012;77:2–11.
33. 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(7):1294–306.
34. Lasalle JM. Autism genes keep turning up chromatin. OA Autism. 2013;1(2):14.
35. Patsialou A, Wilsker D, Moran E. DNA-binding properties of ARID family proteins. Nucleic Acids Res. 2005;33(1):66–80.
36. Clague MJ, Coulson JM, Urbé S. Cellular functions of the DUBs. J Cell Sci. 2012;1251 Clagu(Pt 2):277–86.
37. Hao Z, Zhang H, Cowell J. Ubiquitin-conjugating enzyme UBE2C: molecular biology, role in tumorigenesis, and potential as a biomarker. Tumor Biol. 2012;33(3):723–30.
38. Stouffer MA, Golden JA, Francis F. Neuronal migration disorders: focus on the cytoskeleton and epilepsy. Neurobiol Dis. 2016;92(Pt A):18–45.
39. Curatolo P, Bombardieri R, Jozwiak S. Tuberous sclerosis. Lancet. 2008; 372(9639):657–68.
40. Nodé-Langlois R, Muller D, Boda B. Sequential implication of the mental retardation proteins ARHGEF6 and PAK3 in spine morphogenesis. J Cell Sci. 2006;119:4986–93.
41. 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(13):2524–34.
42. 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 and Development. 2016; 38(1):124–7.
43. 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(11):772–5.
44. 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.
45. 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(38):13485–99.
46. Wu H, Wang X, Liu S, Wu Y, Zhao T, Chen X, et al. Sema4C participates in myogenic differentiation in vivo and in vitro through the p38 MAPK pathway. Eur J Cell Biol. 2007;86(6):331–44.
47. Quick MW. The role of SNARE proteins in trafficking and function of neurotransmitter transporters. Handb Exp Pharmacol. 2006;175:181–96.
48. van Bokhoven H. Genetic and epigenetic networks in intellectual disabilities. Annu Rev Genet. 2011;45(1):81–104.
49. Rippey C, Walsh T, Gulsuner S, Brodsky M, Nord AS, Gasperini M, et al. Formation of chimeric genes by copy-number variation as a mutational mechanism in schizophrenia. Am J Hum Genet. 2013;93(4):697–710.
50. Spielmann M, Lupiáñez DG, Mundlos S. Structural variation in the 3D genome. Nat Rev Genet. 2018;19(7):453–67.
51. van Ravenswaaij-Arts CMA, Kleefstra T. Emerging microdeletion and microduplication syndromes; the counseling paradigm. Eur J Med Genet. 2009;52(2–3):75–6.
52. 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(7):680–5.

Publisher's Note

SUPPLEMENTARY MATERIAL

12

13

PART 1 – Detailed clinical and results description

Patients, inclusion criteria, and clinical characterization

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 RC group, the eligibility criteria used were stricter. In order to be included in the study the patient needed to have (I) documented DD/ 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 syndrome 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. These patients were therefore considered to have idiopathic ID/DD. 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, and their information collected in a database in an anonymous manner. The database for clinical data collection was approved by the Portuguese national data protection committee (CNPD - Comissão Nacional de Proteção de Dados). In the CC, patients were enrolled from a private Genetics laboratory, and had been referenced as having ID, DD and/or congenital anomalies of unknown etiology. Written informed consent was obtained from the parents or persons in charge, for all the participants.

Measures

All participants were given an individually administered IQ test [Portuguese version of Wechsler Intelligence Scale for Children—Third Edition (WISC-III) (1); the participants' primary caregiver had been administered the Vineland Adaptive Behavior Scale (VABS)-Survey form (2), using a semi-structured interview format. All measures were administered by experienced psychologists during routine clinical multidisciplinary assessments. The WISC-III evaluation instrument consists of thirteen subtests (M = 10; SD = 3) spread over two subscales: Verbal and Performance, each one evaluating a different aspect of intelligence (3). The performance of the subjects in the various subtests is clustered in three composite results: a general intelligence measure (FSIQ) and two ratios divided by the nature of its subtests: the VIQ, measurement of verbal intelligence, and the PIQ, a non-verbal intelligence measure (3).

Results

Pathogenic CNVs

For most of these CNVs there are reports in the literature describing the phenotypic and genetic findings for similar patients; therefore, only some particular cases will be discussed here.

1p36.23-p36.21 deletion

Patient R1 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 (MIC), broad nasal bridge, hypoplastic nares, microretrognathia, kyphosis, hypertelorism and telecanthus (Supplementary figure S1). Large terminal deletions in 1p36 are known to cause the 1p36 deletion syndrome (4), nevertheless, interstitial deletions in this region are quite rarely described associated with NDDs.

This deletion affects 126 genes among which four (*PEX14*, *PLOD1*, *NMNAT1* and *MTHFR*) had previously been associated with disease (5) and eight (*MTOR*, *ENO1*, *PIK3CD*, *RERE*,

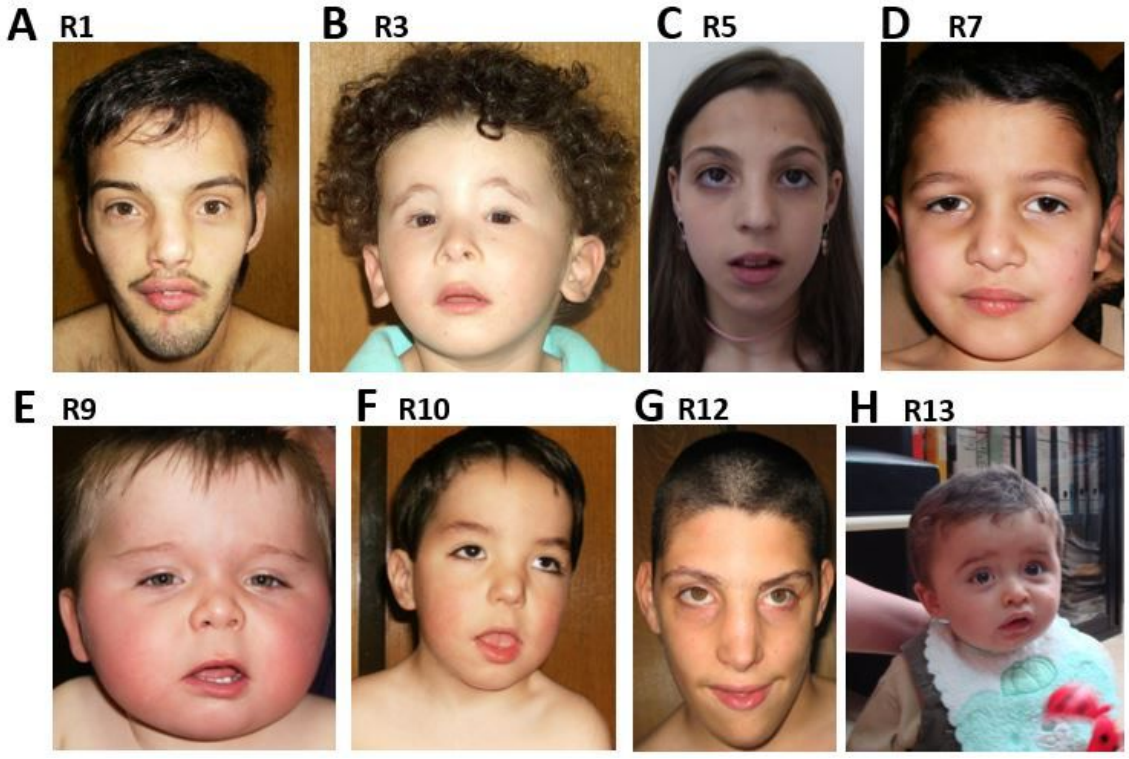
NPPA, *MAD2L2*, *MTHFR* and *KIF1B*) have a high haploinsufficiency score (6), possibly contributing for the phenotype. In the Decipher database (5), there are patients with similar overlapping deletions, of which three (266689, 248448 and 251601) also carry a *de novo* deletion and share similar clinical features, namely ID, MIC (patient 251601) and facial dysmorphisms.

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 (7). 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 neurons in the hippocampus, cardiovascular malformations, hearing loss and renal agenesis (8). More recently, *de novo* heterozygous mutations in *RERE* were described in 10 patients with NDDs/ID, suggesting that *RERE* haploinsufficiency might contribute to ID-related phenotypes (9). Taking this into account, the haploinsufficiency of *RERE* may contribute (alone or in conjunction with the haploinsufficiency of other surrounding genes) for the growth retardation, short stature (height and weight), MIC and ID observed in the patient.

MTOR (Mechanistic Target Of Rapamycin (Serine/Threonine Kinase)) was described as mutated in patients with Smith-Kingsmore (SK) syndrome (AD), characterized by ID, macrocephaly, dysmorphisms and small thoraces (10). MTOR mutations were also described in a patient with epileptic encephalopathy without brain malformation (AD) (11). Besides ID, the patient didn't present any major similarities with SK patients (apart from the alterations in head size which seem 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 others NDDs (Fragile-X syndrome, Down syndrome, Tuberous sclerosis and autism) (12), hence it is a plausible contributor to the phenotype of this patient.

3q22.1-q23 deletions (BPES)

Patient R3 is a 15 year old boy with mild ID (IQ= 54), brachycephaly, cortical atrophy, cardiac interventricular communication, low weight and hypoplastic genitalia with hypospadias. He also presents a peculiar eye dysmorphism characterized by microphtalmia, epicanthus and ptosis (Supplementary figure S1). 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) (13). In recent years it has been shown that large deletions affecting *FOXL2* and its surroundings are often associated with ID combined with BPES, and this patient reinforces the association of deletions in this region with that phenotype.



Supplementary figure 1 – Facial appearance of some patients carrying pathogenic variants. A, patient R1; B, patient R3; C, patient R5; D, patient R7; E, patient R9; F – patient R10; G, patient R12; H, patient R13.

7q11.23 deletions

Two non-related patients with 7q11.23 deletions were detected: C2 and R29.

Patient C2 is a 7 year old boy referred to the consultation due to DD, with previous normal investigations including high-resolution karyotype and Multiplex Ligation-dependent Probe Amplification (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, narrow central incisors, large ears and hyperacusis, narrow girdles, mainly scapular, and large digits. An echocardiography did not reveal any anomaly; presently, he is underweight and does not present behavioral problems. He has a 1.419Mb deletion at 7q11.23 that overlaps the Williams-Beuren syndrome (WBS) critical region and encompasses 30 genes, including the *ELN* (elastin), *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, which, with exception of the *ELN* gene, associated with cardiac problems, were suggested to be linked to the specific cognitive profile and craniofacial features presented by WBS patients (14). Additionally, this patient also carries a 457.8Kb duplication at 15q13.3, that encompasses the *CHRNA7* and *OTUD7A* genes (~30% of the 15q13.3 microdeletion syndrome). It was not possible to test his parents to determine the origin of the CNVs.

Within the same region of chromosome 7, we found another deletion in patient R29, a 19 year old boy with severe ID (IQ=21) whose mother is suspected to have mild ID. He has cerebral atrophy, MIC, epilepsy and recurrent respiratory infections. Behavioral alterations include motor stereotypies, increased activity, sleep disturbances, aggressiveness and frequent verbal preservative repetitions (saying “hello” and “caress”). He carries a 0.4 Mb maternal deletion at 7q11.23 affecting 11 genes, including *BAZ1B*, *STX1A* and *WBSCR22* but not ELN. A neurocognitive assessment of the patient was performed as described previously for WBS

patients (1,15), however, the patient was unable to understand most of the questions or requirement of several tasks, indicating a more severe ID than that expected in a classical WBS patient (16). In fact, he presents concurrent deficits in adaptive functioning, matching a profound ID classification (FSIQ – <20), with severe impairments in verbal comprehension, performance and working memory. This seems unlikely to be caused only by the 7q11.23 deletion.

22q13.3 deletions (Phelan-McDermid syndrome)

Patient C7 is an 8 year 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 behavioral alterations or dysmorphic features. He was thought to have neonatal hypotonia, as he was only able to sit alone at the age of 9 months. Presently, he has mild hypotonia, regular growth, does not have dysmorphic features and wears glasses, and is being followed in the strabismus/ophthalmological consultation. He has also cognitive deficits, scoring below the average in all GMDS sub-scales (locomotor, personal-social, language, eye and hand co-ordination, performance, practical reasoning). aCGH analysis showed that he has a 1.66Mb deletion at 22q13.3 that encompasses 39 genes including the SH3 and multiple ankyrin repeat domains 3 (*SHANK3*) gene, that encodes a multidomain scaffold protein (17) involved in synapsis (18). *SHANK3* is located within the minimum critical region of the Phelan-McDermid syndrome (PMS). Mutations in this gene are associated with several NDDs, including autism spectrum disorder (ASD), presented by more than 50% of the PMS patients (19), schizophrenia and bipolar disease, neurodevelopmental deficits (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) (19). Patient C7 presents some clinical overlap with this presentation, namely the hypotonia, speech impairment and motor development delay, but no significant facial dysmorphisms. The severity

of symptoms from PMS patients were recently 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* (20). This variability in the extent of the mitochondrial dysfunction could contribute for the incomplete overlap of patient's phenotype with the one previously described.

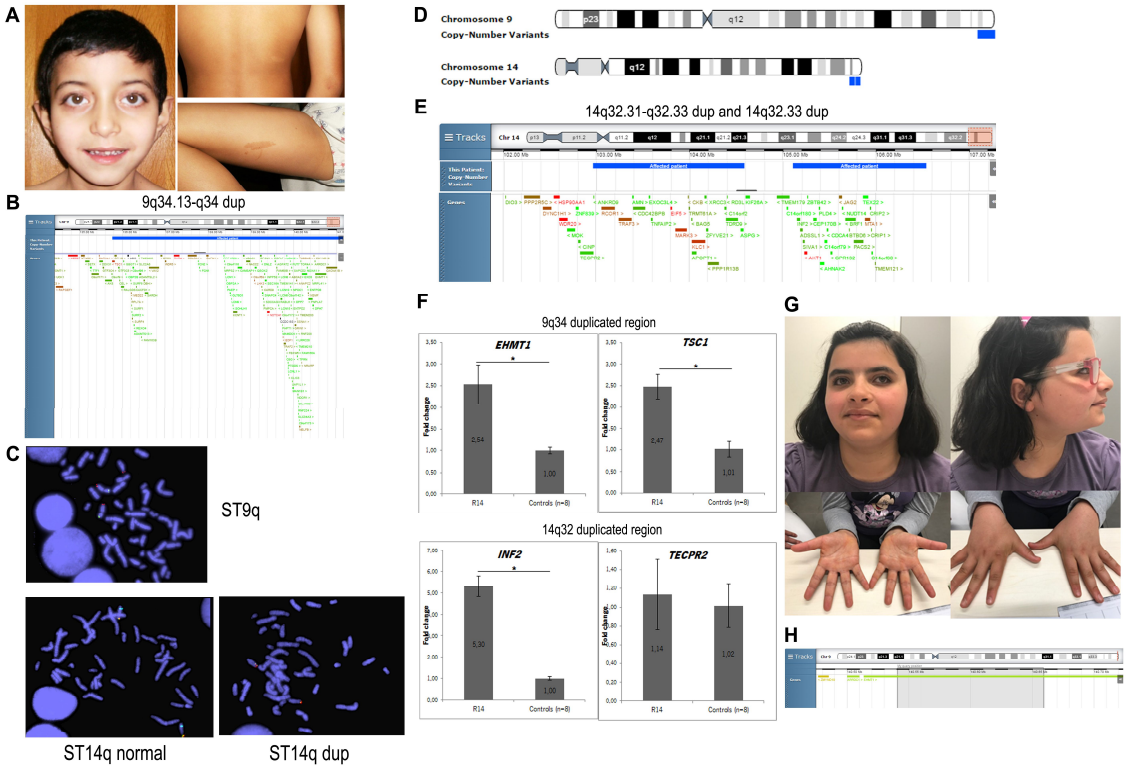
9q34, 14q32.31-q32.33 and 14q32.33 duplications

9q34 duplications were detected in two unrelated patients: C19 and R14 (Supplementary figure S2). Patient C19 is a 12 year old girl with moderate ID, obesity (62.4Kg at 12 years, BMI=26.6 Kg/m^2, P97), hypotonia and facial dysmorphisms, namely a coarse facies, midface hypoplasia (flat face appearance), synophrys, upslanting palpebral fissures, anteverted nares, downturned mouth and everted lower lip. She has a hoarse voice and speech articulation deficits. She presents behavioral problems, namely ADHD, medicated with methylphenidate, an obsessive-compulsive behavior towards food, and sleep disturbances (difficulties in falling asleep). Currently, her EEG analysis is normal, although she had one unprovoked seizure episode. She carries a maternal 118Kb 9q34.3-duplication from, at least, intron 2 to intron 11 of the *euchromatic histone-lysine N-methyltransferase 1* (*EHMT1*) gene. Haploinsufficiency for *EHMT1* causes Kleefstra syndrome (KS, MIM 610253), commonly through deletions or, more rarely, point mutations affecting this gene (21). Until now, only one case of KS caused by an intragenic *EHMT1* duplication was described (22), the duplication being identical to the one presented by this patient; additionally, patient C19´s clinical presentation overlaps the KS core phenotype (21). 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 a mosaic for the deletion (23). Patient C19's mother also

carries the duplication, though not in mosaic state, and has confirmed psychiatric/cognitive problems.

Patient R14 is a 16 year old boy with mild ID (IQ= 57), facial dysmorphisms, hypochromic and *café-au-lait* spots. Behaviorally he presents stereotypies, obsessive and aggressive behavior and ADHD. He carries a rearrangement between chromosome 9 and chromosome 14 in which the 9q duplicated region is located in 14p. Moreover, the duplicated 14q32.31-q32.33 and 14q32.33 regions are likely 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) (Supplementary figure S2). This resulted in three *de novo* duplications: a 5.5Mb 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. All of these might possibly contribute to his phenotype, making it difficult to ascertain the specific role of each imbalance.

Duplications affecting the 5' region of the *EHMT1* gene and duplications or triplications encompassing the entire *EHMT1* gene have been observed in patients with neurodevelopmental impairment, speech delay, and ASD, suggesting that increased *EHMT1* dosage is associated with a neurodevelopmental phenotype (24). Interestingly, the expression of *EHMT1* and *TCS1* (for the 9q dup) was found to be increased in the peripheral blood of patient R14 when compared to controls. Since FISH studies have shown that the 9q34 duplicated region is not located in tandem, but in *trans* (Supplementary figure S2-C), most likely it will not originate any structural effect that could influence the expression of the genes within the region. As for the 14q32 duplicated region: I) the *INF2* mRNA expression was increased in the patient, in agreement with the fact that the entire gene is located inside the duplicated region; II) *TECPR2* expression was not altered. 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 (Supplementary figure S2).



Supplementary figure 2 – Clinical features of patients R14 and C19 and images of their CNVs. (A) appearance of patient R14 (note the hyperpigmented spots); (B) schematic representation of the 9q31.13-q34 duplicated region in patient R14 (the blue bar indicates the duplicated region, adapted from DECIPHER); (C) FISH analysis for patient R14 (9q31.13-q34 – D9S325 - duplication in the top panel; 14q32 – D14S1420 - duplications in the lower panel) revealed the presence of a derivative chromosome in which the duplicated 9q region is located in the 14p. It was not possible to determine if the duplicated 9q portion is located in the duplicated or normal chromosome 14. The duplicated 14q region must be located in tandem. (D) Schematic representation of the duplicated regions in patient R14. (E) Scheme of both duplication in chromosome 14. (F) Expression pattern of *EHMT1*, *TSC1*, *TECPR2* and *INF2* genes. (G) Facial appearance of patient C19. (H) Duplicated 9q region in patient C19. (B2M was used as housekeeping gene; student t-test; $p<0.05$*).

Likely pathogenic CNVs

In this category were included only CNVs whose speculation of pathogenicity is supported in the literature, such as large CNVs that comprise good candidates for disease association. We provide here additional information concerning some CNVs detected.

17p11.2 deletions

There are other genes that could contribute for the phenotype of patient C15.
EPN2 encodes for epsin2, a protein thought to be involved in clathrin-mediated endocytosis, found in a brain-derived clathrin-coated vesicle fraction (25). *RNF112* (ring finger protein 112, also known as *ZNF179*), which encodes a member of the RING finger protein family of transcription factors and is primarily expressed in brain (26) 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* (*C. elegans*) (27). Moreover, there are 3 OMIM genes affected by these deletions: *B9D1*, which encodes a B9 domain-containing protein, associated both with Meckel (28) and Joubert syndromes (both AR) (29); *ALDH3A2*, which encodes for a fatty aldehyde dehydrogenase that causes Sjögren-Larsson syndrome (AR) (30) when mutated and *AKAP10*, which encodes the mitochondrial A-kinase (PRKA) anchor protein 10, which may be associated with increased risk of arrhythmias and sudden cardiac death (31). All these being recessive disorders, the presence of these deletions in heterozygosity is unlikely to cause the phenotype. Nevertheless, a phenotype modifying effect cannot be excluded.

20q13.12-q13.13 deletions

Patient R20 carries a deletion that encompasses several genes, among which the genes *KCNB1, PIGT, CTSA, SLC2A10* and *ARFGEF2* were associated with human disease, whereas *MMP9, CSE1L* and *YWHAB* have very high haploinsufficiency scores in Decipher. *KCNB1* (Potassium Channel, Voltage Gated Shab Related Subfamily B, Member 1) encodes a potassium channel in which *de novo* mutations were found in patients with epilepsy and NDDs (32). 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 (33). Homozygous mutations in this gene have been described in patients with epilepsy, periventricular heterotopia with MIC, movement disorders and ID (34,35). Even though the clinical presentation is not completely overlapping (epilepsy not being a feature of disease in this case), 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 within the deleted region cannot be excluded.

9q33.2-q33.3 triplication

Patient R21 carries a 3.6 Mb *de novo* triplication at 9q33.2-q33.3 that affects 60 genes. *NEK*6 and *PSMB7* were also selected for mRNA expression studies. *NEK6* encodes a protein kinase required for efficient mitotic spindle assembly (36) and found to be a key player in aging and cancer (37) *PSMB7* encodes a proteasome subunit thought to play a role in autophagy inhibition in cardiomyocytes (38). Besides these, two other genes also draw our attention: *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 (39); *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 (40). This CNV may therefore deregulate vesicle cycling and endocytosis in both neurons and in bone.

CNVs of unknown significance

In six cases, the CNV arised *de novo* in the patient; however, this does not reflect the real number of de novo CNVs but instead reflects the unavailability of parent's samples to test inheritance. Several patients presented more than one CNV (6 patients with 2, 1 patient with 3 and 3 patients with 4). In these cases, in which the patient had CNVs inherited from both parents, there is the question of whether these alterations together could lead to the disease in the child through epistatic effects.

Supplementary Table S1 – Patients with altered aCGH results (i.e. with CNVs classified as non-polymorphic).

	Clinical cohort (CC)	**Research cohort (RC)**	**Both cohorts**
Number of patients	137	188	325
Pathogenic	12 (8.8%)	14 (7.4%)	26 (8%)
Likely pathogenic	8 (5.8%)	9 (4.8%)	17 (5.2%)
VOUS	24 (17.5%)	31 (16.5%)	55 (16.9%)
Total with CNVs	**44 (32.1%)**	**54 (28.7%)**	**98 (30.1%)**

Supplementary Table S2 – List of variants of unknown clinical significance (VOUS)

Patients	Gender	Alteration (Hg19)	Type	Size (Kb)	Genes (nº)	Genes (name)	Confirmation	Inheritance	DGV controls	Similar case (Decipher)	Array platform
R24	Male	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
R25	Male	1p34.3(36,775,225-369,17,965)X3	dup	143	8	*LSM10, OSCP1, SH3D21, STK40*	NP	ND	1/270 (del)	No	1
C21	Female	1q44(245,132,097-245,259,567)x3	dup	127.5	1	*EFCAB2*	NP	ND	6/46505	265207	2
C22,C23,C24§	Male, Male, Male	2p15(61,377,041-61,522,171)x3mat	dup	14	3	*C2orf74; AHSA2; USP34*	qPCR	Maternal	No	256542; 279248 (del)	2
R26	Male	2q11.2(96,735,183-98,228,265)X3mat	dup	1496	24	*ARID5A, NEURL3, SEMA4C*	NP	Maternal¥	1/6533	254924, 274288	1
C25	Male	2q12.3q13(109,269,051-110,504,320)x3	dup	1240	7+2	*LIMS1 (PINCH); RANBP2; EDAR; CCDC138; SH3RF3; SEPT10; SOWAHC; MIR4265; SH3RF3-AS1*	NP	ND	No	263424	3
C26	Male	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
C27	Female	2q31.2q31.3(179,933,642-180,709,664)x3	dup	776	2	*SESTD1; ZNF385B*	NP	Paternal¥	No	No	3
R27	Female	2q33.1(203,338,376-203,513,494)X3	dup	175	2	*BMPR2, FAM117B*	NP	ND	No	No	1
C28	Male	3q23(141,021,128-141,154,103)x3	dup	130	1	*ZBTB38 (CIBZ)*	NP	ND	2/18978 (smaller)	No	3
R28	Male	3q26.33(181,357,672-181,466,211)X1	del	108	2	*SOX2; SOX2OT*	qPCR	*de novo*	No	301549 (smaller dup)	1
C29	Female	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
C30	Male	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

C31	Male	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
C32	Male	5q31.1(132,540,588-132,890,257)x3	dup	349.7	2	*FSTL4; MIR1289*	NP	ND	1/17421	No	3
C33	Male	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
R29	Male	7q11.23(72,741,861-73,145,916)x1mat‡	del	0.4	11	*BAZ1B, STX1A, WBSCR22*	qPCR	Maternal	No	No	1
C34	Male	7q36.1(151,768,386-152,077,451)x3	dup	309.1	2	*KMT2C (MLL3); GALNT11*	NP	Paternal¥	4/29957	248263	3
C35	Male	8p23.1(9,687,615-10,112,430)x3	dup	424.8	3	*LINC00599; MIR124; MSRA*	NP	Paternal¥	No	No	2
R30	Male	9p24.1(6,668,082-7,112,536)X1	del	444	1	*KDM4C*	NP	ND	2/ 29084	303548 (smaller)	1
R31	Male	10p11.23(30,659,736-30,761,192)X3	dup	101	7	*MAP3K8*	qPCR	Paternal¥	1/17369	No	1
R32	Male	10q24.2(100,014,985-100,100,985)X3	dup	86	1	*LOXL4*	NP	ND	No	No	1
R33	Male	11p11.2(44,601,486-44,779,120)X3	dup	178	1	*CD82*	NP	ND	No	No	1
R34	Male	11q23.3(119,415,826-119,560,414)X3	dup	145	1	*PVRL1*	NP	ND	No	No	1
R35	Male	12p13.33p13.32(2,248,863-3,497,525)X3pat	dup	1248	9	*CACNA1C, TULP3, FOXM1, TSPAN9*	qPCR	Paternal	2/29084 (smaller)	No	1
R36	Male	13q32.3(99,451,824-99,530,240)X3	dup	78	3	*DOCK9*	NP	ND	1/29084 1/3017 1/10	No	1
R37	Male	14q24.2(71,814,635-71,927,259)X3	dup	113	1	*SNORD56B*	NP	ND	No	No	1
C36	Male	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
C2	Male	15q13.3(31,985,493-32,443,078)x3	dup	457.8	2	*CHRNA7, OTUD7A*	NP	ND	4/181 25/771 1/17421	248307, 248334, 257139,	2

									37/29084	262435, 262984, 337131	
C37	Male	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
C38	Fenale	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
C39	Male	17p13.3(2,287,362-2,287,555)x3	dup	194	1	*MNT*	NP	ND	5/46874 (bigger)	No	2
R38	Male	17p13.1(6,955,115-7,409,331)X1	del	454	30	*DLG4, GABARAP, DULLARD, NEURL4, NLGN2, CHRNB1*	NP	*de novo*	3/19159 (0,000052)	260507, 2346, 3474 all smaller del)	1
R39	Female	17q11.2(27,429,294-27,516,778)X3	dup	87	1	*MYO18A*	qPCR	Paternal¥	No	No	1
R40	Male	17q21.31(43,696,388-43,979,132)X3	dup	283	7	*C17orf69, CRHR1, IMP5, LOC100128977, LOC100130148, MAPT, MGC57346*	NP	ND	2/17421 (smaller)	No	1
R41	Male	19p13.2(7,077,066-7,727,437)X3	dup	605	14	*KHSRP, PSPN, TUBB4, STXBP2*	NP	Maternal	7/29855 (0,00023)	253443	1
		19p13.3-p13.2(6,332,716-6,993,284)X3	dup	660	22	*PNPLA6, ARHGEF18, KIAA1543*	NP	Maternal	3/181 (smaller; half the size)	No	
R42	Female	19q13.12(37,775,477-37,942,465)X3	dup	167	3	*HKR1, ZNF527, ZNF569*	NP	ND	1/270 (del)	No	1
C40	Male	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	Male	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	Female	Xp11.21-p11.1(56,304,820-56,964,968)X3	dup	660	4	*KLF8, UBQLN2*	NP	ND	No	274061 (half)	1
C41	Female	Xp21.1(32,826,352-33,936,518)x3	dup	1110	1	*DMD*	NP	Maternal¥	No	No	3
C42	Female	Xp22.31(6,440,776-8,135,568)x3	dup	1.695	7	*VCX3A, HDHD1, STS, VCX, PNPLA4*	qPCR	ND	1/873 (smaller)	Several	4
C43	Female	Xp22.33(480,164-785,059)x3	dup	304.9	1	*SHOX*	NP	ND	No	279033 (but has more	3

										variants)	
R45	Male	Xq28(153,230,586-153,282,378)X2	dup	52	3	*HCFC1, IRAK1*	NP	ND	1/265 (del)	No	1
C44	Female	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	
R46	Male	5q23.3(128,758,178-129,350,165)X3	dup	592	2	*ADAMTS19, CHSY3*	NP	ND	2/18530 (0,0001)	No	1
		9q33.3(128,474,074-128,515,941)X1	del	419	1	*PBX3*	NP	ND	No	No	1
R47	Male	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¥	2/18978 (0,0001)	248891)	1
R48	Male	3p21.31(48,464,967-48,574,235)X3	dup	109	7	*ATRIP, CCDC51, CCDC72, PFKFB4, PLXNB1, SHISA5, TREX1*	NP	Maternal¥	1/ 17421 (del) 1/2026 (del)	No	1
		7p22.3(1,565,982-1,701,871)X3	dup	136	5	*KIAA1908, MAFK, PSMG3, TFAMP1, TMEM184A*	NP	Paternal	1/270 (del)	No	1
		10p12.31(20,641,191-21,122,699)X3	dup	481	2	*NEBL, PLXDC2*	NP	Maternal¥	No	250836 (smaller)	1
		14q31.3(88,794,387-88,853,440)X3	dup	59	1	*SPATA7*	NP	*de novo*	No	No	1
R49	Female	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 (smaller) 1/3017 (0,0003)	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 ¥	2/29084 (0,00007)	No	1
R50	Male	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	Female	5p15.1(16,112,927-16,260,219)X3	dup	147	1	*MARCH11*	NP	ND	No	289370	1
		13q31.1(84,644,861-84,723,563)X3	dup	79	1	*MIR548F1*	NP	ND	1/29084 (0,00003)	No	1
R52	Male	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	*ARHGAP4, AVPR2, HCFC1, IRAK1, L1CAM, NAA10, RENBP, TMEM187*	NP	ND	1/265 (del)	323738	1
R53	Female	10q25.3(118,404,726-119,052,432)X3	dup	647	7	*C10orf82, HSPA12A, KCNK18, KIAA1598, PDZD8, 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; 1/2026 (smaller)	No	1
		Xp22.33(1,549,311-1,641,335)X3	dup	92	2	*ASMTL, P2RY8*	NP	ND	1/2 (del)	306493, 288916, 288492, 288579	1
R54	Female	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 R24 to R54: from research cohort; Patients C21 to C44: 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; §: siblings; Cut-off for inclusion of a CNV in this table: <1/5000 in DGV.**

PART 2 – Detailed methodology

aCGH

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 other Agilent platforms: according to manufacturer instructions; 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 (GEO accession number GPL15397), 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).

CNV classification criteria

The genomic variants detected were classified using adapted criteria, previously described elsewhere (41–44).

A CNV was classified as pathogenic when a high degree of certainty of their clinically significance is present in the literature. This group includes patients carrying large CNVs that overlap significantly with a region with an established pathogenic effect. This classification was also applied to susceptibility *loci* of variable expressivity and incomplete penetrance (reviewed in (45)).

Likely pathogenic variants include newly described gene rich large CNVs that comprise good candidates for disease association. In this category were included only CNVs whose speculation of pathogenicity is supported in the literature (for example, there is another similar patient described or it includes gene(s) with a compelling function). However, the uncertainty of this claim stills remains. It is possible that variants in this category will be latter classified as pathogenic (with the report of other similar CNVs in patients with overlapping phenotype) or as benign (for example, if the variant is described later on in several unaffected cases).

Variants of unknown clinical significance (VOUS) include variants whose clinical significance is not yet possible to speculate, due to the fact that: 1) there is a lack of overlapping CNVs reported in the literature and/or databases; 2) the CNV contain genes but it is not yet known whether they are dosage sensitive; 3) the CNV is described in multiple contradictory publications and/or databases, and firm conclusions regarding clinical significance are not yet established (44).

Quantitative PCR confirmations

Quantitative PCR reactions were carried out in a 7500-FAST Real Time PCR machine (Thermo Fisher Scientific, Waltham, MA, USA) using Power SYBR Green® (Thermo Fisher Scientific, Waltham, MA, USA), as described elsewhere (46) and following the general recommendations for qPCR (47,48). The specificity of each reaction was verified by the generation of a melting curve for each of the amplified fragments. The primer efficiency was calculated by generation of a standard curve fitting the accepted normal efficiency percentage (primers used for all genes are listed in table S3). Ct values obtained for each test were analyzed in DataAssist™ software (Thermo Fisher Scientific, Waltham, MA, USA).

mRNA expression analysis

Total RNA was isolated from leucocytes of patients and controls (ten healthy controls, five females and five males) using QIAsymphony RNA Kit (QIAGEN GmbH, Germany), according to the manufacturers' protocol. First-strand cDNA, synthesized using SuperScript® III Reverse

Transcriptase (RT) (Thermo Fisher Scientific, Waltham, MA USA). The genes selected to study within each CNV 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 – genes without or with very low peripheral expression were not analyzed [data retrieved from GeneCards database (www.genecards.org) and GTExportal (http://www.gtexportal.org)]. Primers used for all genes are listed on table S4. Quantitative PCR reactions were carried out in a 7500-FAST Real Time PCR machine (Thermo Fisher Scientific, Waltham, MA, USA) using Power SYBR Green® (Thermo Fisher Scientific, Waltham, MA, USA). The expression levels of the genes were normalized to the *B2M*, B-ACTIN, *TRAP1* or *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 (49,50).

Methylation status

Methylation status for *SNRPN* gene (*locus* 15q11-13) was studied by MLPA for patient C10 using the MLPA kit ME028-C1 (MRC Holland) in accordance with manufacturer instructions.

Supplementary Table S3 – Primers used for quantitative PCR confirmation.

Chromosome	Gene	Reference sequence	Primer location	Primer Forward 5'→3'	Primer Reverse 5'→3'	Amplicon size (bp)
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 1	*MAST2*	NM_015112.2	Exon3	AGCTGCTCCCTTTGTCCAG	GCCACCTTTATGAACACTTACCAG	158bp
Chr 1	*PRKAB2*	NM_005399.4	Exon4	AGCCATAATGACTTTGTTGCCA	GCCCATCAGTCTTGACAGAAA	174bp
Chr 1	*FAM69A*	NM_001006605.4	Exon3	AGACTGGAGTTATTGATGGGC	CTGGAATGTTTATTCATAATGGC	130bp
Chr 2	*GPR45*	NM_007227.3	Exon1	ACGTCCCTTGAGGCTTACAC	ACGATGATGCAGACCACAGT	161bp
Chr 2	*ARHGEF4*	NM_015320.3	Exon14	TTCTGGCACAGCATCAGC	CACTTGCAGGCAGAGGAAG	144bp
Chr 2	*USP34*	NM_014709.3	Exon80	ATGAAGGAGCAACTCCCATT	GCTCAGTTCCTGGATCAATAAT	168bp
Chr 3	*SOX2*	NM_003106.3	Exon1	CCCACCTACAGCATGTCCTA	CTGATCATGTCCCGGAGGT	164bp
Chr 3	*ZNF80*	ENSG00000174255	Exon1	GCTACCGCCAGATTCACACT	AATCTTCATGTGCCGGGTTA	182bp

Chr 7	*CNOT4*	NM_001190850.1	Exon10	CACCGAGCGGTTTATAATTCA	AGACCTGTGTTGTGCTGTGG	164bp
Chr 7	*OCM*	NM_001097622.1	Intron1/exon2	CTCTGTTCTTCAGACCCAGACA	GCTTACTTAAGCTCTTCTTCATCCA	152bp
Chr 7	*CALD1*	NM_033139.3	Exon4	GAATGACGATGATGAAGAGGAG	ACAGTACCTGTTCTGGGCATTC	139bp
Chr 7	*BAZ1B*	NM_032408.3	Exon3	TCCTGCCTGGTATGAGAAGC	TCCCACAGCATATTTGGTCA	112bp
Chr 9	*LHX2*	NM_004789.3	Exon3	GCTCGGGACTTGGTTTATCA	GTTGAAGTGTGCGGGGTACT	156bp
Chr 9	*ZNF658*	NM_001317916.1	5'UTR	ACCTCTTTGGTATAAACGTTCCAT	AGGACAGGGAGTCACATCTCTC	119bp
Chr 10	*MAP3K8*	NM_005204.3	Exon3	TGGAGTACATGAGCACTGGAA	TTGACACATGGTCATTAGACTGG	152bp
Chr 10	*EBF3*	NM_001005463.2	Intron15	CTCTCTGCTGGGTGCTGAG	GCGTCCCTTCATACGCTAAC	169bp
Chr 11	*KIRREL3*	NM_032531.3	Exon17/3'UTR	GATGCAGACTCACGTCTAAGGA	CTTGATCAGAGCTTCGAAGGAA	179bp
Chr 12	*TULP3*	NM_003324.4	Exon3	GGCTACTACTTGAGAAGAGGCAAA	TGACATTGCTGTGGGGAGTA	150bp
Chr 12	*MED13L*	NM_015335.4	Exon3/Intron3	GGAAGAAGGACTCTGGGAAAA	CAGGAAACTCTCGGTATCTAGCA	151bp
Chr 15	*SNRPN*	NM_003097.3	5'UTR	CTTTCCTGTCTGTCATTTTGC	GTCCCTTCTCTGTGCAGC	160bp
Chr 17	*MYO18A*	NM_00134765.1	Exon2	TGTCAAGCGCTTTTCCTTCT	AGAGTCCTCACCTCCACCTG	111bp
Chr 20	*SDC4*	ENSG00000124145	Exon4	ACCGAACCCAAGAAACTAGA	GTGCTGGACATTGACACCT	101bp
Chr 20	*EBF4*	NM_001110514.1	Exon16	GCTGCCTCCTCCATGTCC	AAGGCGCTCCTCTGTTTGAC	101bp

Chr X	*ARHGEF6*	NM_004840.2	3'UTR	CTTGAAATGTCCCGCTGAAT	AACAACAGCAAATGCCCAAG	162bp
Chr X	*PNPLA4*	NM_001142389.1	Exon6	CACCAACGCTCTTCCCAT	CACCATGATATCCTGCTTGG	136bp
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

Supplementary table S4 – Primers used for expression studies.

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 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 14	*TECPR2*	NM_014844.3	GGGGAAGACGGAATCTATCA	Exons2/3	GTGACATCAAATCTCCGAAGCT	Exons3/4	149bp
Chr 14	*INF2*	NM_022489.3	GACCACTTCTACCTCCTCCTG	Exon12	TGAGGAAGTTCCCAATTCTC	Exon14	201bp
Chr 15	*B2M*	NM_004048	GATAGTTAAGTGGGATCGAG	Exon 2/3	GCAAGCAAGCAGAATTTGGA	Exon4	93bp
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

PART 3 – Genes within likely pathogenic CNVs: data retrieved from OMIM, Decipher, ClinVar and ExAC databases (6,51)

Supplementary table S5 - OMIM entrance, haploinsufficiency score and constrain metrics for the selected genes in patient R16.

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); DDG2P: Developmental Disorders Genotype-Phenotype Database; 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 constrain metrics for the selected genes in patient C15.

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,	50-60%	Probable	29dels/28dups/23SNVs	-0.27	0.20	0.12	0.42

		Joubert syndrome 27							
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); DDG2P: Developmental Disorders Genotype-Phenotype Database; 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 constrain metrics for the selected genes in patient R20.

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); DDG2P: Developmental Disorders Genotype-Phenotype Database; 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 constrain metrics for the selected genes in patient C16.

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); DDG2P: Developmental Disorders Genotype-Phenotype Database; 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 constrain metrics for the selected genes in patient R21.

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); DDG2P: Developmental Disorders Genotype-Phenotype Database; 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 constrain metrics for the selected genes in patient C19

9q34.3 duplication	List of all the genes affected	*EHMT1*							
Gene	Morbid gene	OMIM	% HI score	DDG2P	ClinVar	Constraint Metrics			
						Synonymous (z)	Missense (z)	LoF (pLI)	CNV (z)
EHMT1	Yes	610253, Kleefstra syndrome	60-70%	Yes	48dels/49dups/3indels/181SNVs/9ins	0.20	2.36	1	-0.38

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); DDG2P: Developmental Disorders Genotype-Phenotype Database; 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 constrain metrics for the selected genes in patients R22 and R23.

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); DDG2P: Developmental Disorders Genotype-Phenotype Database; 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 S12 - OMIM entrance, haploinsufficiency score and constrain metrics for the selected genes in patient C20.

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); DDG2P: Developmental Disorders Genotype-Phenotype Database; 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.

BIBLIOGRAPHY

1. Wechsler D. Wechsler Intelligence Scale for Children—Third Edition. The Psychological Corporation; 1991.
2. Sparrow S, Balla D, Cicchetti D. The Vineland Adaptive Behavior Scales: Interview edition, survey. In: Major psychological assessment instruments. 1984. p. 199–231.
3. Wechsler D. Wechsler intelligence scale for children – Fourth edition (WISC-IV) administration and scoring manual. 2003b, San Antonio, TX Psychol Corp.
4. Jordan VK, Zaveri HP, Scott DA. 1P36 Deletion Syndrome: an Update. Appl Clin Genet [Internet]. 2015;8:189–200. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4555966&tool=pmcentrez&rendertype=abstract
5. Firth H V., 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 [Internet]. 2009;84(4):524–33. Available from: http://dx.doi.org/10.1016/j.ajhg.2009.03.010
6. Huang N, Lee I, Marcotte EM, Hurles ME. Characterising and predicting haploinsufficiency in the human genome. PLoS Genet. 2010;6(10):1–11.
7. 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(18):953–7.
8. 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(2):e57460.
9. 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(5):963–70.
10. 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(7):1659–67.
11. 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(7466):217–21.
12. Costa-Mattioli M, Monteggia LM. mTOR complexes in neurodevelopmental and neuropsychiatric disorders. Nat Neurosci. 2013;16(11):1537–43.
13. Verdin H, Baere E De. FOXL2 Impairment in Human Disease. Horm Res Paediatr. 2012;77:2–11.
14. 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 [Internet]. 2010;18(1):33–8. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2987159&tool=pmcentrez&rendertype=abstract
15. Simões MR, Albuquerque CP, Pinho MS, Pereira M, Seabra-Santos M, Alberto I. Bateria de Avaliação Neuropsicológica de Coimbra (BANC). Cegoc;
16. Mervis CB, Morris CA, Klein-Tasman BP, Bertrand J, Kwitny S, Appelbaum LG, et al. Attentional characteristics of infants and toddlers with Williams syndrome during triadic interactions. Dev Neuropsychol. 2003;23(1–2):243–68.
17. Sheng M, Kim E. The Shank family of scaffold proteins. J Cell Sci. 2000;113:1851–6.
18. 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(1):26–32.
19. Phelan K, Mcdermid HE. The 22q13.3 Deletion Syndrome (Phelan-McDermid Syndrome). Mol Syndromol. 2012;2(3–5):186–201.
20. 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 [Internet]. 2016;6(19544):1–12. Available from: http://dx.doi.org/10.1038/srep19544
21. 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(79):370–7.
22. 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(74):3–7.
23. 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;80:31–8.
24. 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(12):1895–910.
25. 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(48):33959–65.
26. 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.
27. Tomoda T, Kim JH, Zhan C, Hatten ME. Role of Unc51.1 and its binding partners in CNS axon outgrowth. Genes Dev. 2004;18(5):541–58.
28. 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(13):2524–34.
29. 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(72):1–4.
30. 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;12:52–7.
31. Neumann SA, Tingley WG, Conklin BR, Shrader CJ, Peet E, Muldoon MF, et al. AKAP10 (I646V) functional polymorphism predicts heart rate and heart rate variability in apparently healthy, middle-aged European-Americans. Psychophysiology. 2009;46(3):466–72.
32. 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(4):529–40.
33. 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(1):69–76.
34. 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(1):124–7.
35. 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(11):772–5.
36. Prosser SL, O'Regan L, Fry AM. Novel insights into the mechanisms of mitotic spindle assembly by NEK kinases. Mol Cell Oncol. 2016 May;3(3):e1062952.
37. Hirai Y, Tamura M, Otani J, Ishikawa F. NEK6-mediated phosphorylation of human TPP1 regulates telomere length through telomerase recruitment. Genes to Cells Devoted to Mol Cell Mech. 2016 Aug;21(8):874–89.
38. Kyrychenko VO, Nagibin VS, Tumanovska L V., 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(1):8–14.
39. 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(51):13202–12.
40. 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;15:2218–29.
41. Lee C, Iafrate AJ, Brothman AR. Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. Nat Genet. 2007;39(July):48–54.
42. Leung TY, Pooh RK, Wang CC. Classification of pathogenic or benign status of CNVs detected by microarray analysis. Expert Rev Mol Diagn. 2010;10(6):717–21.
43. 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(5):749–64.
44. 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(7):680–5.
45. 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(2):73–90.
46. 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. 2005 Jan;85(1):24–33.
47. Bustin SA, Benes V, Garson J, Hellemans J, Huggett J, Kubista M, et al. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clin Chem. 2009;55(4):611–22.
48. D'haene B, Vandesompele J, Hellemans J. Accurate and objective copy number profiling using real-time quantitative PCR. Methods [Internet]. 2010;50(4):262–70. Available from: http://dx.doi.org/10.1016/j.ymeth.2009.12.007
49. 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(5):345–55.
50. 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 [Internet]. 2010;50(4):S1–5. Available from: http://dx.doi.org/10.1016/j.ymeth.2010.01.005
51. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, et al. A framework for the interpretation of de novo mutation in human disease. Nat Genet. 2014;46(9):944–50.

Sub-Chapter 2.2 – The role of *AKT3* copy number changes in brain abnormalities and neurodevelopmental disorders: four new cases and literature review

Disclaimer:

The results presented in this sub-chapter refer to the collection of four unrelated patients with CNVs affecting *AKT3* gene, and were published in the international peer-reviewed journal ***Frontiers in Genetics*** under the following title and authorship:

The role of *AKT3* copy number changes in brain abnormalities and neurodevelopmental disorders: four new cases and literature review

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

The author of this thesis contributed for all the laboratory work, genomic data acquisition, analysis and interpretation of results as well as for the manuscript preparation and discussion of all the cases (except the aCGH analysis of patients 1 and 4). Patients' clinical data collection: for all the cases this was performed by the referring physician (GS, CDVK, CM, DA, JS, LFB, HVE, ERS).

frontiers in Genetics

ORIGINAL RESEARCH
published: 22 February 2019
doi: 10.3389/fgene.2019.00058

The Role of *AKT3* Copy Number Changes in Brain Abnormalities and Neurodevelopmental Disorders: Four New Cases and Literature Review

*Fátima Lopes[1,2], Fátima Torres[3,4], Gabriela Soares[5], Clara D. van Karnebeek[6,7], Cecília Martins[8], Diana Antunes[9], João Silva[5], Lauren Muttucomaroe[10], Luís Filipe Botelho[11], Susana Sousa[1,2], Paula Rendeiro[3], Purificação Tavares[3], Hilde Van Esch[12], Evica Rajcan-Separovic[13] and Patrícia Maciel[1,2]**

[1] School of Medicine, Life and Health Sciences Research Institute (ICVS), University of Minho, Braga, Portugal, [2] ICVS/3B's - PT Government Associate Laboratory, Guimarães, Portugal, [3] CGC Genetics, Porto, Portugal, [4] Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, Porto, Portugal, [5] Center for Medical Genetics Dr. Jacinto Magalhães, National Health Institute Dr. Ricardo Jorge, Porto, Portugal, [6] Department of Pediatrics, Centre for Molecular Medicine, BC Children's Hospital, University of British Columbia, Vancouver, BC, Canada, [7] Academic Medical Centre, Department of Pediatrics and Clinical Genetics, Amsterdam, Netherlands, [8] Department of Pediatrics, Médio Ave Hospital Center, Vila Nova de Famalicão, Portugal, [9] Medical Genetics Department, Hospital D. Estefânia, Centro Hospitalar Lisboa Central, Lisbon, Portugal, [10] Department of Pediatrics, University of British Columbia, Vancouver, BC, Canada, [11] Department of Neuroradiology, Hospital de Santo António, Porto Hospital Center, Porto, Portugal, [12] Laboratories for Center for Human Genetics, University Hospitals Leuven, Leuven, Belgium, [13] Department of Pathology, Children's Hospital of British Columbia, Vancouver, BC, Canada

OPEN ACCESS

Edited by:
Erica E. Davis,
Duke University, United States

Reviewed by:
Richard S. Nowakowski,
Florida State University College of Medicine, United States
Gemma Louise Carvill,
Northwestern University, United States

***Correspondence:**
Patrícia Maciel
pmaciel@med.uminho.pt

Specialty section:
This article was submitted to Genetic Disorders, a section of the journal Frontiers in Genetics

Received: *10 April 2018*
Accepted: *24 January 2019*
Published: *22 February 2019*

Citation:
Lopes F, Torres F, Soares G, van Karnebeek CD, Martins C, Antunes D, Silva J, Muttucomaroe L, Botelho LF, Sousa S, Rendeiro P, Tavares P, Van Esch H, Rajcan-Separovic E and Maciel P (2019) The Role of AKT3 Copy Number Changes in Brain Abnormalities and Neurodevelopmental Disorders: Four New Cases and Literature Review. Front. Genet. 10:58. doi: 10.3389/fgene.2019.00058

Microdeletions at 1q43-q44 have been described as resulting in a clinically recognizable phenotype of intellectual disability (ID), facial dysmorphisms and microcephaly (MIC). In contrast, the reciprocal microduplications of 1q43-q44 region have been less frequently reported and patients showed a variable phenotype, including macrocephaly. Reports of a large number of patients with copy number variations involving this region highlighted the *AKT3* gene as a likely key player in head size anomalies. We report four novel patients with copy number variations in the 1q43-q44 region: one with a larger deletion (3.7 Mb), two with smaller deletions affecting *AKT3* and *SDCCAG8* genes (0.16 and 0.18 Mb) and one with a quadruplication (1 Mb) that affects the entire *AKT3* gene. All patients with deletions presented MIC without structural brain abnormalities, whereas the patient with quadruplication had macrocephaly, but his carrier father had normal head circumference. Our report also includes a comparison of phenotypes in cases with 1q43-q44 duplications to assist future genotype-phenotype correlations. Our observations implicate *AKT3* as a contributor to ID/development delay (DD) and head size but raise doubts about its straightforward impact on the latter aspect of the phenotype in patients with 1q43-q44 deletion/duplication syndrome.

Keywords: 1q43-q44 CNVs, *AKT3*, microcephaly, macrocephaly, *ZBTB18*, *SDCCAG8*, phenotypic expressivity

INTRODUCTION

The 1q43-q44 microdeletion syndrome is characterized by variable degrees of intellectual disability (ID), growth retardation, microcephaly (MIC), *corpus callosum* anomalies (CCAs), 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). With the development of microarray technology, many cases with submicroscopic deletions in this region were reported, with the consequent identification of the genes associated with the 1q43-q44 deletion syndrome (Boland et al., 2007; Hill et al., 2007; van Bon et al., 2008; Orellana et al., 2009; Caliebe et al., 2010; Lall et al., 2011; Nagamani et al., 2012; Wang et al., 2013). In Ballif et al. (2012) defined three potentially critical regions for MIC, CCAs, and SZR, proposing that MIC is associated with deletions of the *AKT3* (AKT serine/threonine kinase 3) gene (in 93% of the cases); CCAs with deletions affecting *ZNF238* (gene zinc finger and BTB domain containing 18, also called *ZBTB18*) (in 86% of the cases) and SZR with deletions affecting the *FAM36A* (also called *COX20*, cytochrome c oxidase assembly factor) and *HNRNPU* (heterogeneous nuclear ribonucleoprotein U) genes (in 87% of the patients) (Ballif et al., 2012). In the same year, Nagamani et al ruled out the implication of *AKT3* gene in CCAs, because patients 5 and 6 of their series, which have, respectively, an intragenic deletion of *AKT3* and a small deletion affecting *AKT3* and *SDCCAG8* (serologically defined colon cancer antigen 8) genes, did not have CCAs (Nagamani et al., 2012). The *HNRNPU* (heterogeneous nuclear ribonucleoprotein U) and the *FAM36A* (family with sequence similarity 36, also known as *COX20*—cytochrome c oxidase assembly factor) genes, 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; Leduc et al., 2017). Even though the vast majority of the 1q43-q44 deletion cases described so far with MIC do carry genomic rearrangements that disrupt the *AKT3* gene, there are patients described in the literature with *AKT3* disruption that do not display MIC (Ballif et al., 2012). Conversely, there are also patients with 1q43-q44 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; Raun et al., 2017). In this perspective, the description of more patients with 1q43-q44 copy number variants (CNVs) may help to define more precise phenotype-genotype correlations. 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). In contrast with the deletions, there are only a few cases with pure gains in the region and detailed phenotypes. Copy number gains were described in patients with macrocephaly together with development delay, and also paired with speech and motor delay, hypotonia, and mild facial dysmorphisms (Wang et al., 2013; Chung et al., 2014).

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 of the findings in humans (Easton et al., 2005). Akt3-null and heterozygous mice also have an impairment in angiogenesis, showing a dose-dependent reduction in vessel number (5-fold decrease in homozygous and 2.5-fold decrease in heterozygous), an aspect of the phenotype that was never evaluated in any of the reported patients (Corum et al., 2014).

We describe four patients with 1q43-q44 CNVs, detected by array comparative genomic hybridization (aCGH), and attempt to establish genotype-phenotype correlations, aiming to bring further insight into the role of *AKT3* in brain abnormalities and ID.

METHODS

Ethical Procedures

Patients 1–3 were ascertained in the context of a larger study of neurodevelopmental disorders in Portugal, by the referring doctor. Clinical information was gathered in an anonymous database authorized by the Portuguese Data Protection Authority (CNPD). The study was approved by the ethics committee of Center for Medical Genetics Dr. Jacinto Magalhães, National Health Institute Dr. Ricardo Jorge. Written informed consent for sample collection, genetic studies and publication was obtained for all participants (signed by their parents). Informed consent for publication of photos was obtained from the parents for patients 1, 2, and 3 only.

Molecular Karyotyping

Genomic DNA was extracted from peripheral blood using the Citogene® DNA isolation kit (Citomed, Portugal) for patient 1, QIAsymphony SP (QIAGEN GmbH, Germany) for patients 2 and 3 and DNeasy (QIAGEN GmbH, Germany) for patient 4. The aCGH hybridization and analysis was performed using: Patient 1—aCGH Agilent 180 K 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 750 K Platform (750.000 markers distributed throughout the genome, with a medium resolution of 8–20 Kb), Chromosome Analysis Suite (ChAS 3.0) software (Affymetrix); Patient 4—Affymetrix Cytoscan HD array.

Quantitative PCR Confirmations

Primers for quantitative PCR (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).

For dosage quantification of *AKT3*, primers were designed for exons 8, 9, and 10 of the *AKT3* gene (ENSG00000117020). The reference genes used were *SDC4* (ENSG00000124145) and *ZNF80* (ENSG00000174255) localized in the 20q12-q13 and 3p12 regions, respectively (primers designed for all regions are listed **Supplementary Table 1**). All qPCR reactions were carried out in a 7,500-FAST Real Time PCR machine (Thermo Fisher Scientific, Waltham, MA, USA) using Power SYBR Green® (Thermo Fisher Scientific, Waltham, MA, USA). 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 (Thermo Fisher Scientific, Waltham, MA, USA).

RESULTS

Clinical Description and Molecular Findings

Patient 1 (DECIPHER #272238)

This girl was evaluated at 9 years of age for learning disabilities and MIC [occipitofrontal circumference (OFC) −2.5 SD]. Height and weight were at the 25th centile. She had a mildly sloping forehead and large upper central incisors (**Figure 1A**). Evaluation with the Wechsler Intelligence Scale for Children (third edition) (Wechsler, 1991) showed a full scale IQ of 63. Brain magnetic resonance imaging (MRI) was normal except for a discrete global atrophy. Pregnancy and delivery were uncomplicated at 35 weeks. OFC at birth was at the 3rd centile. Family history is unremarkable. Parents are young, healthy and non-consanguineous. The patient has a healthy younger brother. Congenital cytomegalovirus infection was excluded using PCR on the DNA obtained from the newborn metabolic disease screening Guthrie card. Informed consent was obtained from the child's parents for blood sampling and genetic analyses. Peripheral blood chromosome analysis demonstrated a normal 46,XX karyotype.

Subsequent array CGH revealed a 0.18 Mb *de novo* deletion at chromosome region 1q43-q44 (chr1:243,552,007–243,738,675) containing the *AKT3* and *SDCCAG8* genes (**Figure 2**). 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*. Sanger sequencing of the *AKT3* coding region revealed no variants.

Patient 2 (DECIPHER # 367117)

This boy was referred for evaluation of MIC and learning difficulties, associated with global DD in the past. Development evaluation (Griffiths) performed in 2011 reported all developmental areas within the low inferior range. He has an OFC of 48 cm ($P < 1$, −3.9 SD) and presents slightly dysmorphic features, including long philtrum and thin upper lip with cupid's bow (**Figure 1B**). Brain MRI performed at age 4 years was normal. Presently, 12 years old, he attends a school with an adapted curriculum, maintaining some learning difficulties. He was treated with risperidone for aggressive behavior in the past, but treatment has been discontinued. No seizures or other behavior anomalies were reported. The patient has a healthy older sister who was not genetically tested. Learning difficulties were also reported in the paternal side of the family: the father has MIC (OFC of 52.34 cm, P3), presents a rather long face, and has mild learning difficulties; one paternal uncle can't read or write and has a son with learning difficulties; a sister of the great-great-grandmother was always assistance-dependent due to a supposed ID and MIC. Apart from the father, none of the affected family members was genetically tested.

Array CGH revealed a 0.16 Mb deletion (chr1:243,592,147–243,749,968) affecting both the *AKT3* and *SDCCAG8* genes (**Figure 2**). The presence of the deletion was confirmed by qPCR for the *AKT3* gene (exon 10), which also revealed paternal inheritance.

Patient 3 (Decipher # 367116)

This boy was referred for evaluation around 3–4 months of age for evident MIC. At 3 years of age he presented weight and height within the normal range, but his OFC has been in P1 (−2.6 SD) since he was 8 months old. The patient had a hyperkinetic behavior and global DD, the language delay being the most striking. At the age of 5 years he was undertaking speech therapy. He didn't know colors or numbers, was described by parents as "clumsy" and by the teacher as aggressive. The patient has type B bilateral tympanogram. MRI evaluation of the brain at 3 years and 8 months old showed a cerebral volume in accordance with a decreased cephalic perimeter, without enlarged cerebrospinal fluid spaces [**Figure 1C**, (1–3)]. The cerebral hemispheres appeared otherwise unremarkable without noticeable malformations of cortical development, no signs of hypoxic-ischemic or infectious lesions. The *corpus callosum* was completely formed and displayed a normal thickness. No other abnormalities were seen. There was no evidence of significant skull abnormalities, other than the identified smaller dimensions and a slight left positional plagiocephaly; the electroencephalogram (EEG) was also normal. Concerning family history, he has a maternal uncle with epilepsy.

Array CGH revealed a 3.7 Mb deletion (chr1:240,366,425–244,111,022) (**Figure 2**), that proved to be *de novo* after qPCR confirmation using primers for the *AKT3* gene (exon 10).

Patient 4 (DECIPHER # 367118)

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 at a later age (all symptoms appeared before 5 years of age). He had a suspected IQ below 70 (although no formal evaluation was performed). He also presents macrocephaly (with OFC of 55.5 cm at 7 years old, corresponding to the 99th centile, +2.53 SD) and white matter lesions of brain including thalamic lesions. aCGH revealed a 1 Mb paternal quadruplication (chr1:243,415,063–244,478,355) (**Figure 2**) affecting the *CEP170*, *AKT3*, *SDCCAG8,* and *ZBTB18*

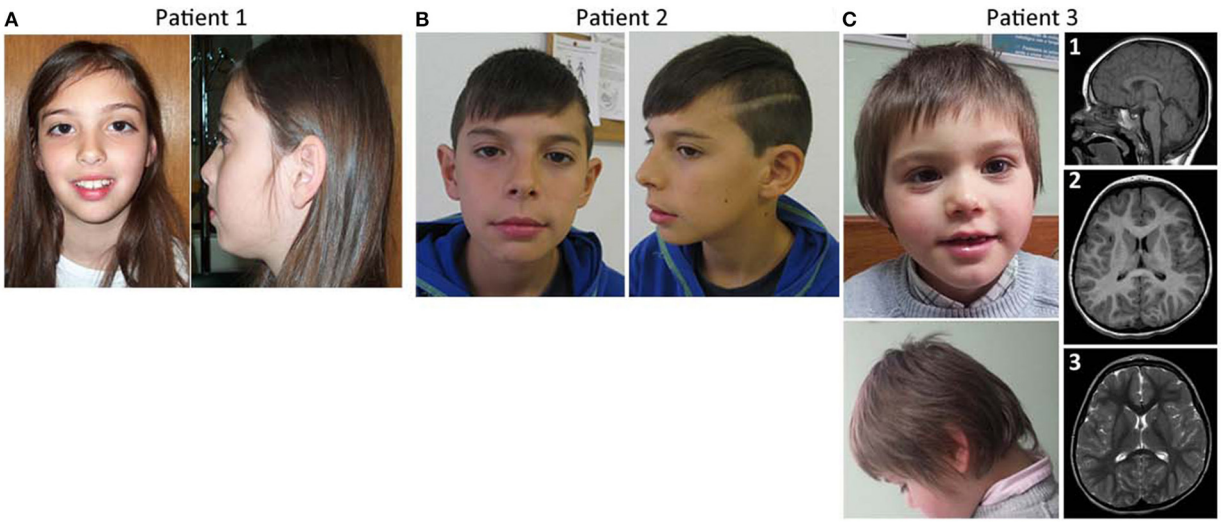


FIGURE 1 | Facial features and brain imaging of the patients. **(A)** Patient 1 facial features, **(B)** patient 2 facial features, **(C)** Patient 3 facial features, and MRI brain imaging.

FIGURE 2 | Schematic representations of the CNVs found in the patients and overlap with the critical regions proposed by Ballif et al. (2012). A 2.2 Mb genomic portion encompassing cytobands 1q43-q44 is shown. RefSeq genes present within the genomic region are shown in pink and the transcriptional direction is shown by the arrows. Shaded in gray is the proposed critical region for microcephaly (MIC) (affecting the *AKT3* gene), in blue the critical region for *corpus callosum* anomalies (CCAs) (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.

genes; the father is phenotypically normal, even though he is a carrier of this genetic variant. Meanwhile, patient 4 also underwent trio whole exome sequencing analysis, which retrieved no diagnosis.

DISCUSSION

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. Comparing patient 6 described by Nagamani et al. (2012) with patients 1 and 2 of this report, MIC in our patients 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 in the patient presented by Nagamani the deletion only affects the last 2 exons of the gene. Patient 3 in the current study fits quite well in the established 1q43-q44 microdeletion syndrome regarding the phenotype and size of the deletion. This patient, despite having a larger deletion that includes not only the *AKT3* gene but also the *CEP170* gene, has evident MIC but an apparently normal *corpus callosum* (**Figure 1C**).

The association of *AKT3* copy number gains with the mirror phenotype (macrocephaly) has also been reported in literature (Wang et al., 2013; Chung et al., 2014; Conti et al., 2015; Hemming et al., 2016). *AKT3* is one of the genes in which this type of mutations was found in patients with severe overgrowth syndromes (Tatton-Brown and Weksberg, 2013). Patient 4 in our series does present macrocephaly, supporting this model. However, his father becomes the first reported case of a quadruplication affecting *AKT3* in an asymptomatic individual.

In 2008, van Bon et al. described a pair of sisters with ID and MIC who inherited an *AKT3*-affecting deletion from their healthy mother (patient 11 and 12 from their series; the mother had normal IQ and normal OFC). In Ballif et al. (2012) also described a patient series in which three patients had inherited 1q43-q44 deletions (patient 10 which has MIC, 21 which has no information regarding OFC and 22 which doesn't have MIC). However, only patient 10 carried a deletion that affected *AKT3* and it was maternally inherited. Reports of cases carrying 1q43-q44 CNVs with different clinical outcomes can also be found in the DECIPHER database (Firth et al., 2009). There are three DECIPHER cases with *AKT3*-affecting deletions inherited from parents (#317423, #252432, and #277172, the two latter ones being reportedly healthy progenitors). However, it is important to highlight the difficulties in interpreting this data given the often incomplete and/or imperfect annotation of DECIPHER entrances, particularly regarding clinical description and follow up studies.

A possible contributing factor for the incomplete penetrance and clinical variability associated with *AKT3* genetic variants may be somatic mosaicism. This could lead to the heterogeneous distribution of AKT3 and/or modifying gene variants at the somatic level. In fact, somatic mutations activating the mTOR pathway have been shown to cause a continuum of cortical dysplasias; ultra-deep sequencing on DNA extracted from surgically resected brain, blood, and/or buccal samples from patients with several (mostly focal or asymmetric) cortical malformations led to the identification of somatic activating mutations in several mTOR genes, including *AKT3* (Lee et al., 2012; Poduri et al., 2012; Alcantara et al., 2017; D'Gama et al., 2017). Moreover, D'Gama et al. have proposed a "two-hit" model in a patient with both germline and somatic *TSC2* gene mutations (D'Gama et al., 2017). The presence of "second-hit" mutations, undetectable by targeted sequencing or present at a level below the detection limit of the techniques commonly used, could therefore contribute to the variability of the phenotypes presented. This may be paricularly relevant in a situation of increased gene dosage, which *per se* could have a similar but more subtle impact than *AKT3* somatic gain of function mutations.

A comparison between the "core" 1q43-q44 deletions and duplications phenotype and the four reported cases is made in **Table 1**. Of notice, the number of cases reported with duplications affecting the *AKT3* is quite reduced, making it difficult to establish a core symptomatology.

The lack of objective phenotypic measures among 1q43-q44 CNVs reports is an important limitation, however, to the adequate establishment of genotype-phenotype correlations. This was recently exposed by Raun et al who, by using more rigorous measures of head size deviation, showed that *AKT3* deletion is associated with more severe forms of MIC, while deletions in 1q43-q44 not affecting *AKT3* resulted in less severe MIC (probably because, as suspected before, *AKT3* is unlikely to be the only gene modulating head size at the 1q43-q44 region) (Raun et al., 2017). However, this might not be the case for all the cases since our patients don't seem to follow this pattern. Patient 3 has a MIC with −2.6 SD below the mean even though he has a deletion altering several gene besides *AKT3*, whereas patient 1, who only has the *AKT3* and *SDCCAG8* genes deleted, has a very similar MIC (−2.5 SD) to that of patient 3. Patient 2 is the case with the more severe MIC, with a SD of −3.9 even though he is the one with the smallest deletion of our series. *AKT3* partial or pure deletions may thus be subject to incomplete penetrance and/or differential expressivity driven by different genetic and epigenetic backgrounds of the individuals (the resulting phenotype not necessarily related to the deletion size). Although never reported in the 1q43-q44 region nor affecting *AKT3* specifically (the *AKT3* gene is not listed in the Geneimprint database on April 2018) (Geneimprint : Genes), imprinting alterations constitute another mechanism of differential growth (dy)sregulation which could be of relevance in these patients (reviewed in Choufani et al., 2013; Geneimprint, 2018).

Recently, single nucleotide variations in *ZBTB18* were identified in two patients, one with developmental and speech delay, MIC and dysmorphic features and the other with severe ID, breathing disturbances and MIC without structural anomalies (de Munnik et al., 2014; Lopes et al., 2016). This shows that *ZBTB18* mutation is sufficient to cause MIC, contradicting the exclusive contribution of *AKT3*. Although, this gene is not directly involved in the three patients with deletions here described, we cannot exclude that the deletions occurring in patient 1, 2, and 3 don't have an effect in *ZBTB18* expression in the nervous system.

The minimal overlapping region of all the patients with 1q43-q44 CNVs described in the current study encompasses only one additional gene to *AKT3*: the *serological defined colon cancer antigen 8* (*SDCCAG8*). 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 those patients include ID, a feature that is present in our patient, the alterations causing disease in those cases are associated with a recessive model of inheritance (Otto et al., 2010).

TABLE 1 | Comparison of the clinical features of the patients in the current series with patients with *AKT3* deletions described in the literature.

Clinical feature		Core phenotype	Patient 1	Patient 2	Patient 3	Core phenotype*	Patient 4
CNV		CN Loss	Deletion	Deletion	Deletion	CN Gain (pure)	Quadruplication
Clinical Overview	Gender	NR	♀	♂	♂	NR	♂
	Consanguinity	NR	No	No	No	NR	No
	Birth	NR	35 w (uneventful)	41 w	40 w	NR	To term
	Measurements at birth (heigh/weigth/OFC)	NR	NA	49cm (P7)/3240g(P10)/ 33cm(P2)	47cm(P1)/2710g(P2)/ 32.5cm(P1)	NR	NA
	Age at observation	NR	9y	12y	3y (Evident MIC)	NR	Before 5y
	ID	Moderate to severe	Mild (IQ = 63)	Mild (IQ NA)	Mild (IQ = 61)	Moderate	Mild
	Weigth (centile) at observation	NR	P25	$P < 5$	Within normal range for age¥	NR	P50
	Heigth (centile) at observation	Short stature	P25	P3	Within normal range for age¥	NR	P20
Cranio-facial abnormalities	Head/ OFC (centile)	MIC	MIC ($P < 3$); −2,5SD	MIC ($P < 5$)	MIC ($P < 3$)	MAC	MAC
	Structural brain abnormalities	Corpus callosum abnormalities (CCA) (agenesis/hypogenesis)	No	No	No	No	No
	Facial dysmorphisms	Round face, prominent forehead, flat nasal bridge, hypertelorism, epicanthal folds, malformed and low-set ears, hand and foot anomalies,	Mildly sploping forehead; large upper incisors	No major dysmorphisms	Upward palpebral fissures; retrognatia; poor hearing (conduction); dental caries	Prominent forehead, hypertelorism, wide nasal bridge, horizontal palpebral fissures, low set protruding ears.	NA
Others	Behavior	No alterations	No alterations	Aggressivity	Aggressivity; Auto-mutilation hyperactivity	No alterations	No alterations
	Seizures	Yes	No	No	No	No	No
	Hypotonia	Yes	No	No	No	Yes	No
	Genitalia	No alteration pattern	Normal	NA	Normal	No alteration pattern	Normal
	Heart	Minor heart conditions (usually resolve naturally)	NA	NA	Never evaluated	No	No
Molecular findings	Inheritance	–	*de novo*	Paternal	*de novo*	–	Paternal
	Confirmation	–	qPCR	qPCR	qPCR	–	NP
	Start (hg19)	–	243,552,007	243,592,147	240,366,425	–	243,415,063
	End (hg19)	–	243,738,675	243,749,968	244,111,022	-	244,478,355
	Size (Mb)	–	0.18	0.16	3.7	–	1
	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*

**Core phenotype of patients with 1q43–q44 duplications should be interpreted with care as there are very few cases reported. ¥ Specific measure not available; CNV, copy number variants; NA, not available; OFC, occipitofrontal circumference; y, years; m, months; MIC, microcephaly; ID, intellectual disability; DD, developmental delay; NR, not representative; qPCR, quantitative PCR; ND, not determined; NP, not performed.*

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). Additionally, *SDCCAG8* has been described to play a role in neuronal polarization and migration in the developing mouse cortex (Insolera et al., 2014), which would be consistent with the described genetic effects in humans. However, and given the presence of *SDCCAG8* deletions and quadruplication in heterozygosity in our patients, it most likely is not contributing to the MIC phenotype.

In summary, we describe four patients with 1q43-q44 CNVs, three of which with outcomes that are quite consistent with those of the "core" 1q43-q44 deletions affecting the *AKT3* gene, whereas the last (patient 4), in combination with other previously reported cases, highlights the not so straightforward and isolated implication of *AKT3* CNVs in human OFC determination. 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 the consequent of different genetic or epigenetic backgrounds of the individuals. This variability has important implications in the clinical practice in the context of the genetic counseling. The implication of *AKT3* in head size appears to be clear for the vast majority of the cases, even though not absolute. For this reason, the reporting of more patients with 1q43-q44 CNVs, their clinical and genetic features and their variable phenotypic expressivity is important and should be continued.

AUTHOR CONTRIBUTIONS

FL, FT, SS, PM, and PR performed the molecular studies and analyzed the molecular data. GS, CvK, CM, DA, JS, LM, and LB collected and analyzed clinical data. FL, FT, ER-S, HV, and PM drafted the paper. PT, ER-S, and PM obtained funding for this study. The study was performed under the direction of PM. All authors have agreed with and approved the final version.

FUNDING

This work has been funded by FEDER funds, through the Competitiveness Factors Operational Programme (COMPETE), and by National funds, through the Foundation for Science and Technology (FCT), under the scope of the projects: PIC/IC/83026/2007, PIC/IC/83013/2007 and POCI-01-0145-FEDER-007038. The work has also been funded by the project NORTE-01-0145-FEDER-000013, supported by the Northern Portugal Regional Operational Programme (NORTE 2020), under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (FEDER). FL was supported by Foundation for Science and Technology (FCT) through the fellowship SFRH/BD/90167/2012.

ACKNOWLEDGMENTS

We would like to thank all the patients and their families for participation in the genetic studies and for allowing this publication. To thank Marta Daniela Costa for the help with experiments. We would also like to acknowledge the DECIPHER Consortium, Database of Genomic Variants and OMIM since this study makes use of data made available by these platforms.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.00058/full#supplementary-material

REFERENCES

Alcantara, D., Timms, A. E., Gripp, K., Baker, L., Park, K., Collins, S., et al. (2017). Mutations of AKT3 are associated with a wide spectrum of developmental disorders including extreme megalencephaly. *Brain J. Neurol.* 140, 2610–2622. doi: 10.1093/brain/awx203

Ballif, B. C., Rosenfeld, J. A., Traylor, R., Theisen, A., Bader, P. I., Ladda, R. L., et al. (2012). 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.* 131, 145–156. doi: 10.1007/s00439-011-1073-y

Boland, E., Clayton-Smith, J., Woo, V. G., McKee, S., Manson, F. D., Medne, L., et al. (2007). 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.* 81, 292–303. doi: 10.1086/519999

Buffart, T. E., Israeli, D., Tijssen, M., Vosse, S. J., Mrsić, A., Meijer, G. A., et al. (2008). Across array comparative genomic hybridization: a strategy to reduce reference channel hybridizations. *Genes. Chromosomes Cancer* 47, 994–1004. doi: 10.1002/gcc.20605

Caliebe, A., Kroes, H. Y., van der Smagt, J. J., Martin-Subero, J. I., Tönnies, H., van't Slot, R., et al. (2010). 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.* 53, 179–185. doi: 10.1016/j.ejmg.2010.04.001.

Choufani, S., Shuman, C., and Weksberg, R. (2013). Molecular Findings in Beckwith–Wiedemann Syndrome. *Am. J. Med. Genet. C Semin. Med. Genet.* 163, 131–140. doi: 10.1002/ajmg.c.31363

Chung, B. K., Eydoux, P., An Karnebeek, C. D., and Gibson, W. T. (2014). Duplication of AKT3 is associated with macrocephaly and speech delay. *Am. J. Med. Genet. A.* 164, 1868–1869. doi: 10.1002/ajmg.a.36521

Conti, V., Pantaleo, M., Barba, C., Baroni, G., Mei, D., Buccoliero, A. M., et al. (2015). Focal dysplasia of the cerebral cortex and infantile spasms associated with somatic 1q21.1-q44 duplication including the AKT3 gene. *Clin. Genet.* 88, 241–247. doi: 10.1111/cge.12476

Corum, D. G., Tsichlis, P. N., and Muise-Helmericks, R. C. (2014). 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.* 28, 395–407. doi: 10.1096/fj.13-235382

de Munnik, S. A., García-Miñaúr, S., Hoischen, A., van Bon, B. W., Boycott, K. M., Schoots, J., et al. (2014). A de novo non-sense mutation in ZBTB18 in a patient with features of the 1q43q44 microdeletion syndrome. *Eur. J. Hum. Genet.* 22, 844–846. doi: 10.1038/ejhg.2013.249

De Vries, B. B., Knight, S. J., Homfray, T., Smithson, S. F., Flint, J., and Winter, R. M. (2001). Submicroscopic subtelomeric 1qter deletions: a recognisable phenotype? *J. Med. Genet.* 38, 175–178. doi: 10.1136/jmg.38.3.175

D'Gama, A. M., Woodworth, M. B., Hossain, A. A., Bizzotto, S., Hatem, N. E., LaCoursiere, C. M., et al. (2017). Somatic mutations activating the mTOR pathway in dorsal telencephalic progenitors cause a continuum of cortical dysplasias. *Cell Rep.* 21, 3754–3766. doi: 10.1016/j.celrep.2017.11.106

Easton, R. M., Cho, H., Roovers, K., Shineman, D. W., Mizrahi, M., Forman, M. S., et al. (2005). Role for Akt3/protein kinase Bgamma in attainment of normal brain size. *Mol. Cell. Biol.* 25, 1869–1878. doi: 10.1128/MCB.25.5.1869-1878.2005

Firth, H. V., Richards, S. M., Bevan, A. P., Clayton, S., Corpas, M., Rajan, D., et al. (2009). DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am. J. Hum. Genet.* 84, 524–533. doi: 10.1016/j.ajhg.2009.03.010

Gai, D., Haan, E., Scholar, M., Nicholl, J., and Yu, S. (2015). Phenotypes of AKT3 deletion: a case report and literature review. *Am. J. Med. Genet. A.* 167A, 174–179. doi: 10.1002/ajmg.a.36710

Geneimprint (2018). *Genes.* Available online at: http://www.geneimprint.com/site/genes-by-species (Accessed January 8, 2018).

Hamshere, M. L., Walters, J. T., Smith, R., Richards, A. L., Green, E., Grozeva, D., et al. (2013). Genome-wide significant associations in schizophrenia to ITIH3/4, CACNA1C and SDCCAG8, and extensive replication of associations reported by the Schizophrenia PGC. *Mol. Psychiatry* 18, 708–712. doi: 10.1038/mp.2012.67

Hemming, I. A., Forrest, A. R., Shipman, P., Woodward, K. J., Walsh, P., Ravine, D. G., et al. (2016). Reinforcing the association between distal 1q CNVs and structural brain disorder: a case of a complex 1q43-q44 CNV and a review of the literature. *Am. J. Med. Genet. Part B Neuropsychiatr.* 171B, 458–467. doi: 10.1002/ajmg.b.32427

Hill, A. D., Chang, B. S., Hill, R. S., Garraway, L. A., Bodell, A., Sellers, W. R., et al. (2007). A 2-Mb critical region implicated in the microcephaly associated with terminal 1q deletion syndrome. *Am. J. Med. Genet. A.* 143A, 1692–1698. doi: 10.1002/ajmg.a.31776

Hoebeeck, J., van der Luijt, R., Poppe, B., De Smet, E., Yigit, N., Claes, K., et al. (2005). Rapid detection of VHL exon deletions using real-time quantitative PCR. *Lab. Invest. J. Tech. Methods Pathol.* 85, 24–33. doi: 10.1038/labinvest.3700209.

Insolera, R., Shao, W., Airik, R., Hildebrandt, F., and Shi, S. H. (2014). SDCCAG8 regulates pericentriolar material recruitment and neuronal migration in the developing cortex. *Neuron* 83, 805–822. doi: 10.1016/j.neuron.2014.06.029

Jovanovic, L., Delahunt, B., McIver, B., Eberhardt, N. L., and Grebe, S. K. G. (2003). Optimising restriction enzyme cleavage of DNA derived from archival histopathological samples: an improved HUMARA assay. *Pathology* 35, 70–74. doi: 10.1080/0031302021000062370

Kenedy, A. A., Cohen, K. J., Loveys, D. A., Kato, G. J., and Dang, C. V. (2003). Identification and characterization of the novel centrosome-associated protein CCCAP. *Gene* 303, 35–46. doi: 10.1016/S0378-1119(02)01141-1

Krijgsman, O., Israeli, D., van Essen, H. F., Eijk, P. P., Berens, M. L. M., Mellink, C. H. M., et al. (2013). Detection limits of DNA copy number alterations in heterogeneous cell populations. *Cell. Oncol. Dordr.* 36, 27–36. doi: 10.1007/s13402-012-0108-2

Lall, M., Thakur, S., Puri, R., Verma, I., Mukerji, M., and Jha, P. (2011). A 54 Mb 11qter duplication and 0.9 Mb 1q44 deletion in a child with laryngomalacia and agenesis of corpus callosum. *Mol. Cytogenet.* 4:19. doi: 10.1186/1755-8166-4-19

Leduc, M. S., Chao, H. T., Qu, C., Walkiewicz, M., Xiao, R., Magoulas, P., et al. (2017). Clinical and molecular characterization of de novo loss of function variants in HNRNPU. *Am. J. Med. Genet. A.* 173, 2680–2689. doi: 10.1002/ajmg.a.38388

Lee, J. H., Huynh, M., Silhavy, J. L., Kim, S., Dixon-Salazar, T., Heiberg, A., et al. (2012). *De novo* somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *Nat. Genet.* 44, 941–945. doi: 10.1038/ng.2329

Lopes, F., Barbosa, M., Ameur, A., Soares, G., de SÁ, J., Dias, A. I., et al. (2016). Identification of novel genetic causes of Rett syndrome-like phenotypes. *J. Med. Genet.* 53, 190–199. doi: 10.1136/jmedgenet-2015-103568

Mankinen, C. B., Sears, J. W., and Alvarez, V. R. (1976). Terminal (1)(q43) long-arm deletion of chromosome no. 1 in a three-year-old female. *Birth Defects Orig. Artic. Ser.* 12, 131–136.

Nagamani, S. C. S., Erez, A., Bay, C., Pettigrew, A., Lalani, S. R., Herman, K., et al. (2012). Delineation of a deletion region critical for corpus callosal abnormalities in chromosome 1q43-q44. *Eur. J. Hum. Genet.* 20, 176–179. doi: 10.1038/ejhg.2011.171

Nakatani, K., Sakaue, H., Thompson, D. A., Weigel, R. J., and Roth, R. A. (1999). Identification of a human Akt3 (protein kinase B gamma) which contains the regulatory serine phosphorylation site. *Biochem. Biophys. Res. Commun.* 257, 906–910. doi: 10.1006/bbrc.1999.0559

Orellana, C., Roselló, M., Monfort, S., Oltra, S., Quiroga, R., Ferrer, I., et al. (2009). Corpus callosum abnormalities and the controversy about the candidate genes located in 1q44. *Cytogenet. Genome Res.* 127, 5–8. doi: 10.1159/000279261

Otto, E. A., Hurd, T. W., Airik, R., Chaki, M., Zhou, W., Stoetzel, C., et al. (2010). Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy. *Nat. Genet.* 42, 840–850. doi: 10.1038/ng.662

Poduri, A., Evrony, G. D., Cai, X., Elhosary, P. C., Beroukhim, R., Lehtinen, M. K., et al. (2012). Somatic activation of AKT3 causes hemispheric developmental brain malformations. *Neuron* 74, 41–48. doi: 10.1016/j.neuron.2012.03.010

Poot, M., and Kas, M. J. (2013). Antisense may make sense of 1q44 deletions, seizures, and HNRNPU. *Am. J. Med. Genet. A.* 161A, 910–912. doi: 10.1002/ajmg.a.35770.

Poot, M., Kroes, H. Y., and Hochstenbach, R. (2008). AKT3 as a candidate gene for corpus callosum anomalies in patients with 1q44 deletions. *Eur. J. Med. Genet.* 51, 689–690. doi: 10.1016/j.ejmg.2008.07.001

Raun, N., Mailo, J., Spinelli, E., He, X., McAvena, S., Brand, L., et al. (2017). Quantitative phenotypic and network analysis of 1q44 microdeletion for microcephaly. *Am. J. Med. Genet. A.* 173, 972–977. doi: 10.1002/ajmg.a.38139

Schaefer, E., Zaloszyc, A., Lauer, J., Durand, M., Stutzmann, F., Perdomo-Trujillo, Y., et al. (2011). Mutations in SDCCAG8/NPHP10 cause bardet-biedl syndrome and are associated with penetrant renal disease and absent polydactyly. *Mol. Syndromol.* 1, 273–281. doi: 10.1159/000331268

Tatton-Brown, K., and Weksberg, R. (2013). Molecular mechanisms of childhood overgrowth. *Am. J. Med. Genet. C Semin. Med. Genet.* 163C, 71–75. doi: 10.1002/ajmg.c.31362

Thierry, G., Bénéteau, C., Pichon, O., Flori, E., Isidor, B., Popelard, F., et al. (2012). Molecular characterization of 1q44 microdeletion in 11 patients reveals three candidate genes for intellectual disability and seizures. *Am. J. Med. Genet. A.* 158A, 1633–1640. doi: 10.1002/ajmg.a.35423

van Bon, B. W., Koolen, D. A., Borgatti, R., Magee, A., Garcia-Minaur, S., Rooms, L., et al. (2008). Clinical and molecular characteristics of 1qter microdeletion syndrome: delineating a critical region for corpus callosum agenesis/hypogenesis. *J. Med. Genet.* 45, 346–354. doi: 10.1136/jmg.2007.055830

Wang, D., Zeesman, S., Tarnopolsky, M. A., and Nowaczyk, M. J. (2013). Duplication of AKT3 as a cause of macrocephaly in duplication 1q43q44. *Am. J. Med. Genet. A.* 161A, 2016–2019. doi: 10.1002/ajmg.a.35999

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children, 3rd Edn*. San Antonio, TX: The Psychological Corporation.

Conflict of Interest Statement: FT, PR, and PT were employed by the company CGC Genetics.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary material to the article:

The role of *AKT3* copy number changes in brain abnormalities and neurodevelopmental disorders: four new cases and literature review

Fátima Lopes[1,2], Fátima Torres[3,4], Gabriela Soares[5], Clara D. van Karnebeek[6,7], Cecília Martins[8], Diana Antunes[9], João Silva[5], Lauren Muttucomaroe[10], Luís Filipe Botelho[11], Susana Sousa[1,2], Paula Rendeiro[3], Purificação Tavares[3], Hilde Van Esch[12], Evica Rajcan-Separovic[13], Patrícia Maciel[1,2*]

[1]Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal
[2]ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal
[3]CGC Genetics, Porto, Portugal
[4]Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, Porto, Portugal;
[5]Center for Medical Genetics Dr. Jacinto Magalhães, National Health Institute Dr. Ricardo Jorge, Praça Pedro Nunes, Porto, Portugal
[6]Department of Pediatrics, BC Children's Hospital & Centre for Molecular Medicine, University of British Columbia, Vancouver, Canada
[7]Department of Pediatrics, Médio Ave Hospital Center, Vila Nova de Famalicão, Portugal
[8]Medical Genetics Department, Hospital D. Estefânia, Centro Hospitalar Lisboa Central, Lisboa, Portugal
[9]Department of Neuroradiology, Hospital de Santo António, Porto Hospital Center, Porto, Portugal
[10]Laboratories for Center for Human Genetics, University Hospitals Leuven, Leuven, Belgium.

Correspondence to:

Patricia Maciel, PhD

Life and Health Sciences Research Institute (ICVS) - School of Medicine, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
Phone: +351 253604824
Email: pmaciel@ecsaude.uminho.pt

Supplementary Table S1 – Primers used for quantitative PCR confirmation.

Chromosome	Gene	Reference sequence	Primer location	Primer Forward 5'→3'	Primer Reverse 5'→3'	Amplicon size (bp)
Chr 1	*AKT3*	ENSG00000117020	Exon7	TCTGGGCTTAACCTCTTCCA	TGTTAAAAAGGGATGTCTAGTGTTC	162bp
Chr 1	*AKT3*	ENSG00000117020	Exon8	CCTTGAAATATTCCTTCCAGACA	CCATGCAAATACTGGATTTACTTCT	101bp
Chr 1	*AKT3*	ENSG00000117020	Exon9	AGAGAGCGGGTGTTCTCTGA	CCTTGAGATCACGGTACACAA	106bp
Chr 1	*AKT3*	ENSG00000117020	Exon10	CAGTTGGAGAATCTAATGCTGGA	AATGGAACCGAAGCCTACCT	150bp
Chr 3	*ZNF80*	ENSG00000174255	Exon1	GCTACCGCCAGATTCACACT	AATCTTCATGTGCCGGGTTA	182bp
Chr 20	SDC4	ENSG00000124145	Exon4	ACCGAACCCAAGAAACTAGA	GTGCTGGACATTGACACCT	101bp

Sub-Chapter 2.3 – 2p15p16.1 microduplication: case report and review

Disclaimer:

The results presented in this sub-chapter refer to the collection of four patients from the same family, all with a 2p15 microduplication. These results will be published in an international peer-reviewed journal under the following title and authorship:

2p15p16.1 microduplication: case report and review

*Qiao Y, *Torres F, Bagheri H, Tena J, Santos-Pereira JM, Martell S, O'Driscoll M, Gomez-Skarmeta JL, Rocha F, Melo C, Lewis S, Maciel P and Rajcan-Separovic E

In preparation

*both authors contributed equally for this work

The author of this thesis contributed for genomic data acquisition, analysis and interpretation of results for the aCGH for the reported patients, as well as for the manuscript preparation and discussion of the cases.

2p15p16.1 microduplication: case report and review

*Qiao Y, *Torres F, Bagheri H, Tena J, Santos-Pereira JM, Martell S, O'Driscoll M, Gomez-Skarmeta JL, Rocha F, Melo C, Lewis S, Maciel P and Rajcan-Separovic E

In preparation

Abstract

The 2p15p16.1 microdeletion syndrome is characterized by intellectual disability (ID), microcephaly, speech delay and distinct dysmorphic features. Four genes have been proposed to play an important role in developmental anomalies (*BCL11A*, *XPO1*, *REL*, and *USP34*), three of which caused microcephaly and abnormal development in zebrafish knockdown (*BCL11A*, *XPO1*, and *REL*). Microduplications in the 2p15p16.1 region have been rarely reported so far. Here, we describe a family with three children affected with ID and congenital anomalies and their mother with learning difficulty who all carry a small 150Kb duplication in the proximal 2p15p16 microdeletion region (61376462-61527143bp, hg19) involving the *C2orf74* (disrupted), *AHSA2* and *USP34* (disrupted) genes. Our review of 15 other microduplication cases which had more detailed phenotype information (10 from publications and 5 from the Decipher database) and including at least one of the above 3 genes (*USP34*, *C2orf74* and *AHSA2*), shows that ID, speech delay, variable congenital anomalies and facial dysmorphisms exist commonly in cases with 2p15p16.1 microduplications, despite variable size/breakpoints of the duplications. Expression, and topography associated domains were not altered in patient lymphoblasts. We conclude that duplications in the 2p15p16.1 chromosomal region are associated with ID and variable abnormalities, distinct from the deletions, and that head size is not typically affected (2/18 cases). In the future, functional studies in patient induced pluripotent

stem cells would be helpful in determining their effect in a functionally more relevant biological system.

Running title: 2p15p16.1 microduplication

Key words: 2p15p16.1 microduplication; 2p15p16.1 microdeletion; DNA copy number variants (CNVs); intellectual disability (ID)

Introduction

The 2p15p16.1 microdeletion syndrome (OMIM 612513) is a rare genetic disorder caused by *de novo* deletions of variable size (0.1 to 9Mb) within the 2p15p16.1 region (from 55,500,000 to 65,300,000, hg19). Thus far, more than 36 subjects with this microdeletion have been published with common clinical features including intellectual disability (ID), microcephaly, developmental delay (DD), speech delay, hypotonia, distinctive facial dysmorphism and digital anomalies (Bagheri et al. 2016; Lévy et al. 2017). Four genes (*USP34*, *BCL11A*, *REL* and *XPO1*) were considered critical for developmental abnormalities due to their deletion in most cases and presence in the smallest deletions; however, only *BCL11A*, *REL* and *XPO1* showed an abnormal phenotype in zebrafish knock-down models (Bagheri et al. 2016). In contrast to the microdeletion syndrome, cases with 2p15p16.1 microduplications with phenotype descriptions are rarely described (Mimouni-Bloch et al. 2015; Borlot et al. 2017; Pavone et al. 2018; Lovrecic et al. 2018; Chen et al. 2018). Here we present a family with three children and their mother who carry a 150Kb microduplication in the 2p15p16 microdeletion region. We provide their detailed phenotypes and assess protein expression of the genes within the duplication (*USP34*, *C2orf74* and *AHSA2*) and the structure of the topologically associated domains using Hi-seq. We also review other cases with 2p15p16 microduplication reported from publications and Decipher database (https://decipher.sanger.ac.uk/syndrome/70#overview) (Firth et al. 2009).

Clinical description

The proband is a 14-year old boy who has older twin brothers, currently aged 18-years old. They all had normal prenatal development with no history of intrauterine growth restriction. They present the normal height for their age and are normocephalic. All of them have attention deficit hyperactivity disorder (ADHD) and are medicated with methylphenidate. They have a personal history of speech delay, although presently they do not show speech nor language delay. They have a cognitive deficit corresponding to mild ID (according to DSM-IV-TR), being slightly better in verbal tasks than in performance tasks. The learning difficulties are more evident in the twins, who attend a specialized school, while the proband attends the regular school, although with special support. Several consistent facial dysmorphisms were noted: bulbous nose; prominent nasal philtrum; dysplastic ears; big mouth; thick lips; eyebrows with an unusual shape and retrognatia. The twins have myopia and astigmatism. The proband has a peculiar hair implantation (almost a widow peak). They present a deviation of the hallux and camptodactyly of all 5 hand fingers. Their behaviour is considered normal with no autistic features. Neurological/ genitourinary/endocrinological/cardiac anomalies are also absent. There was no history of recurrent infections.

Their mother seems to have some cognitive deficit, but no ADHD, although there was no formal cognitive evaluation. She has been taking care of her children, so despite some limitations she apparently functions in daily life. She attended the regular school and completed grade 9 (secondary school); she failed once in grade 8 and described herself as having learning difficulties. Physically, all of them (the mother and the children) have similar facial and finger characteristics (Figure 1). The father's phenotype and family history are unknown.

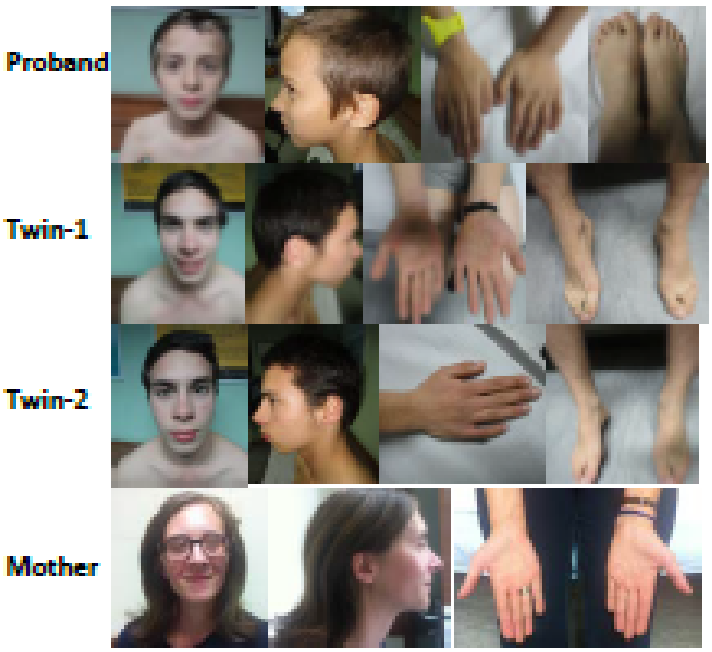


Figure 1: Photos of the face, hand and foot dysmorphisms of the patients.

Methods

Genome-wide chromosome microarray analysis was performed in the family using KaryoArray®v3.0 (Agilent 8x60k, Santa Clara, CA) platform. High resolution Affymetrix Cytoscan SNP microarray was performed in one of the twins (http://www.affymetrix.com) (Affymetrix Inc., Santa Clara, CA). Quantitative multiplex polymerase chain reaction (PCR) of short fluorescent fragments (QMPSF) was used for the validation of the duplicated genes (*C2orf74*, *AHSA2*, and *USP34*) (primers on supplementary table S1) (Qiao et al. 2013). The protein expression levels of the duplicated genes were detected by Western blot from the subject-derived Epstein-Barr Virus (EBV)-transformed lymphoblastoid cell lines (LCLs) (Wen et al. 2015), using Anti-AHSA2 (ab107074, Abcam) and Anti-C2orf74

(ab188356, Abcam), Anti-Actin (A2066, Sigma) and as secondary antibody, Goat Anti-Rabbit IgG H&L (HRP) (ab97051, Abcam). For USP34, urea-based whole cell extracts (WCEs) were prepared: 9M urea, 50mM Tris-HCL at pH7.5 and 10mM 2-mercaptoethanol with sonication (15sec at 30% amplitude using a micro-tip; Sigma-Aldrich). WCEs were blotted using an Anti-USP34 (A300-824, Bethyl). Protein concentration was determined using the Bradford Assay. Samples were then stored at -20°C or immediately boiled in 2X SDS-loading buffer (5% SDS, 10% glycerol, 10% 2-β-mercaptoethanol, 125mM Tris-HCl, pH 6.8 and 0.2% bromophenol blue), loaded onto SDS-PAGE gels and then semi-dry blotted onto polyvinyl difluoride (PVDF) membranes.

Hi-seq experiments were performed in one of the twins using the established method (Fernández-Miñán et al. 2016; Irastorza-Azcarate et al. 2018).

Results

A 145Kb (61,377,041-61,522,171) (hg19) duplication was identified in the two affected siblings and their mother. The presence of duplication (61376462-61527143bp) was confirmed in one of the twins using a high-resolution platform (Affymetrix Cytoscan SNP microarray). No other copy number changes were found in the family. This small duplication only involved 3 genes: *C2orf74* (disrupted), *AHSA2*, and *USP34* (disrupted) (Figure 2). The duplication of the three genes was validated in the family by QMPSF. However, the protein expression for the 3 genes from the subject-derived LCLs did not show statistically significant changes compared to the normal controls (Figure 3). The topologically associated domains were also comparable to controls (data not shown).

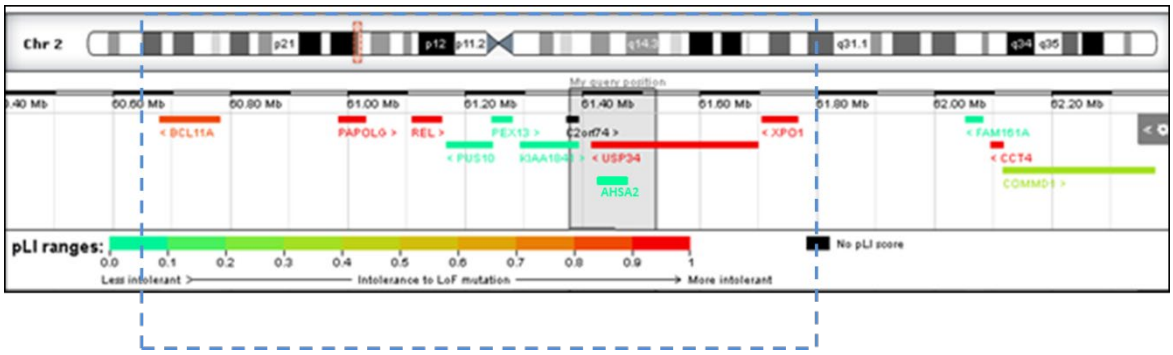


Figure 2: The genomic profile of the 2p15p16 microduplication region: the dotted blue line rectangle encompasses the potential critical region of the microduplication of 2p15-p16, as defined by Pavone, P., European Journal of Medical Genetics (2018), https://doi.org/10.1016/j.ejmg.2018.05.001. The grey rectangle defines the area duplicated in our patients, encompassing genes *C2orf74* (disrupted), *AHSA2* and *USP34* (disrupted).

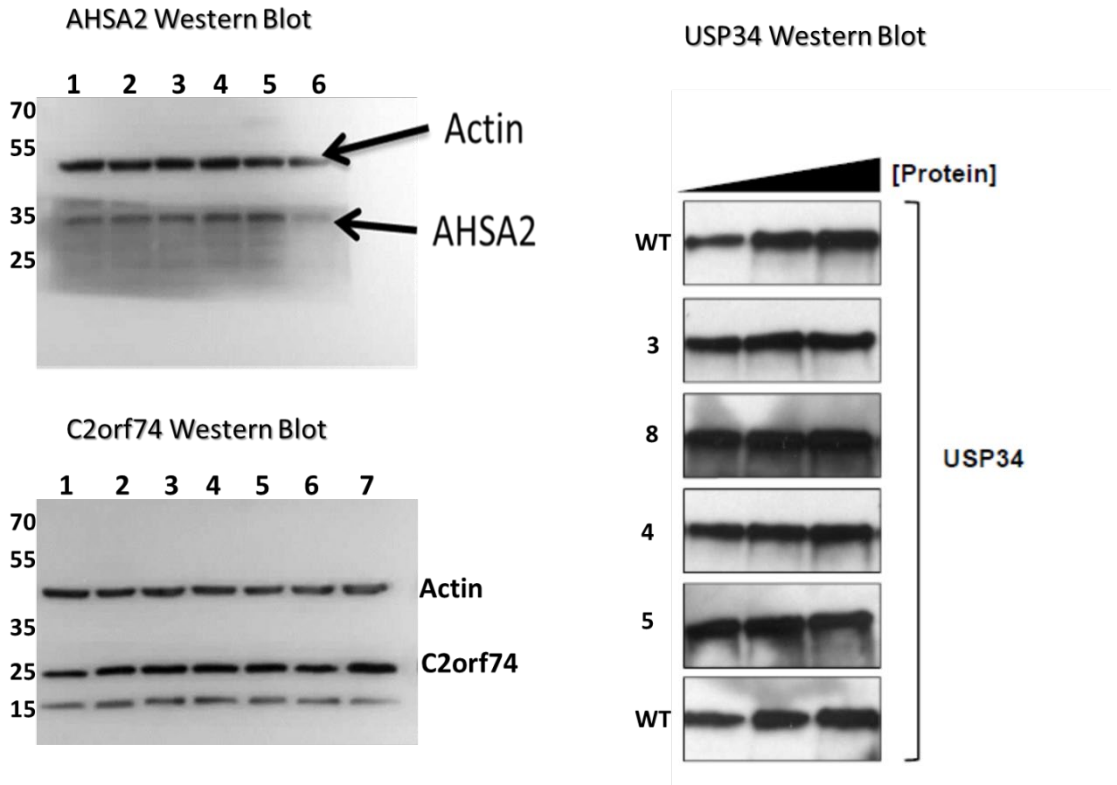


Figure 3: Protein expression for AHSA2, C2orf74 and USP34 from the subject-derived LCLs. 1 – male control; 2 – female control; 3 – proband; 4 – twin-2; 5 – mother; 6 – control 1 (2pdel); 7 – control 2 (2pdel); 8 – twin-1; WT – wild type.

Discussion

We report a family with 4 individuals with developmental delay/learning difficulties and dysmorphic features who carry the smallest duplication described so far in the 2p15p16.1 region. It includes 3 genes: *C2orf74*, *AHSA2* and *USP34*. *C2orf74* (chromosome 2 open reading frame 74) is a protein coding gene with undetermined function and no established link to nervous system; nevertheless, it is widely expressed in the brain [data retrieved from NCBI Gene, https://www.ncbi.nlm.nih.gov/gene/339804 and Bgee database (Bastian et al. 2008)]. It was recently described, together with the *USP34* gene, as a novel susceptibility gene for ankylosing spondylitis, in the Han Chinese population (Wang et al. 2017). *AHSA2*, presently designated by *AHSA2P* (Activator of HSP90 ATPase Homolog 2, Pseudogene) encodes a pseudogene also with undetermined function and/or link to nervous system. Their protein expression in patient lymphoblasts is unchanged. *USP34* regulates axin stability and Wnt/CTNNB signalling which play a very important role in ID and autism (Lui et al. 2011; Oliva, Vargas, and Inestrosa 2013; Kwan, Unda, and Singh 2016). In fact, Wnt signalling influences different aspects of neuronal circuit assembly through changes in gene expression and/or cytoskeletal modulation (Salinas and Zou 2008). Furthermore, many of the proteins in both signalling pathways localize to the synapse and play important functions in synaptic growth and maturation, reviewed by (Kwan, Unda, and Singh 2016). This gene was included in the smallest 2p15-p16 deletions (Bagheri et al. 2016; Mallett et al. 2018) and when deleted showed altered protein and RNA expression in patient lymphoblasts but no functional consequences in patient cells or abnormal phenotype in zebrafish (Bagheri et al. 2016).

The duplication in our cases involves exons 34-80 of this gene, overlapping with the critical region of the 2p15p16 Microdeletion region (Lévy et al. 2017). However, the protein expression of this gene in patients' LCLs is unchanged. It was previously reported that microduplications can alter the chromosome topology and result in abnormal expression of distant genes (Zlotorynski 2018), even without affecting the gene expression of genes within the duplication. However, we have not found evidence of altered chromosome topology using Hi-seq in one of the sibs. The possibility of the duplication having a functional effect in a cell type (e.g. in brain cells) and transcript expression specific manner remains open.

We reviewed the clinical and genomic information of fourteen more patients with duplications in this region (Table 1) which had phenotype description and a duplication <3.5Mb including at least one of the three genes from our patients' duplication (*AHSA2*, *Corf74* and *USP34*) and/or any of the three candidate genes for the microdeletion syndrome which cause microcephaly/developmental anomalies in zebrafish when knocked down (*BCL11A*, *REL* and *XPO1*). The duplication size in the 19 cases varied from 0.15Mb to 3.24 Mb and most of them included the key genes (18/19 cases with *USP34*, 13 cases with *BCL11A*, and 14 cases with *XPO1* and/or *REL*).

Table 1. Genomic and phenotype data comparison among our cases and reported ones with 2p15p16 microduplication (hg19)

Group	Subject	Genomic Position	Size	Inheritance	Genes*	Major Phenotypes
Our cases	Proband	2:61376462-61527143	150kb	Maternal	***C2orf74*, AHSA2P, USP34****	Mild ID, ADHD, speech delay, facial dysmorphisms (retrognatia; bulbous nose; dysplastic ears; thick lips; big mouth; prominent nasal filter; eyebrows with a different (weird) shape), fingers dysmorphism (a deviation of the halux and camptodactyly of all 5 hand fingers)
	Twins	2:61376462-61527143	150kb	Maternal	***C2orf74*, AHSA2P, USP34****	Mild ID, ADHD, speech delay, facial dysmorphisms (retrognatia; bulbous nose; dysplastic ears; thick lips; big mouth; prominent nasal filter; eyebrows with a different (weird) shape), fingers dysmorphism (a deviation of the halux and camptodactyly of all 5 hand fingers), myopia and astigmatism
	Mother	2:61376462-61527143	150kb	Unknown**	***C2orf74*, AHSA2P, USP34****	Cognitive deficit (not formally evaluated), learning difficulties, fingers dysmorphism, similar facial characteristics
Published Cases	(Pavone et al. 2018)	2:60294104–62030285	1.73Mb	*de novo*	*BCL11A, PAPOLG, REL, PUS10, PEX13, KIAA1841,* ***C2orf74, AHSA2, USP34,*** *XPO1*	ID, hypotonia, motor dyspraxia, epileptic seizures, behavioral anomalies; scoliosis, obesity, and minor facial dysmorphisms (dolicocephaly, a short receding forehead, puffy cheeks, protruding middle face, retrognathia, swollen hands with bilateral fifth fingers clinodactyly, and bilateral second-third toes syndactyly, and nail dysplasia).
	(Lovrecic et al. 2018)	2:60113626-62111114	2.0Mb	*de novo*	*BCL11A, PAPOLG, REL, PUS10, PEX13, KIAA1841,* ***C2orf74, AHSA2, USP34****, XPO1, SNORA70B, FAM161A, CCT4**	Mild ID, DD, mild dysmorphic features (receding forehead, broad and high nasal bridge, sparse eyebrows, epicanthal folds, straight eyelashes and pronounced philtrum; bilateral clinodactyly of the 5th finger and bilateral 2nd-3rd toes syndactyly).
		2:60308869-62368583	2.06Mb	*de novo*	*BCL11A, PAPOLG, REL, PUS10, PEX13, KIAA1841,* ***C2orf74, AHSA2, USP34****, XPO1, SNORA70B, FAM161A, CCT4, COMMD1**	Mild hypertonus and related mild motor delay; macrocephaly
		2:60236241-61848845	1.61Mb	Maternal	*BCL11A, PAPOLG, REL, PUS10, PEX13, KIAA1841,* ***C2orf74, AHSA2, USP34****, XPO1*	Learning difficulties and speech delay

		2:59938734-62025519	2.09Mb	*de novo*	*BCL11A, PAPOLG, REL, PUS10, PEX13, KIAA1841,* ***C2orf74, AHSA2, USP34****, XPO1*	Mild ID, speech delay, with small ears and small hands; recurrent infections with neutropenia and thrombocytopenia.
	(Chen et al. 2018) (fetus)	2:58288588-61532538	3.24Mb	Maternal	*VRK2, FANCL, MIR4432, BCL11A, PAPOLG, FLJ16341, REL, PUS10, PEX13, KIAA1841, LOC339803,* ***C2orf74, AHSA2, USP34****	Unknown; evaluated at 8 months and without evidence of psychomotor developmental abnormality.
	(Chen et al. 2018) (pregnant woman)	2:58288588-61532538	3.24Mb	Unknown***	*VRK2, FANCL, MIR4432, BCL11A, PAPOLG, FLJ16341, REL, PUS10, PEX13, KIAA1841, LOC339803,* ***C2orf74, AHSA2, USP34****	Moderate ID, without behavioural disorders.
	(Chen et al. 2018) (pregnant' s sister)	2:58288588-61532538	3.24Mb	Unknown***	*VRK2, FANCL, MIR4432, BCL11A, PAPOLG, FLJ16341, REL, PUS10, PEX13, KIAA1841, LOC339803,* ***C2orf74, AHSA2, USP34****	Moderate ID, without behavioural disorders.
	(Mimouni-Bloch et al. 2015)	2:60150427-61816209	1.66Mb	*de novo*	*MIR4432, BCL11A, PAPOLG, FLJ16341, REL, PUS10, PEX13, KIAA1841, LOC339803,* ***C2orf74, AHSA2, USP34****, SNORA70B, XPO1*	Mild ID, mild language delay and attention deficit behavior, mild dysmorphic features (puffy eyelids, broad philtrum and right earlobe sinus), motor dyspraxia and mild generalized hypotonia, sensory processing difficulties, lacrimal duct stenosis, moderate bilateral hearing loss, Iron deficiency anemia
	(Borlot et al. 2017)	2:60326674-62025420	1.7Mb	*de novo*	*MIR4432, BCL11A, PAPOLG, FLJ16341, REL, PUS10, PEX13, KIAA1841, LOC339803,* ***C2orf74, AHSA2, USP34****, SNORA70B, XPO1*	Mild ID, Retinitis pigmentosa; seizure
	258333	2:59792238-61879023	2.09Mb	*de novo* constitutive	*MIR4432, BCL11A, PAPOLG, FLJ16341, REL, PUS10, PEX13, KIAA1841, LOC339803,* ***C2orf74, AHSA2, USP34****, SNORA70B, XPO1*	ID, macrocephaly, abnormality of erythrocytes, recurrent infections

Decipher Cases (including at least one gene present in our patients)	265052	2:60541781-61952880	1.41Mb	*de novo*	*MIR4432*, BCL11A, PAPOLG, FLJ16341, REL, PUS10, PEX13, KIAA1841, LOC339803,* ***C2orf74, AHSA2, USP34****, SNORA70B, XPO1*	DSLD, obesity, arthritis, precocious puberty in females
	1570	2:60648296-61568645	920.35 kb	*de novo*	*BCL11A, PAPOLG, FLJ16341, REL, PUS10, PEX13, KIAA1841, LOC339803,* ***C2orf74, AHSA2, USP34****	ID, DSLD, blepharophimosis, delayed cranial suture closure, facial dysmorphysm (frontal bossing, low-set ears, micrognathia, thin lower lip vermilion, thin upper lip vermilion, upslanted palpebral fissure, wide nasal bridge, abnormality of the nasal alae), atria septal defect, hyperextensibility of the finger joints, proportionate short stature, short palm, short phalanx of finger, tapered finger, 2-3 toe syndactyly
	290412	2:61246086-61391564	145.48 kb	Unknown	*PEX13*, KIAA1841, LOC339803,* ***C2orf74****	ID, developmental regression, clinodactyly of the 5th toe, hirsutism, hypertrichosis
	256542	2:61297652-61785065	487.41 kb	*de novo*	*KIAA1841*, LOC339803,* ***C2orf74, AHSA2, USP34****, SNORA70B, XPO1*	ID, feeding difficulties in infancy, abnormality of the upper respiratory tract, facial dysmorphysm (almond-shaped palpebral fissure, epicanthus, hypertelorism, low-set ears), muscular hypotonia, patent ductus arteriosus, vesicoureteral reflux

Legend: ID – Intellectual Disability; ADHD – Attention Deficit Hyperactivity Disorder; DSLD – Delayed Speech and Language Development; GDD – Global Developmental Delay; * disrupted by the CNV breakpoint; ** the parents (maternal grandparents of the siblings) were not studied; *** probably inherited from the father.

Despite the variability in breakpoints and duplication sizes, all patients had ID and variable morphological anomalies. Speech delay, hypotonia, dysmorphic features, behaviour abnormalities, and digital anomalies were also noted and are also present in the cases with deletions. The head size was not reported to be abnormal in duplications except in one case with macrocephaly (DECIPHER #258333). Similarly, by comparing the head size in 10 cases with 2p15p16 microduplications and 26 with microdeletions, Loviglio and colleagues demonstrated normal head size in cases with duplications while significant microcephaly was noted in cases with microdeletions (Loviglio et al. 2017).

Why duplications of variable sizes affecting different genes in the 2p15p16.1 region cause an abnormal phenotype including ID remains puzzling. The possibility that genes involved in the duplications participate in similar or same pathways, which can be disturbed by imbalance of any of the components, was proposed. *USP34* has a known function in NF-kB signalling and enrichment of genes from 2p15p16.1 deletion in the NF-kB pathway was noted (Bagheri et al. 2016), although the pathway also did not seem to be altered in EBV transformed lymphoblasts. However, studying pathways in EBV transformed lymphoblasts may not reflect the real biological effect of the gene copy number change and future studies of the challenging and genomically very unstable 2p15p16.1 region should also be performed in neurons derived from patients' pluripotent stem cells. In addition, a much larger number of cases with detailed clinical description will be required to improve our understanding of the role of duplications in the 2p15p16.1 region.

Acknowledgements

This work has been funded by FEDER funds, through the Competitiveness Factors Operational Programme (COMPETE), and by National funds, through the Foundation for Science and Technology (FCT), under the scope of the projects: PIC/IC/83026/2007, PIC/IC/83013/2007 and POCI-01-0145-FEDER-007038. Additionally, this work was funded by the project NORTE-01-0145-FEDER-000013, supported by the Northern Portugal Regional Operational Programme (NORTE 2020), under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (FEDER). This work was supported by funding from the Canadian Institutes for Health Research (CIHR) (MOP 74502; PI: ERS). MESL and ERS are Career Scholars supported by the Michael Smith Foundation for Health Research. We also appreciate the support of the family members involved in this study.

There is no conflict of interest of our work.

Bibliography

Bagheri, Hani, Chansonette Badduke, Ying Qiao, Rita Colnaghi, Iga Abramowicz, Diana Alcantara, Christopher Dunham, et al. 2016. “Identifying Candidate Genes for 2p15p16.1 Microdeletion Syndrome Using Clinical, Genomic, and Functional Analysis.” *JCI Insight* 1 (3). https://doi.org/10.1172/jci.insight.85461.

Bastian, Frederic, Gilles Parmentier, Julien Roux, Sebastien Moretti, Vincent Laudet, and Marc Robinson-Rechavi. 2008. “Bgee: Integrating and Comparing Heterogeneous Transcriptome Data among Species.” *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5109 LNBI:124–31. https://doi.org/10.1007/978-3-540-69828-9_12.

Borlot, Felippe, Brigid M. Regan, Anne S. Bassett, D. James Stavropoulos, and Danielle M. Andrade. 2017. “Prevalence of Pathogenic Copy Number Variation in Adults with Pediatric-Onset Epilepsy and Intellectual Disability.” *JAMA Neurology* 74 (11):1301–11. https://doi.org/10.1001/jamaneurol.2017.1775.

Chen, Chih Ping, Schu Rern Chern, Peih Shan Wu, Shin Wen Chen, Shih Ting Lai, Tzu Yun Chuang, Wen Lin Chen, Chien Wen Yang, and Wayseen Wang. 2018. “Prenatal Diagnosis of a 3.2-Mb 2p16.1-P15 Duplication Associated with Familial Intellectual Disability.” *Taiwanese Journal of Obstetrics and Gynecology* 57 (4). Elsevier Ltd:578–82. https://doi.org/10.1016/j.tjog.2018.06.018.

Fernández-Miñán, A., J. Bessa, J. J. Tena, and J. L. Gómez-Skarmeta. 2016. “Assay for Transposase-Accessible Chromatin and Circularized Chromosome Conformation Capture, Two Methods to Explore the Regulatory Landscapes of Genes in Zebrafish.” *Methods in Cell Biology* 135:413–30. https://doi.org/10.1016/bs.mcb.2016.02.008.

Firth, Helen V., Shola M. Richards, A. Paul Bevan, Stephen Clayton, Manuel Corpas, Diana Rajan, Steven Van Vooren, Yves Moreau, Roger M. Pettett, and Nigel P. Carter. 2009. “DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources.” *American Journal of Human Genetics* 84 (4). The American Society of Human Genetics:524–33. https://doi.org/10.1016/j.ajhg.2009.03.010.

Irastorza-Azcarate, Ibai, Rafael D. Acemel, Juan J. Tena, Ignacio Maeso, José Luis Gómez-Skarmeta, and Damien P. Devos. 2018. “4Cin: A Computational Pipeline for 3D Genome Modeling and Virtual Hi-C Analyses from 4C Data.” *PLoS Computational Biology* 14 (3):1–20. https://doi.org/10.1371/journal.pcbi.1006030.

Kwan, Vickie, Brianna K. Unda, and Karun K. Singh. 2016. “Wnt Signaling Networks in Autism Spectrum Disorder and Intellectual Disability.” *Journal of Neurodevelopmental Disorders* 8 (1). Journal of Neurodevelopmental Disorders:1–10.

https://doi.org/10.1186/s11689-016-9176-3.

Lévy, Jonathan, Aurélie Coussement, Céline Dupont, Fabien Guimiot, Clarisse Baumann, Géraldine Viot, Sandrine Passemard, et al. 2017. “Molecular and Clinical Delineation of 2p15p16.1 Microdeletion Syndrome.” *American Journal of Medical Genetics, Part A* 173 (8):2081–87. https://doi.org/10.1002/ajmg.a.38302.

Loviglio, M. N., M. Leleu, K. Männik, M. Passeggeri, G. Giannuzzi, I. Van Der Werf, S. M. Waszak, et al. 2017. “Chromosomal Contacts Connect Loci Associated with Autism, BMI and Head Circumference Phenotypes.” *Molecular Psychiatry* 22 (6). Nature Publishing Group:836–49. https://doi.org/10.1038/mp.2016.84.

Lovrecic, Luca, Chiara Gnan, Federica Baldan, Alessandra Franzoni, Sara Bertok, Giuseppe Damante, Bertrand Isidor, and Borut Peterlin. 2018. “Microduplication in the 2p16.1p15 Chromosomal Region Linked to Developmental Delay and Intellectual Disability.” *Molecular Cytogenetics* 11 (1). Molecular Cytogenetics:1–6. https://doi.org/10.1186/s13039-018-0388-y.

Lui, Tony T H, Celine Lacroix, Syed M Ahmed, Seth J Goldenberg, Craig A Leach, Avais M Daulat, and Stephane Angers. 2011. “The Ubiquitin-Specific Protease USP34 Regulates Axin Stability and Wnt / β -Catenin Signaling The Ubiquitin-Specific Protease USP34 Regulates Axin Stability And.” *Molecular and Cell Biology* 31 (10):2053. https://doi.org/10.1128/MCB.01094-10.

Mallett, Teresa, Michelle McElroy, Kali Swift, and Suzanne Reuter. 2018. “Chromosome 2 Microdeletion Syndrome in a Newborn with Amniotic Band Sequence.” *South Dakota Medicine : The Journal of the South Dakota State Medical Association*. United States.

Mimouni-Bloch, Aviva, Josepha Yeshaya, Sarit Kahana, Idit Maya, and Lina Basel-Vanagaite. 2015. “A De-Novo Interstitial Microduplication Involving 2p16.1-P15 and Mirroring 2p16.1-P15 Microdeletion Syndrome: Clinical and Molecular Analysis.” *European Journal of Paediatric Neurology* 19 (6). Elsevier Ltd:711–15. https://doi.org/10.1016/j.ejpn.2015.07.013.

Oliva, Carolina A., Jessica Y. Vargas, and Nibaldo C. Inestrosa. 2013. “Wnts in Adult Brain: From Synaptic Plasticity to Cognitive Deficiencies.” *Frontiers in Cellular Neuroscience* 7 (December):1–16. https://doi.org/10.3389/fncel.2013.00224.

Pavone, Piero, Raffaele Falsaperla, Renata Rizzo, Andrea D. Praticò, and Martino Ruggieri. 2018. “Chromosome 2p15-P16.1 Microduplication in a Boy with Congenital Anomalies: Is It a Distinctive Syndrome?” *European Journal of Medical Genetics*, no. April:1–8. https://doi.org/10.1016/j.ejmg.2018.05.001.

Qiao, Y., C. Tyson, M. Hrynchak, E. Lopez-Rangel, J. Hildebrand, S. Martell, C. Fawcett, et al. 2013. “Clinical Application of 2.7M Cytogenetics Array for CNV Detection in Subjects with Idiopathic Autism and/or Intellectual Disability.” *Clinical Genetics* 83

(2):145–54. https://doi.org/10.1111/j.1399-0004.2012.01860.x.

Salinas, Patricia C., and Yimin Zou. 2008. “Wnt Signaling in Neural Circuit Assembly.” *Annual Review of Neuroscience* 31 (1):339–58. https://doi.org/10.1146/annurev.neuro.31.060407.125649.

Wang, Mengmeng, Lihong Xin, Guoqi Cai, Xu Zhang, Xiao Yang, Xiaona Li, Qing Xia, et al. 2017. “Pathogenic Variants Screening in Seventeen Candidate Genes on 2p15 for Association with Ankylosing Spondylitis in a Han Chinese Population.” *PLoS ONE* 12 (5):1–12. https://doi.org/10.1371/journal.pone.0177080.

Wen, Jiadi, Courtney W. Hanna, Sally Martell, Peter C.K. Leung, Suzanne M.E. Lewis, Wendy P. Robinson, Mary D. Stephenson, and Evica Rajcan-Separovic. 2015. “Functional Consequences of Copy Number Variants in Miscarriage.” *Molecular Cytogenetics* 8 (1). https://doi.org/10.1186/s13039-015-0109-8.

Zlotorynski, Eytan. 2018. “Gene Expression: Developmental Enhancers in Action.” *Nature Reviews Genetics* 19 (4). Nature Publishing Group:187. https://doi.org/10.1038/nrg.2018.13.

Supplementary data to the article:

2p15p16.1 microduplication: case report and review

*Qiao Y, *Torres F, Bagheri H, Tena J, Santos-Pereira JM, Martell S, O'Driscoll M, Gomez-Skarmeta JL, Rocha F, Melo C, Lewis S, Maciel P and Rajcan-Separovic E

Sequence of qPCR Primers (Real-time PCR) (5'---3'):

C2orf74 Ex1-2_Forward: GGCCACACCCTTAGAAAACA

C2orf74 Ex1-2_Reverse: TGGCTTTCAACAACTGAAGG

AHSA2 Ex1-2_Forward: GCAACCCAGGAATTGACTGT

AHSA2 Ex1-2_Reverse: GAAGAGCCACAGTGGGAATC

Purchased from IDT Prime time Hs.PT.51.2086875

UPS34 Ex 1-3_Forward: GTGTCAGTACTTCTTGAGCCA

UPS34 Ex 1-3_Reverse: GACCATGCAGCTCGTCAG

Purchased from IDT Prime time Hs.PT.51.2568294

UPS 34 Ex5-7_Forward: TGTCGTAACTCCTGATCCGA

UPS 34 Ex5-7_Reverse: AGCACATGCGTTTATTACAGTTG

Sub-Chapter 2.4 – The contribution of 7q33 copy number variations for intellectual disability

Disclaimer:

The results presented in this sub-chapter were published in ***Neurogenetics***:
Lopes F*, Torres F*, *et al*. (2018) *The contribution of 7q33 copy number variations for intellectual disability*. Neurogenetics. Jan;19(1):27-40. doi: 10.1007/s10048-017-0533-5. Epub 2017 Dec 19.
*both authors contributed equally for this work.

The results presented in this sub-chapter refer to publication of seven patients from different laboratories with 7q33 CNVs. The author of this dissertation 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;
- The initial aCGH analysis of patient 4 was not performed by Torres F.

neurogenetics
https://doi.org/10.1007/s10048-017-0533-5

ORIGINAL ARTICLE

The contribution of 7q33 copy number variations for intellectual disability

Fátima Lopes[1,2] · Fátima Torres[3,4] · Sally Ann Lynch[5] · Arminda Jorge[6,7] · Susana Sousa[1,2] · João Silva[8] · Paula Rendeiro[3] · Purificação Tavares[3] · Ana Maria Fortuna[8] · Patrícia Maciel[1,2]

Received: 22 May 2017 / Revised: 28 November 2017 / Accepted: 29 November 2017

Abstract

Copy number variations (CNVs) at the 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 CNVs in 7q33 alone. All the patients presented mild to moderate intellectual disability (ID), dysmorphic features, and a behavioral phenotype characterized by aggressiveness and disinhibition. One family presents a small duplication in *cis* affecting *CALD1* and *AGBL3* genes, while the other four patients carry two larger deletions encompassing *EXOC4*, *CALD1*, *AGBL3*, and *CNOT4*. This work helps to refine the phenotype and narrow the minimal critical region involved in 7q33 CNVs. Comparison with similar cases and functional studies should help us clarify the relevance of the deleted genes for ID and behavioral alterations.

Keywords 7q33 CNVs · *CALD1* · *AGBL3* · *EXOC4* · *CNOT4* · Duplication

Introduction

Interstitial CNVs in 7q are a rare event and, consequently, poorly characterized. Specifically, there are only 10 reports in the literature of interstitial deletions involving 7q33. Two cases are deletions (7.6 and 7 Mb) derived from chromosomal translocations [1–3]; one case is a small deletion (100 kb) affecting only two genes, *AKR1B1* and *SLC35B4*, in a patient with PHACE syndrome [4]; seven cases show large deletions ranging from cytoband 7q32 to 7q35 [5–11]. A deletion affecting 7q33 only was reported as an abstract but made no mention of the deletion size and genes affected [7]. The two most recent reports in the literature regarding interstitial 7q deletions describe genomic losses in a patient with intellectual disability (ID), language delay, and microcephaly [12] and in a patient with ID and dysmorphisms [11]. Not surprisingly, given the variable sizes of the deletions and duplications in all the reported cases, there is a widely variable phenotypic presentation, most likely due to the large number of genes involved in these variants. A summary of these reports is presented in Table 1.

It is possible to identify several interesting genes that could account for the ID/developmental delay (DD) phenotype associated with 7q33 CNVs, among which are *EXOC4* (exocyst complex component 4), *CNOT4* (CCR4-NOT transcription

Fátima Lopes and Fátima Torres contributed equally

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10048-017-0533-5) contains supplementary material, which is available to authorized users.

✉ Patrícia Maciel
pmaciel@med.uminho.pt

1 Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal

2 ICVS/3B's—PT Government Associate Laboratory, Braga/Guimarães, Portugal

3 CGC Genetics, Porto, Portugal

4 Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, Porto, Portugal

5 Temple Street Hospital, Dublin, Ireland

6 Development Unit, Pediatrics Service, Hospital Centre of Cova da Beira, Covilhã, Portugal

7 CICS—Health Sciences Research Centre, University of Beira Interior, Covilhã, Portugal

8 Center for Medical Genetics Dr. Jacinto Magalhães, Centro Hospitalar do Porto, Porto, Portugal

Published online: 19 December 2017

Table 1 Summary of literature reports of CNVs affecting the 7q33 cytoband

Publication	Affected cases (*n*)	CNV	Description	Size	Genes affected	Phenotype	Inheritance
Malmgren et al. 2005 [2]	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)
Malmgren et al. 2005 [2]	3	Trans + dup	Der(6) ins (6;7) carriers: 7q33-q34 duplication	Dup	Up to 68	Delay speech development, ID, difficulties in school	Inherited (familial translocation leading to del/dup)
Yue et al. 2005 [3]	1	Trans + del	t(7;10)(q33;q23) Der(7) carriers: 7q34-q35 deletion	Del 7 Mb	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 [10]	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 [7]	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 [9]	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 [8]	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 [5]	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 [6]	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 [4]	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

Table 1 (continued)

Publication	Affected cases (*n*)	CNV	Description	Size	Genes affected	Phenotype	Inheritance
Dilzell et al. 2015 [11]	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
Bartsch et al. 1990 [13]	2	Dup	7q33-qter duplication	ND	ND	DD, feeding difficulties (at birth), macrocephaly, chronic obstipation, facial dysmorphisms (high forehead, frontal bossing, deep nasal bridge, epicanthic folds, down-slanting palpebral fissures, microretrognathia), low-set ears, macroglossia, short neck, hypotonia, enlarged subarachnoid spaces and cisterns.	Inherited (maternal translocation leading to dup)

ND not described, *ID* intellectual disability, *DD* developmental delay, *Mb* megabase, *kb* kilobase, *Del* deletion, *Dup* duplication, *Trans* translocation

complex, subunit 4), *CALD1* (Caldesmon 1), and *AGBL3* (ATP/GTP binding protein-like 3). Genotype-phenotype correlations in patients can help define the most relevant genes in this perspective.

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 that *CALD1* and *AGBL3* are major contributors for the ID phenotype of these patients.

Methods

Patients

Patients 1–3 and 5–7 were ascertained within a large study of neurodevelopmental disorders in Portugal, in which the enrollment of the patients and families was done by the referring doctor. Clinical information was gathered in an anonymous database authorized by the Portuguese Data Protection Authority (CNPD). The study was approved by the ethics committee of Center for Medical Genetics Dr. Jacinto Magalhães, National Health Institute Dr. Ricardo Jorge.

Written informed consent was obtained for all participants involved in this publication for the genetic and gene expression studies, blood collection, and for publication of results (including photos).

Molecular karyotyping

Genomic DNA was extracted from peripheral blood using Citogene® DNA isolation kit (Citomed, Portugal) for patients 1, 2, and 3 and QIAsymphony SP (QIAGEN GmbH, Germany) for patients 5, 6, and 7. The aCGH analysis was performed using aCGH Agilent 180 K 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 [14, 15]; Agilent 44 K oligo for patient 4; Affymetrix CytoScan 750 K platform 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 [16]. 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. qPCR reactions were carried out in a 7500-FAST Real-Time PCR machine (Thermo Fisher Scientific, Waltham, MA, USA) using Power SYBR Green® (Thermo Fisher Scientific, Waltham, MA, USA). 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 [17]. Ct values obtained for each test were analyzed in DataAssist ™ software (Thermo Fisher Scientific, Waltham, MA, USA). First-strand complementary DNA (cDNA) was synthesized using SuperScript® III Reverse Transcriptase (RT) (Thermo Fisher Scientific, Waltham, MA, USA).

FISH analysis

FISH was performed in metaphase chromosome spreads from cultured peripheral blood cells from patient 6. The FISH probe was generated using the BAC clone RP11-615F13 (Empire Genomics, Buffalo, NY, USA) and labeled with Green 5-Fluorescein dUTP. Analysis was performed according to the manufacturer's indication, and the fluorescence signals were captured using an Isis Fluorescence Imaging System, MetaSystems (Altlussheim, Germany).

Gene fusion exploratory analysis

Total RNA isolation and cDNA synthesis was performed as described above. In order to determine the presence of a fusion gene at the breakpoints of the 7q33 duplication described in patients 5 and 6, a set of primers were designed for amplification and sequencing of possible gene fusions, namely those linking *CALD1* exon 4 with *AGBL3* exon 16 and *AGBL3* exon 16 with *CALD1* exon 4. The fragments were amplified by PCR and sequenced on an automated DNA-sequencer ABI 3730 XL DNA Analyzer (Thermo Fisher Scientific, Waltham, MA, 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 the occipitofrontal circumference (OFC) was in the 75th centile. Evaluation with the Wechsler Intelligence Scale for Children (third edition) [18] was performed at childhood and showed a full-scale IQ of 42, associated with behavioral changes such as aggressiveness, hyperactivity, and disinhibition. The patient is currently 24 years old. He is dysmorphic, with a bulbous and snub nose (with concave root of the nose), down-slanting palpebral fissures, epicanthic folds, deep set eyes, thin upper lip, poor dental implantation and narrow cleft palate, dysplastic and posteriorly rotated ears, and prognathism (Fig. 1a). Additionally, he also has bushy eyebrows, spiky hair with a frontal cowlick, and two hair whorls at the forehead. The hands present light membranous syndactyly of the second to third digits and feet with brachydactyly, sandal gap, and fetal pads. Brain magnetic resonance imaging (MRI) detected a perivascular space enlargement while the 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. 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 hair whorls at forehead, deep set eyes, epicanthic folds, thin upper lip, and poor dental implantation (Fig. 1b). She also had a short neck; narrow palate; and small dysplastic ears, posteriorly rotated. Abnormalities of the hands and feet included light membranous syndactyly and brachydactyly, respectively. Computed tomography (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 patients 1 and 2. She has some clinical features similar to the daughter, such as facial dysmorphic features (milder) and brachydactyly (Fig. 1c). She has mild ID, although no formal neuropsychological evaluation was performed; she did not complete the fourth grade of school but she has the ability to do household chores.

Patient 4

Patient 4 was born at term to unrelated parents that are phenotypically normal. He was noted to be dysmorphic at birth and was admitted to the hospital because of

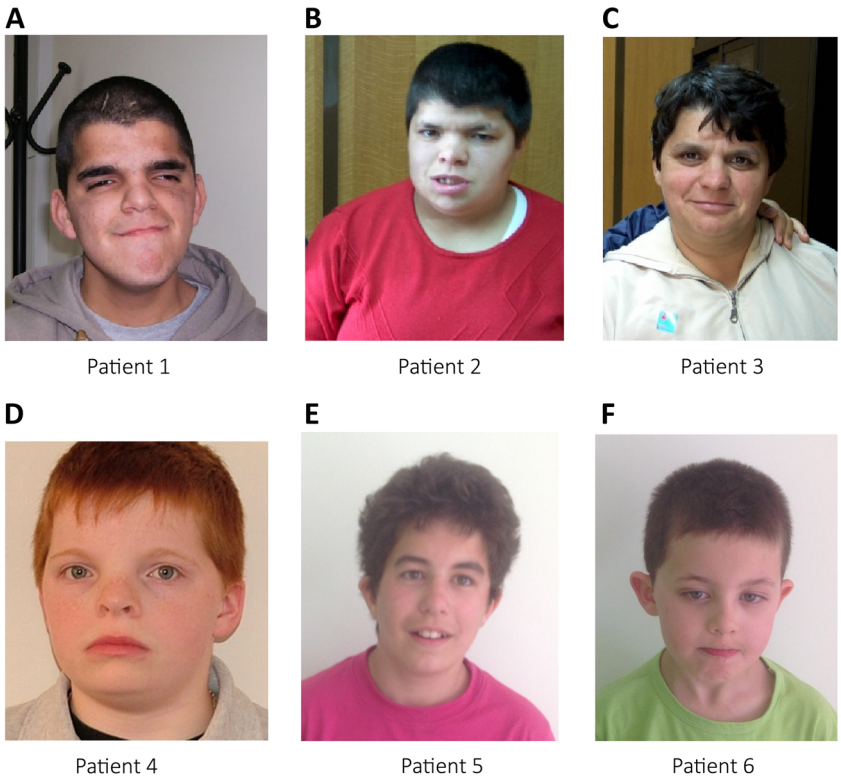


Fig. 1 Facial features of the patients

respiratory grunting. He had feeding problems early on. At 4 months of age, a right inguinal hernia was detected. He was noted to have a wide open anterior fontanelle at 8 months. Otitis media developed and a congenital meatal stenosis required meatoplasty at age 4 years. An evaluation at 10 years old revealed that he weighed 39.75 kg (centile 75) and had a height of 139.8 cm (centile 50) and an OFC of 57.2 cm (all within normal parameters) (Fig. 1d). Currently, he has hypertelorism and myopia. Behavioral issues were noticed at 4 years of age and he was referred to Child 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, lack of attention, impulsiveness, and sexual disinhibition). He does not have significant facial dysmorphisms besides strabismus.

Patient 6 (patient 5's brother) is a 9-year-old boy with mild ID (IQ = 67) and aggressive behavior. He presents normal weight, height, and OFC (at birth and currently) and does not 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 < 60 at 20 years of age), a psychiatric disorder (emotional lability, obsessive behavior), and epilepsy. Although the father was not formally evaluated in the consultation by the responsible physician, he is described as healthy. Due to the mother's health condition, patients 5 and 6 currently live in an institution, since the mother does not have the intellectual and behavioral ability to take care of them. The facial appearance of patients 5 and 6 is presented in Fig. 1e, f.

A clinical comparison between the cases is presented in Table 2.

Molecular findings

aCGH

aCGH in patient 1 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) [19]. A qPCR assay for the *CNOT4*

Table 2 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	Normal	Normal	ND
EEG/seizures	Seizures (single 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
Behavior phenotype	Aggressiveness, hyperactivity, and disinhibition	Aggressiveness, disinhibition	ND	Obsessive-compulsive disorder, emotional lability, aggressiveness, low frustration threshold, impulsiveness, disinhibition	Opposition behavior, lack of attention, impulsiveness, disinhibition	Aggressiveness	Emotional lability, obsessive behavior
Dysmorphisms	Yes	Yes	Yes	Yes	No	No	No
Eyes/ophtalmological examination	Strabismus, epicanthus, sunken eyes	Epicanthus, sunken eyes	ND	Down-slanting palpebral; hypertelorism; Normal vision	Strabismus	Myopia	ND
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	Hair whorls at the forehead, prognathism	Hair whorls at 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)	Family history of ID (brother and mother)	Mother with psychiatric disorder, although without formal assessment

NA not available, *ND* not described, *ID* intellectual disability, *DD* developmental delay

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.

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 three 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 of the CNV in his brother, confirming the presence of three 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). Patient 5 also performed a targeted exome sequencing comprising 4813 genes associated with known clinical phenotypes based on the OMIM database, but no significant pathogenic variants were identified.

A comparison between the molecular alterations identified in the reported patients is presented in Fig. 2 and Table 3.

FISH results

FISH analysis in patient 6 revealed a signal in chromosome 7 that is indicative of the presence of the duplication in tandem, excluding a location in another chromosome (Fig. 3d).

Fusion transcript results

Considering that, according to Newman and colleagues, most of the duplications' CNVs are in tandem and could originate fusion genes at the breakpoints [20], we have designed a set of assays in order to test for the presence of such chimeric transcripts. A fusion transcript between *AGBL3* exon16 and *CALD1* exon4 was detected in patients 5 and 6 (Fig. 3e). This finding is in agreement with the FISH analysis and also indicates that the duplication is not inverted. This hypothesis was also reinforced by the fact that it was not possible to amplify any PCR products indicative for an inverted duplication in patient 5 (Fig. 3e). According to our analysis, the identified *AGBL3-CALD1* gene fusion transcript would lead to an out of frame protein from *CALD1* on (Fig. 3f).

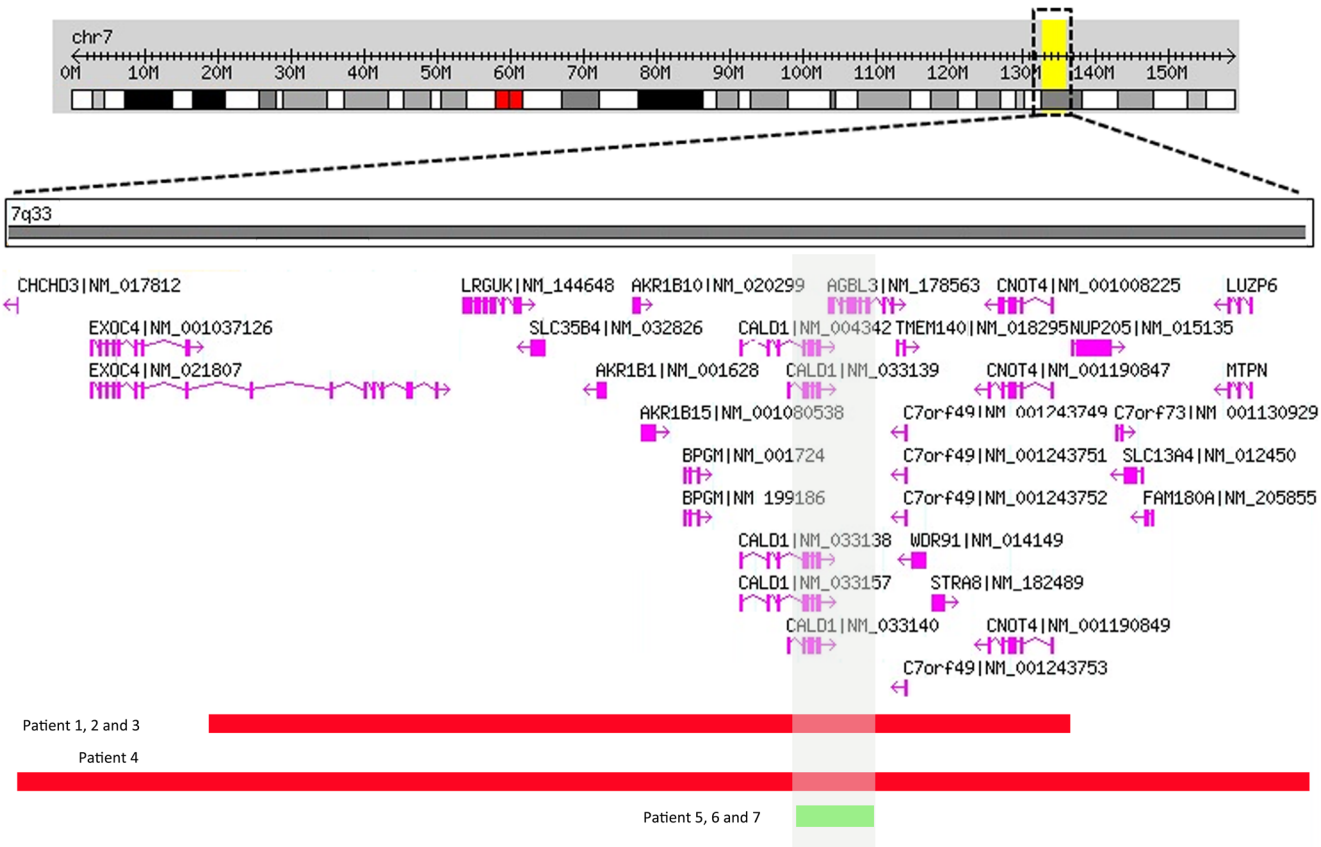


Fig. 2 Schematic representations and overlap of the CNVs found in the patients. A 3 Mb genomic portion of the cytoband 7q33 is shown. RefSeq genes present within the genomic region (in pink; transcriptional direction represented by the arrows) are shown. The overlapping deleted region for all the patients is shaded in gray. Individual red horizontal bars represent deletions. In each CNV, the corresponding patient is indicated

Table 3 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 180 K	–	–	Agilent 44 K	CytoScan 750 K	–	CytoScan 750 K
Confirmation	qPCR (*CNOT4*)	qPCR (*CNOT4*)	qPCR (*CNOT4*)	NP	qPCR (*CALD1*); gene fusion	qPCR (*CALD1*); FISH (in *cis*); gene fusion	qPCR (*CALD1*)
CNV size	2.07 Mb	–	–	3,04 Mb	216 kb	–	209 kb
Interval coordinates (Hg19)	chr7:133,176,651–135,252,871	–	–	chr7:132,766,430–135,802,894	chr7:134,598,205–134,807,358	–	chr7:134,598,205–134,807,358
Inheritance	Maternal	Maternal	ND	de novo	Maternal	Maternal	ND
Genes affected	*AGBL3*, *AKR1B1*, *AKR1B10*, *AKR1B15*, *BPGM*, *C7orf49*, *CALD1*, *CNOT4*, *EXOC4*, *LRGUK*, *NUP205*, *SLC35B4*, *STRA8*, *TMEM140*, *WDR91*, *LOC653739*	*CNOT4* (performed by qPCR; presumably the same ones as patient 1)	*CNOT4* (confirmed 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*, *LOC653739*	*CALD1*, *AGBL3*, *C7orf49*, *LOC653739*	*CALD1* (confirmed by qPCR; presumably the same ones as patient 5)	*CALD1*, *AGBL3*, *C7orf49*, *LOC653739*

NP not performed, *qPCR* quantitative PCR

Constraint metrics

Several constraint metrics for all the 21 genes affected in patient 4 (with the larger CNV) are presented in Table 4. *EXOC4* and *CNOT4* are two of the genes with the highest ranks in the haploinsufficiency score (predicted probability of exhibiting haploinsufficiency); the data was retrieved from DECIPHER database, where the score was determined using the classification model published by Huang et al. [21]).

Discussion

While 7q33 CNVs are rare events, several interstitial deletions of chromosome 7q have been described in the recent past ranging from 7.6 to 13.8 Mb in size all [2, 5, 6, 11, 12]. In this work, we report seven patients (from three families) with 7q33 CNVs, all affecting at least the *CALD1* and *AGBL3* genes (Fig. 2 and Tables 2 and 3). Patients 5, 6, and 7 all present a small 216 kb duplication affecting the *CALD1*, *AGBL3*, and *C7orf49* genes, confirmed by FISH analysis to be in tandem and to lead to the formation of a fusion gene (Fig. 3e). This type of chimeric genes can be related to clinical phenotypes [20]. In fact, an enrichment of rare, brain-expressed chimeric genes was observed in individuals with schizophrenia, with functional studies suggesting a disrupting effect of these fusion genes in critical neuronal pathways [22]. Because both breakpoints occur in intronic regions, the genes are fused by *AGBL3* intron 15 and *CALD1* intron 3, leading to a fused transcript between *AGBL3* exon 16 and *CALD1* exon 4 without any apparent compromise of exonic regions (Fig. 3b, f). The variable amplification of the chimeric messenger RNA (mRNA) indicates that the fused transcript is likely to be degraded, though not completely, since we were able to amplify it in one of the samples collected from patient 5, but not in the other. Also, for patient 6, the fusion transcript was only possible to amplify in the first sample collected using a nested PCR protocol (Fig. 3e). The transcript was also not possible to detect in cultured blood cells for patient 6. This variability in the degradation of the product is not surprising, as it has been described before in a very similar study [23].

Nevertheless, and since the degradation of the *AGBL3-CALD1* chimeric gene does not appear to be complete, it is plausible that it might also contribute for the phenotype, since it could interfere with parent gene function [22].

Additionally, although the pathogenic contribution of the chimeric *AGBL3-CALD1* gene cannot be excluded, the detected rearrangement could also impair the individual expression of both the *AGBL3* and *CALD1* genes.

The *AGBL3* (ATP/GTP binding protein-like 3) gene encodes a cytosolic carboxypeptidase (CCP3) that is able to mediate both the deglutamylation and deaspartylation of tubulin [24]. The deglutamylation of tubulin plays an important

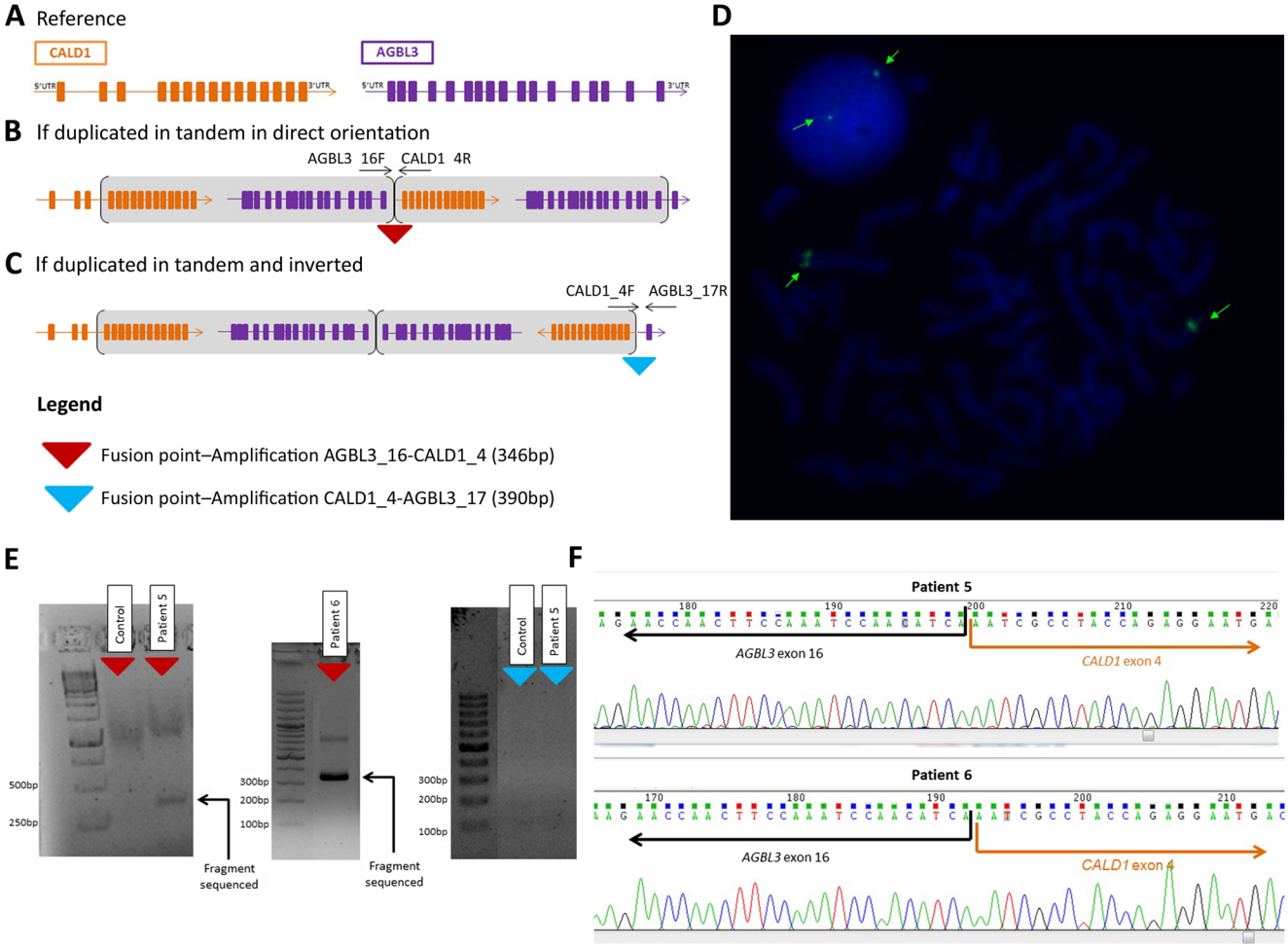


Fig. 3 Schematic representation of *CADL1* and *AGBL3* genes at the 7q33 cytoband. **a** Schematic representation of the normal location of *CADL1* and *AGBL3* genes. **b** The hypothesis that the duplicated region (highlighted in gray) is located in tandem and not inverted is represented, which would hypothetically lead to the formation of a gene fusion between *AGBL3* and *CALD1*. **c** The hypothesis that the duplicated region (highlighted in gray) is located in tandem and inverted is represented, which would hypothetically lead to the formation of a gene fusion between *AGBL3* and *AGBL3* and between *CALD1* and *AGBL3*. The red triangle represents the possible fusion between *AGBL3* and *CALD1* (in the "in tandem not inverted" scenario), the green triangle the possible fusion between *AGBL3* and *AGBL3*, and the blue triangle the possible fusion between *CALD1* and *AGBL3* (in the "in tandem inverted" scenario). **d** FISH analysis for the duplicated region using the BAC clone FISH probe RP11-615F13 located in *CALD1* gene where it is possible to observe a signal indicative of the presence of the duplication in tandem (arrow). **e** PCR amplification of potential fusion products from patients' and controls' cDNA. Only the PCR product corresponding to the *AGBL3* and *CALD1* fusion transcript was possible to amplify (indicated by the arrow) in patients 5 and 6. The absence of a PCR product for both control and patient 5's cDNA on the right (blue triangle) is not in support of the presence of this fusion product ("duplication inverted"). **f** Sanger sequencing of the PCR fragment amplified by AGBL3_16F and CALD1_4R revealed that *AGBL3* exon 16 and *CALD1* exon 4 were fused at the cDNA pf patient 5

role in regulation of the microtubule cytoskeleton, of known relevance for neurons; in fact, the control of the length of the polyglutamate side chains linked to tubulin was shown to be critical for neuronal survival [25], which would make this gene a possible contributor to the patients' phenotype. However, and although tubulin is a key protein in regulation of the microtubule cytoskeleton and this is of known relevance for neurons [25], there is not enough evidence that *AGBL3* does have a function in cytoskeleton regulation in neurons. In fact, according to the GTEx portal [26], *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 [24, 26]. Therefore, the contribution of this gene to the ID phenotype is actually unclear.

The other gene affected by this rearrangement is *CALD1* (Caldesmon1) which encodes for the caldesmon protein and 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 [27]. Since caldesmon has numerous functions in cell motility (such as migration, invasion, and proliferation), executed through the reorganization of the actin cytoskeleton [28], its alteration is likely to have a functional contribution for ID

Table 4 OMIM entrance, haploinsufficiency score, and constraint 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	–	64.11%	–	10dels/11dups	–	–	–	–
AKR1B1	No	–	25.64%	–	9dels/10dups	− 0.32	0.27	0	− 2.25
AKR1B10	No	–	77.14%	–	9dels/11dups	− 0.28	− 0.27	0	− 4.12
AKR1B15	No	–	85.40%	–	9dels/11dups	0.02	− 1.03	0	− 3.36
BPGM	Yes	222800, erythrocytosis due to bisphosphoglycerate mutase deficiency, AR	22.09%	–	11dels/11dups/1SNV	0.16	0.77	0.13	0.5
C7orf49	No	–	80.37%	–	10dels/11dups	− 0.13	− 0.36	0.34	0.56
C7orf73	No	–	24.71%	–	11dels/11dups	–	–	–	–
CALD1	No	–	20.28%	–	10dels/11dups	1.02	− 0.14	1	0.73
CHCHD3	No	–	6.30%	–	9dels/13dups	0.18	0.15	0.04	− 0.13
CNOT4	No	–	6.19%	–	11dels/12dups	0.14	3.38	1	0.81
EXOC4	No	–	4.22%	–	18dels/18dups/1SNV	− 0.09	− 0.27	0	− 1.74
FAM180A	No	–	63.78%	–	11dels/11dups	− 0.26	− 0.33	0.34	1.16
LRGUK	No	–	71.82%	–	10dels/12dups	0.6	− 1.63	0	− 1.4
MTPN	No	–	15.71%	–	11dels/11dups	0.57	2.05	0.75	0.98
NUP205	Yes	616893, nephrotic syndrome, type 13	11.40%	–	11dels/12dups/1SNV	− 0.77	0.87	1	0.18
SLC13A4	No	–	40.17%	–	11dels/11dups	0.64	2.16	0.92	− 0.96
LUZP6	No	–	86.19%	–	11dels/11dups	–	–	–	–
STRA8	No	–	56.99%	–	10dels/11dups	1.42	0.74	0	0.51
TMEM140	No	–	83.19%	–	10dels/11dups	− 0.01	− 0.05	0.04	–
WDR91	No	–	46.24%	–	10dels/11dups	0.7	1.12	0	0.51
SLC35B4	No	–	21.16%	–	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 that a gene is more likely to exhibit haploinsufficiency, and low ranks (e.g., 90–100%) indicate that a gene is more likely NOT to 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

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 through the reorganization of the cytoskeleton and impact on nervous system structure and function [29, 30]. Caldesmon is an important regulator of axon development [31] and may also play a role in synaptogenesis, synaptic plasticity, and dendritic arborization [32].

Four of the patients presented larger deletions also affecting the *EXOC4* and *CNOT4* genes. Considering that patients 1 and 2 are siblings and present the same deletion and very similar phenotypes, the main comparison should be made with patient 4. Concerning the behavioral phenotype, 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 [11]. The deletions' overlap for these four patients is defined by the deletion of patients 1, 2, and 3, resulting in a 2.08-Mb region that includes 15 genes. *EXOC4* (EXOCYST COMPLEX COMPONENT 4) is one of the common genes deleted among the first four patients. *EXOC4* is the human homolog of Sec8 in yeast. *EXOC4*/Sec8 encodes a member of the exocyst complex, broadly expressed in rat brain, localized in the synapses, and which plays a role in neurotransmitter release [33]. Sec8 was described to be involved in the directional movement of

Table 5 Summary of the DECIPHER patients with relatively small (< 2.6 Mb) and overlapping CNVs in 7q33

DECIPHER number	CNV	Size	Genes affected[a]	Inheritance	Pathogenicity	Index phenotype	Parent phenotype
280,233	del	178 kb	*EXOC4*	paternal	ND	ID	ND
253,613	del	45 kb	*EXOC4*	ND	ND	ID, autism, speech delay, hypotonia, obesity, puberty delay, limb abnormalities (short foot and tapered finger)	ND
262,735	del	259 kb	*EXOC4*	ND	ND	ID, behavioral abnormalities, hypotonia, atopic dermatitis	ND
271,567	del	160 kb	*EXOC4*	ND	ND	ND	ND
273,272	del	139 kb	*AKR1B1*	ND	ND	ID	ND
333,171	del	121 kb	*EXOC4*	ND	ND	Behavioral abnormality, language impairment	ND
338,702	del	468 kb	*EXOC4*	ND	ND	Behavioral abnormality, delayed speech and language development	ND
331,287	del	585 kb	*EXOC4*	maternal	Likely pathogenic	Developmental delay	ND
267,399	del	123 kb	*EXOC4*, *LRGUK*	ND	ND	ND	ND
328,659	del	2.6 Mb	*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
282,285	dup	487 kb	*EXOC4*, *LRGUK*, *SLC35B4*	maternal	Uncertain (has a larger de novo pathogenic del in chr9)	Autism	ND
305,865	dup	346Kb	*EXOC4*, *LRGUK*, *SLC35B4*	ND	Uncertain	Autism, global developmental delay	ND
255,520	dup	719 kb	*CHCHD3*, *EXOC4*	Inherited from normal parent	ND	ND	Healthy
251,768	dup	828 kb	*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
256,271	dup	1 Mb	*CHCHD3*, *EXOC4*, *PLXNA4*	De novo	ND	ND	ND

ND not described, *del* deletion, *dup* duplication, *ID* intellectual disability

[a] Genes affected in the DECIPHER patient

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) [34, 35]. Sec8 was also described to bind to postsynaptic density protein-95 (PSD-95), essential for synaptic function [36].

Yue and colleagues reported a patient with DD and macrocephaly who presented a de novo translocation t(7;10)(q33;q23), together with a paternal 7-Mb 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 [3]. 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 allow 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 patients here described and to the patient reported by Yue and colleagues. At this point, we can only hypothesize that *EXOC4* haploinsufficiency 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 [37–39]. The disruption of the proper methylation state of several genes has been shown to be associated with several neurodevelopmental disorders (see [40] 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* [41]. 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 [42–44]. Mersman and colleagues demonstrated that in the yeast, JARID1C homolog protein (Jhd2) levels are also regulated by *CNOT4* via a polyubiquitin-mediated degradation process [41]. 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* [39]. The International Mouse Phenotyping Consortium [45] reports that mice carrying a homozygous intragenic deletion in Cnot4 present preweaning lethality (with complete penetrance), while the heterozygous mice have an abnormal caudal vertebrae morphology, hematopoiesis, and immune system defects [46]. No mention is made to central nervous system (CNS) or cognitive deficits, or craniofacial features in these mice. However, the literature reports 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 reinforces the possible contribution of *CNOT4* for the phenotype in patients 1, 2, 3, and 4.

Besides the analysis of the candidate genes in the 7q33 affected region, it is also important to take into account the patients described in DECIPHER database [19], with deletions and duplications that partially overlap the 7q33 affected region, summarized in Tables 4 and 5. 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 transmitting progenitor is not provided. Additionally, we became aware of the existence of at least two more patients (unrelated, one with speech delay and the other with ID and hypotonia) carrying small deletions affecting only *EXOC4* gene that are inherited from the presumably healthy parents (personal communication by Audrey Briand-Suleau, Cochin Hospital, Paris, France). 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 tandem or not, 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 must be interpreted with caution. In the literature, there are few reports of duplication affecting the 7q33 cytoband [2, 13]. Although their size is significantly larger than that of the duplication in patients 5, 6, and 7, the patients with duplications in this region reported by Malmgren and colleagues appear to have a lighter phenotype than those with the corresponding deletion. As for the report of Bartsch and colleagues, both reported patients have a very severe presentation, which might be due to the duplicated region being very large, encompassing the entire genomic region from 7q33 until the telomere. The difference in size makes the cases reported in these two publications very difficult to compare with patients 5, 6, and 7.

In the Database of Genomic Variants (DGVs), there are no deletions as large as the one present in patient 4. As for the duplicated region, there are no similar duplications in DGV. There are three small deletions in this region (affecting *AGBL3*, *CALD1*, and *TMEM140* genes); however, the presence of these deletions should be interpreted with care, as many of these large studies of control populations might have false calls and/or affected individuals as controls and they cannot be the basis of exclusion of a candidate alteration, especially in the light of other genetic and functional evidence supporting its relevance.

Nevertheless, these six cases raise doubts about the straightforward contribution of *EXOC4* for the NDD phenotype, leaving *AGBL3*, *CNOT4*, and *CALD1* as the more promising candidates.

In summary, this work presents seven patients with interstitial 7q33 CNVs and suggests that *EXOC4*, *CNOT4*, *AGBL3*, and *CALD1* genes are likely contributing for ID and a behavioral phenotype, characterized by aggressiveness and disinhibition. CNVs could impact the phenotype observed in these patients not only by means of haploinsufficiency but also due to the formation of chimeric genes, as the one observed in the patients with the duplication. Chimeras may disrupt critical brain processes, including neurogenesis, neuronal differentiation, and synapse formation, supporting the idea that chimeric genes play a role in the illness, at least in a small number of affected individuals, as recent publications have illustrated [22, 47]. Further studies need to be performed in order to better understand the contribution of each gene within this region to the phenotype.

Acknowledgements We would like to thank all the patients and their families for their participation in the genetic studies and for allowing this publication. We would also like to acknowledge the DECIPHER Consortium, Database of Genomic Variants, and OMIM since this study makes use of data generated by these platforms.

Author contribution FL, FT, SS, and PR performed the molecular studies and analyzed the molecular data. AMF, SAL, AJ, and JS collected clinical data. FL, FT, and PM drafted the paper. PM obtained funding for this study. The study was performed under the direction of PM.

Funding information This work has been funded by FEDER funds, through the Competitiveness Factors Operational Programme (COMPETE), and by National funds, through the Foundation for Science and Technology (FCT), under the scope of the projects PIC/IC/83026/2007, PIC/IC/83013/2007, and POCI-01-0145-FEDER-007038. This work has also been funded by the project NORTE-01-0145-FEDER-000013, supported by the Northern Portugal Regional Operational Programme (NORTE 2020), under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (FEDER). FL was supported by Foundation for Science and Technology (FCT) through the fellowship SFRH/BD/90167/2012.

Compliance with ethical standards

Competing interests The authors declare that they have no conflict of interest.

References

1. Xu Z, Geng Q, Luo F, Xu F, Li P, Xie J (2014) Multiplex ligation-dependent probe amplification and array comparative genomic hybridization analyses for prenatal diagnosis of cytogenomic abnormalities. Mol Cytogenet 7(1):84. https://doi.org/10.1186/s13039-014-0084-5
2. Malmgren H, Malm G, Sahlén S, Karlsson M, Blennow E (2005) 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 139(1):25–31. https://doi.org/10.1002/ajmg.a.30983
3. Yue Y, Grossmann B, Holder SE, Haaf T (2005) De novo t(7;10)(q33;q23) translocation and closely juxtaposed microdeletion in a patient with macrocephaly and developmental delay. Hum Genet 117(1):1–8. https://doi.org/10.1007/s00439-005-1273-4
4. Mitchell S, Siegel DH, Shieh JTC, Stevenson DA, Grimmer JF, Lewis T, Metry D, Frieden I, Blei F, Kayserili H, Drolet BA, Bayrak-Toydemir P (2012) Candidate locus analysis for PHACE syndrome. Am J Med Genet A 158A(6):1363–1367. https://doi.org/10.1002/ajmg.a.35341
5. Rossi E, Verri AP, Patricelli MG, Destefani V, Ricca I, Vetro A, Ciccone R, Giorda R, Toniolo D, Maraschio P, Zuffardi O (2008) A 12Mb deletion at 7q33-q35 associated with autism spectrum disorders and primary amenorrhea. Eur J Med Genet 51(6):631–638. https://doi.org/10.1016/j.ejmg.2008.06.010
6. Petrin AL, Giacheti CM, Maximino LP, Abramides DVM, Zanchetta S, Rossi NF, Richieri-Costa A, Murray JC (2010) Identification of a microdeletion at the 7q33-q35 disrupting the CNTNAP2 gene in a Brazilian stuttering case. Am J Med Genet A 152A(12):3164–3172. https://doi.org/10.1002/ajmg.a.33749
7. Ponnala R, Dalal A (2011) Partial monosomy 7q. Indian Pediatr 48(5):399–401
8. Verma RS, Conte RA, Sayegh SE, Kanjilal D (1992) The interstitial deletion of bands q33-35 of long arm of chromosome 7: a review with a new case report. Clin Genet 41(2):82–86
9. Stallard R, Juberg RC (1981) Partial monosomy 7q syndrome due to distal interstitial deletion. Hum Genet 57(2):210–213
10. Nielsen KB, Egede F, Mouridsen I, Mohr J (1979) Familial partial 7q monosomy resulting from segregation of an insertional chromosome rearrangement. J Med Genet 16(6):461–466. https://doi.org/10.1136/jmg.16.6.461
11. Dilzell K, Darcy D, Sum J, Wallerstein R (2015) Deletion of 7q33-q35 in a patient with intellectual disability and dysmorphic features: further characterization of 7q interstitial deletion syndrome. Case Rep Genet 2015:131852–131855. https://doi.org/10.1155/2015/131852
12. Kale T, Philip M (2016) An interstitial deletion at 7q33-36.1 in a patient with intellectual disability, significant language delay, and severe microcephaly. Case Rep Genet 2016:6046351. https://doi.org/10.1155/2016/6046351
13. Bartsch O, Kalbe U, Ngo TK et al (1990) Clinical diagnosis of partial duplication 7q. Am J Med Genet 37(2):254–257. https://doi.org/10.1002/ajmg.1320370218
14. Krijgsman O, Israeli D, van Essen HF, Eijk PP, Berens MLM, Mellink CHM, Nieuwint AW, Weiss MM, Steenbergen RDM, Meijer GA, Ylstra B (2013) Detection limits of DNA copy number alterations in heterogeneous cell populations. Cell Oncol Dordr 36(1):27–36. https://doi.org/10.1007/s13402-012-0108-2
15. Buffart TE, Israeli D, Tijssen M, Vosse SJ, Mršić A, Meijer GA, Ylstra B (2008) Across array comparative genomic hybridization: a strategy to reduce reference channel hybridizations. Genes Chromosomes Cancer 47(11):994–1004. https://doi.org/10.1002/gcc.20605
16. Jovanovic L, Delahunt B, McIver B et al (2003) Optimising restriction enzyme cleavage of DNA derived from archival histopathological samples: an improved HUMARA assay. Pathology (Phila) 35:70–74
17. Hoebeeck J, van der Luijt R, Poppe B, de Smet E, Yigit N, Claes K, Zewald R, de Jong GJ, 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(1):24–33. https://doi.org/10.1038/labinvest.3700209
18. Wechsler D (1991) Wechsler intelligence scale for children—third edition

19. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Vooren SV, Moreau Y, Pettett RM, Carter NP (2009) DECIPHER: database of chromosomal imbalance and phenotype in humans using Ensembl resources. Am J Hum Genet 84(4):524–533. https://doi.org/10.1016/j.ajhg.2009.03.010
20. Newman S, Hermetz KE, Weckselblatt B, Rudd MK (2015) Next-generation sequencing of duplication CNVs reveals that most are tandem and some create fusion genes at breakpoints. Am J Hum Genet 96(2):208–220. https://doi.org/10.1016/j.ajhg.2014.12.017
21. Huang N, Lee I, Marcotte EM, Hurles ME (2010) Characterising and predicting haploinsufficiency in the human genome. PLoS Genet 6(10):e1001154. https://doi.org/10.1371/journal.pgen.1001154
22. Rippey C, Walsh T, Gulsuner S, Brodsky M, Nord AS, Gasperini M, Pierce S, Spurrell C, Coe BP, Krumm N, Lee MK, Sebat J, McClellan JM, King MC (2013) Formation of chimeric genes by copy-number variation as a mutational mechanism in schizophrenia. Am J Hum Genet 93(4):697–710. https://doi.org/10.1016/j.ajhg.2013.09.004
23. Córdova-Fletes C, Domínguez MG, Delint-Ramirez I, Martínez-Rodríguez HG, Rivas-Estilla AM, Barros-Núñez P, Ortiz-López R, Neira VA (2015) A de novo t(10;19)(q22.3;q13.33) leads to ZMIZ1/PRR12 reciprocal fusion transcripts in a girl with intellectual disability and neuropsychiatric alterations. Neurogenetics 16(4):287–298. https://doi.org/10.1007/s10048-015-0452-2
24. Tort O, Tanco S, Rocha C, Bieche I, Seixas C, Bosc C, Andrieux A, Moutin MJ, Aviles FX, Lorenzo J, Janke C (2014) The cytosolic carboxypeptidases CCP2 and CCP3 catalyze posttranslational removal of acidic amino acids. Mol Biol Cell 25(19):3017–3027. https://doi.org/10.1091/mbc.E14-06-1072
25. Rogowski K, Van DJ, Magiera MM et al (2010) A family of protein-Deglutamylating enzymes associated with neurodegeneration. Cell 143(4):564–578. https://doi.org/10.1016/j.cell.2010.10.014
26. GTEx Portal. http://www.gtexportal.org/home/. Accessed 25 Feb 2017
27. Lin JJ-C, Li Y, Eppinga RD et al (2009) Chapter 1: roles of caldesmon in cell motility and actin cytoskeleton remodeling. Int Rev Cell Mol Biol 274:1–68. https://doi.org/10.1016/S1937-6448(08)02001-7
28. Mayanagi T, Sobue K (2011) Diversification of caldesmon-linked actin cytoskeleton in cell motility. Cell Adhes Migr 5(2):150–159. https://doi.org/10.4161/cam.5.2.14398
29. Fukumoto K, Morita T, Mayanagi T, Tanokashira D, Yoshida T, Sakai A, Sobue K (2009) Detrimental effects of glucocorticoids on neuronal migration during brain development. Mol Psychiatry 14(12):1119–1131. https://doi.org/10.1038/mp.2009.60
30. Mayanagi T, Morita T, 'ichiro HK et al (2008) Glucocorticoid receptor-mediated expression of caldesmon regulates cell migration via the reorganization of the actin cytoskeleton. J Biol Chem 283(45):31183–31196. https://doi.org/10.1074/jbc.M801606200
31. Morita T, Mayanagi T, Sobue K (2012) Caldesmon regulates axon extension through interaction with myosin II. J Biol Chem 287(5):3349–3356. https://doi.org/10.1074/jbc.M111.295618
32. Sobue K, Fukumoto K (2010) Caldesmon, an actin-linked regulatory protein, comes across glucocorticoids. Cell Adhes Migr 4(2):185–189. https://doi.org/10.4161/cam.4.2.10886
33. Hsu SC, Ting AE, Hazuka CD, Davanger S, Kenny JW, Kee Y, Scheller RH (1996) The mammalian brain rsec6/8 complex. Neuron 17(6):1209–1219. https://doi.org/10.1016/S0896-6273(00)80251-2
34. Gerges NZ, Backos DS, Rupasinghe CN, Spaller MR, Esteban JA (2006) Dual role of the exocyst in AMPA receptor targeting and insertion into the postsynaptic membrane. EMBO J 25(8):1623–1634. https://doi.org/10.1038/sj.emboj.7601065
35. Sans N, Prybylowski K, Petralia RS, Chang K, Wang YX, Racca C, Vicini S, Wenthold RJ (2003) NMDA receptor trafficking through an interaction between PDZ proteins and the exocyst complex. Nat Cell Biol 5(6):520–530. https://doi.org/10.1038/ncb990
36. Riefler GM, Balasingam G, Lucas KG et al (2003) Exocyst complex subunit sec8 binds to postsynaptic density protein-95 (PSD-95): a novel interaction regulated by cypin (cytosolic PSD-95 interactor). Biochem J 373(1):49–55. https://doi.org/10.1042/BJ20021838
37. Kruk JA, Dutta A, Fu J, Gilmour DS, Reese JC (2011) The multifunctional Ccr4-not complex directly promotes transcription elongation. Genes Dev 25(6):581–593. https://doi.org/10.1101/gad.2020911
38. Collart MA (2003) Global control of gene expression in yeast by the Ccr4-not complex. Gene 313:1–16. https://doi.org/10.1016/S0378-1119(03)00672-3
39. Grönholm J, Kaustio M, Myllymäki H et al (2012) Not4 enhances JAK/STAT pathway-dependent gene expression in drosophila and in human cells. FASEB J Off Publ Fed Am Soc Exp Biol 26(3):1239–1250. https://doi.org/10.1096/fj.11-195875
40. Rudenko A, Tsai L-H (2014) Epigenetic modifications in the nervous system and their impact upon cognitive impairments. Neuropharmacology 80:70–82. https://doi.org/10.1016/j.neuropharm.2014.01.043
41. Mersman DP, Du H-N, Fingerman IM, South PF, Briggs SD (2009) Polyubiquitination of the demethylase Jhd2 controls histone methylation and gene expression. Genes Dev 23(8):951–962. https://doi.org/10.1101/gad.1769209
42. Abidi F, Holloway L, Moore CA et al (2009) Novel human pathological mutations. Gene symbol: JARID1C. Disease: mental retardation, X-linked. Hum Genet 125:344
43. Ounap K, Puusepp-Benazzouz H, Peters M et al (2012) A novel c.2T > C mutation of the KDM5C/JARID1C gene in one large family with X-linked intellectual disability. Eur J Med Genet 55(3):178–184. https://doi.org/10.1016/j.ejmg.2012.01.004
44. Brookes E, Laurent B, Õunap K, Carroll R, Moeschler JB, Field M, Schwartz CE, Gecz J, Shi Y (2015) Mutations in the intellectual disability gene KDM5C reduce protein stability and demethylase activity. Hum Mol Genet 24(10):2861–2872. https://doi.org/10.1093/hmg/ddv046
45. IMPC | International Mouse Phenotyping Consortium. http://www.mousephenotype.org/. Accessed 21 Sep 2017
46. Cnot4 MGI Mouse Gene Detail - MGI:1859026 - CCR4-NOT transcription complex, subunit 4. http://www.informatics.jax.org/marker/MGI:1859026. Accessed 14 Apr 2017
47. Mayo S, Monfort S, Roselló M, Orellana C, Oltra S, Caro-Llopis A, Martínez F (2017) Chimeric genes in deletions and duplications associated with intellectual disability. Int J Genomics 2017:4798474. https://doi.org/10.1155/2017/4798474

Supplementary material to the article:

The contribution of 7q33 copy number variations for intellectual disability

Fátima Lopes[1,2]*, Fátima Torres[3,4]*, Sally Ann Lynch[5], Arminda Jorge[6,7], Susana Sousa[1,2], João Silva[8], Paula Rendeiro[3], Purificação Tavares[3], Ana Maria Fortuna[8], Patricia Maciel[1,2]

[1]Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal;

[2]ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal;

[3]CGC Genetics, Porto, Portugal;

[4]Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, Porto, Portugal;

[5]Temple Street Hospital, Dublin, Ireland;

[6]Development Unit, Pediatrics Service, Hospital Centre of Cova da Beira, Covilhã, Portugal;

[7]CICS - Health Sciences Research Centre, University of Beira Interior, Covilhã, Portugal;

[8]Center for Medical Genetics Dr. Jacinto Magalhães, Centro Hospitalar do Porto, Praça Pedro Nunes, Porto, Portugal

*Authors contributed equally

Correspondence to:

Patricia Maciel, PhD

Life and Health Sciences Research Institute - School of Medicine

University of Minho

Campus de Gualtar

4710-057 Braga

Portugal

Phone: +351 253604824

Email: pmaciel@ecsaude.uminho.pt

Supplementary Table S1 – Primers used for quantitative PCR confirmation.

Chromosome	Gene	Reference sequence	Primer location	Primer Forward 5'→3'	Primer Reverse 5'→3'	Amplicon size (bp)
Chr 3	*ZNF80*	ENSG00000174255	Exon1	GCTACCGCCAGATTCACACT	AATCTTCATGTGCCGGGTTA	182bp
Chr 7	*CALD1*	ENST00000393118	Exon4	GAATGACGATGATGAAGAGGAG	ACAGTACCTGTTCTGGGCATTC	139bp
Chr 7	*CNOT4*	ENST00000541284	Exon10	CACCGAGCGGTTTATAATTCA	AGACCTGTGTTGTGCTGTGG	164bp
Chr 20	*SDC4*	ENSG00000124145	Exon4	ACCGAACCCAAGAAACTAGA	GTGCTGGACATTGACACCT	101bp

Supplementary Table S2 – Primers used for gene fusion exploratory analysis

Chromosome	Gene	Reference sequence	Primer location	Primer 5'→3'	Fw/Rv	Amplicon	Amplicon size (bp)
Chr 7	*AGBL3*	ENST00000436302	Exon 16	TGCACCACAACTTAAAAAGCA	Fw	AGBL3_16-CALD1_4	346bp
Chr 7	*CALD1*	ENST00000393118	Exon 4	GTCACCTGTCCCAAGGATTC	Rv		
Chr 7	*CALD1*	ENST00000393118	Exon 4	GCCTACCAGAGGAATGACGA	Fw	CALD14-AGBL3_17	390bp
Chr 7	*AGBL3*	ENST00000436302	Exon 17	AGAAGCAGTCTCCCACTTGC	Rv		

CHAPTER 3

The contribution of WES to the identification of epilepsy-causing genes

Disclaimer:

The work presented in this chapter refers to the aCGH and WES analysis performed in patients from the same family diagnosed with epilepsy. These results will be published in an international peer-reviewed journal under the following title and authorship:

Identification of *FERMT2* as a novel epilepsy candidate gene by whole exome sequencing

Torres F, Lopes F, Beleza P, Rouleau G, Maciel P

In preparation

The author of this thesis contributed for the analysis and interpretation of the results of the WES analysis and Sanger sequencing analysis for the reported patients, as well as for the manuscript preparation and discussion of the cases.

Introduction

Epilepsy is a frequent feature of NDs and one of the main contributors to the global burden of disease for neurological disorders (Heyne, Singh, Stamberger, Abou Jamra, Caglayan, Craiu, et al. 2018; Devinsky et al. 2018). Currently, more than 40 genes are considered *bona fide* causes of genetic epilepsies, most of them being linked to developmental and epileptic encephalopathies, severe epilepsies with an early age of onset, and multiple associated comorbidities (Helbig, Riggs, et al. 2018). Nevertheless, the most common group of epilepsies, accounting for approximately one third of all cases, are IGEs, also called GGEs (Helbig et al. 2009). These epilepsies present a complex pattern of inheritance and, to date, only a small fraction of the susceptibility genes has been identified (Hildebrand et al. 2013; Sirven 2015). Some of these genes were identified by positional cloning of rare multi-generational autosomal dominant families, being the ion channel genes, such as GABA receptors (*GABRA1*, *GABRB2*, *GABRB3*, *GABRD*, *GABRG2*), potassium channels (*KCNQ2*, *KCNQ3*), and sodium channels (*SCN1B*, *SCN1A*, *SCN2A*), a substantial group (Hildebrand et al. 2013; Spillane, Kullmann, and Hanna 2016).

In this work we searched for the genetic etiology of disease in a large family with idiopathic generalized epilepsy (IGE) using genome wide approaches.

Subjects and methods

Subjects

The aCGH and WES analysis were performed in patients from the same family diagnosed with epilepsy, according to the ILAE criteria (Scheffer et al. 2017), and/or presenting generalized epileptiform activity (GEE), according to the EEG analysis. Figure 3.1 summarizes the segregation of the disease in family E1, which is, apparently, consistent with an autosomal dominant pattern of inheritance.

The family is constituted by 39 individuals; blood samples were collected and DNA extracted from the following 23 individuals: five affected with epilepsy or epilepsy + GEE (II2, II9, II14, III1, III2); three with GEE (I2, II11, III8); two with epileptiform activity in the right and left temporal lobe (I3, I5) and 13 unaffected (I1, II7, II12, II16, II18, II19, III3, III7, III9, III10, III12, III13, III14). Molecular karyotyping (aCGH) was performed in nine individuals (I1, I2, I3, II2, II9, II14, III1, III2, III8), while WES analysis was performed in three (II14, III1, III2); Sanger sequencing analysis was

performed in the 23 individuals for selected variants, according to the description below.

Molecular karyotyping

Genomic DNA was extracted from peripheral blood using the Citogene® DNA isolation kit (Citomed, Portugal). The aCGH hybridization and analysis was performed using Agilent 180K custom array (GEO GL15397, across-array methodology) (Buffart et al. 2008; Krijgsman et al. 2013) and Nexus Copy Number 5.0 software (BioDiscovery, California, USA) with FASST Segmentation algorithm for data analysis.

Exome sequencing

Exon-containing fragments were captured using the Agilent SureSelect Human All Exon Capture V4 Kit (Santa Clara, California, USA). Sequencing was done using paired-end 2x100 bp read chemistry, with 3 exomes/lane format using the Illumina HiSeq2000 system. Data processing was done 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/) (K. Wang, Li, and Hakonarson 2010).

Analysis pipeline

Single Nucleotide Variants (SNVs) and indels were filtered as follows: 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 and variants with ≥20% reads (≥0.2) were used in downstream analysis. After that, the following variants were removed from the analysis: a) intergenic, intronic and synonymous variants other than those in the consensus splice sites (keeping only coding and canonic splicing occurring variants); b) variants present at a frequency of >0.5% in 1000Genome database; c) variants present at a frequency of >0.5% in ExAC database (GnomAD); d) variants present at >0.1% in an in-house dataset (Dr. Guy

Rouleau's lab database) of 100 similarly sequenced exomes from patients with rare-monogenic diseases unrelated to the clinical phenotypes of our patients, keeping the variant if present in up to one other control in in-house database. Finally, preference was given to variants present in the three family members studied by WES, II14, III1 and III2.

Prioritization of variants and candidate genes

Variants were selected for Sanger confirmation and segregation analysis if (I) Integrative Genomics Viewer (IGV) (J. T. Robinson et al. 2011) visualization confirmed that the position where the variant occurred was sufficiently covered and appeared not to be an artifact, (II) the variant occurred in a candidate gene of apparent functional relevance for the nervous system, (III) the effect of the variant on the gene was predicted to be deleterious. Since only missense variants were detected, further evaluation was performed using Align GVGD (v2007), SIFT (v6.2.0), Mutation Taster (v2013), PolyPhen – 2, as well as nucleotide (phyloP) and amino-acid conservation data. Gene prioritization took into consideration (I) the biological function according to information available in the literature [PubMed database (www.ncbi.nlm.nih.gov/pubmed/), Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), (https://omim.org/), Gene Entrez (www.ncbi.nlm.nih.gov/gene) and GeneCards – the human gene database (www.genecards.org) databases]; (II) brain expression according to the Bgee database (Bastian et al. 2008) and the database from bioproject PRJEB4337 (Fagerberg et al. 2014) (preference was given to genes expressed in the brain); (III) the existence of animal models (preference was given for those associated and/or predicting an epileptic phenotype), and (IV) a positive-match in GeneMatcher software (Sobreira, Schiettecatte, Valle, et al. 2015).

Databases consulted

a. PubMed, www.ncbi.nlm.nih.gov/pubmed/.

b. DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources (DECIPHER), https://decipher.sanger.ac.uk/: repository of an international community of academic departments of clinical genetics and rare disease genomics (Firth et al. 2009).

c. Database of Genomic Variants (DGV), http://dgv.tcag.ca/: curated catalogue of structural variation found in the genomes of control individuals from worldwide populations; only represents structural variations (defined as genomic alterations that involve segments of DNA that are larger than 50bp) identified in healthy control samples (MacDonald et al. 2014).

d. Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), (https://omim.org/).

e. ClinVar, www.ncbi.nlm.nih.gov/clinvar/: freely accessible, public archive of reports of the relationships among human variations and phenotypes, with supporting evidence (Landrum et al. 2014, 2016).

f. The Genome Aggregation Database (gnomAD), http://gnomad.broadinstitute.org/.

g. dSNP: www.ncbi.nlm.nih.gov/projects/SNP/

h HGMD: HGMD (http://www.hgmd.org) (Stenson et al. 2014)

i. GeneMatcher, https://genematcher.org/statistics/: freely accessible Web-based tool developed with the goal of identifying additional individuals with rare phenotypes who had variants in the same candidate disease gene (Sobreira, Schiettecatte, Valle, et al. 2015; Sobreira, Schiettecatte, Boehm, et al. 2015).

j Orphanet, https://www.orpha.net/: reference source of information on rare diseases.

k Bgee database: https://bgee.org/ (Bastian et al. 2008): database to retrieve and compare gene expression patterns in multiple animal species, produced from multiple data types (RNA-Seq, Affymetrix, *in situ* hybridization and EST data). Bgee is based exclusively on curated "normal", healthy, expression data (e.g., no gene knock-out, no treatment, no disease), to provide a comparable reference of normal gene expression.

l. https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJEB4337 (HPA RNA-seq normal tissues): data was obtained by RNA-seq of tissue samples from 95 human individuals representing 27 different tissues in order to determine tissue-specificity of all protein-coding genes (Fagerberg et al. 2014).

m. Data obtained from private laboratory databases, through inter-lab collaboration: ICVS – Universidade do Minho (Braga, Portugal), CGC Genetics (Porto, Portugal), Neurogene Laboratório (Florianópolis, Santa Catarina, Brasil),

AMC – Academisch Medisch Centrum (Amsterdam, The Netherlands), Department of Pathology and Laboratory Medicine, UBC (Vancouver, Canada).

Epilepsy NGS-based panel analysis

Before proceed to Sanger validation and segregation analysis of selected candidate variants, an analysis of 343 genes commonly included in epilepsy NGS-based panels was performed. The coverage of these genes as well as the presence of deleterious variants was checked using either the IGV and/or the software http://gene.iobio.io/ (Miller et al. 2014). The full list of the genes of the panel is described in Supplementary data.

Sanger validation of selected candidate variants

Validation of the selected candidate variants was performed using standard Sanger sequencing. Primers were designed to surround the candidate variant using Primer3Plus software (Untergasser et al. 2007) and evaluated for specificity and SNPs using BiSearch (Tusnády et al. 2005; Arányi et al. 2006) and SNPCheck (https://secure.ngrl.org.uk/SNPCheck/snpcheck.htm) software, respectively.

Primers sequences are detailed in Table S1. PCR reactions were performed using HotStarTaq Master Mix (QIAGEN, Germany) according to the following conditions: 6.3µL HotStar MasterMix, 3.2µL nuclease-free Ambion® DEPC-treated water (ThermoFisher, Waltham, Massachusetts, USA), 2µL of primers mix (10nmol/µL of each primer), 1µL of DNA (at 20ng/µL). The amplification parameters were: activation for 15min at 95ºC; (94ºC for 30 seconds; 58ºC for 1 min; 72ºC for 1 min)x 30 cycles; final extension 10 min at 72ºC; hold at 10ºC. PCR products were purified using ExoProStarTM 1-Step (GE Healthcare Life Sciences, Chicago, Illinois, USA) according to the manufacturer's instructions. Sequencing reactions were performed using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (ThermoFisher, Waltham, Massachusetts, USA), according to the manufacturer's instructions and using 0.5µL of universal primer (M13). The cycling parameters for sequencing reaction: denaturing for 1min at 96ºC; (96ºC for 10 seconds; 50ºC for 5 seconds; 60ºC for 4 min)x 30 cycles; hold at 10ºC. Removal of unincorporated dye terminators was performed using DyeEx 96 Kit (QIAGEN, Germany), and the purified sequencing reactions were loaded onto the 3500xL Genetic Analyzer

(ThermoFisher, Waltham, Massachusetts, USA), according to the manufacturer's instructions. The sequence files generated were analyzed using Mutation Surveyor® software (SoftGenetics LLC, State College Pennsylvania, USA.

Results

Family E1 is constituted by 39 individuals, of which:

- six have epilepsy (II2, II3, II9, II14, III1, III2); four of them have also registered GEE (II2, II14, III1, III2); one has epilepsy but no data concerning EEG (II3);
- three have GEE (I2, II11, III8), but no clinical manifestations of epilepsy; one had also febrile seizures and photoparoxysmal response (III8);
- two have epileptiform activity in the right and left temporal lobe (I3, I5);
- six are presumably healthy but without clinical observation (II4, II5, II6, III4, III5, III6);
- the remaining 22 individuals are healthy;

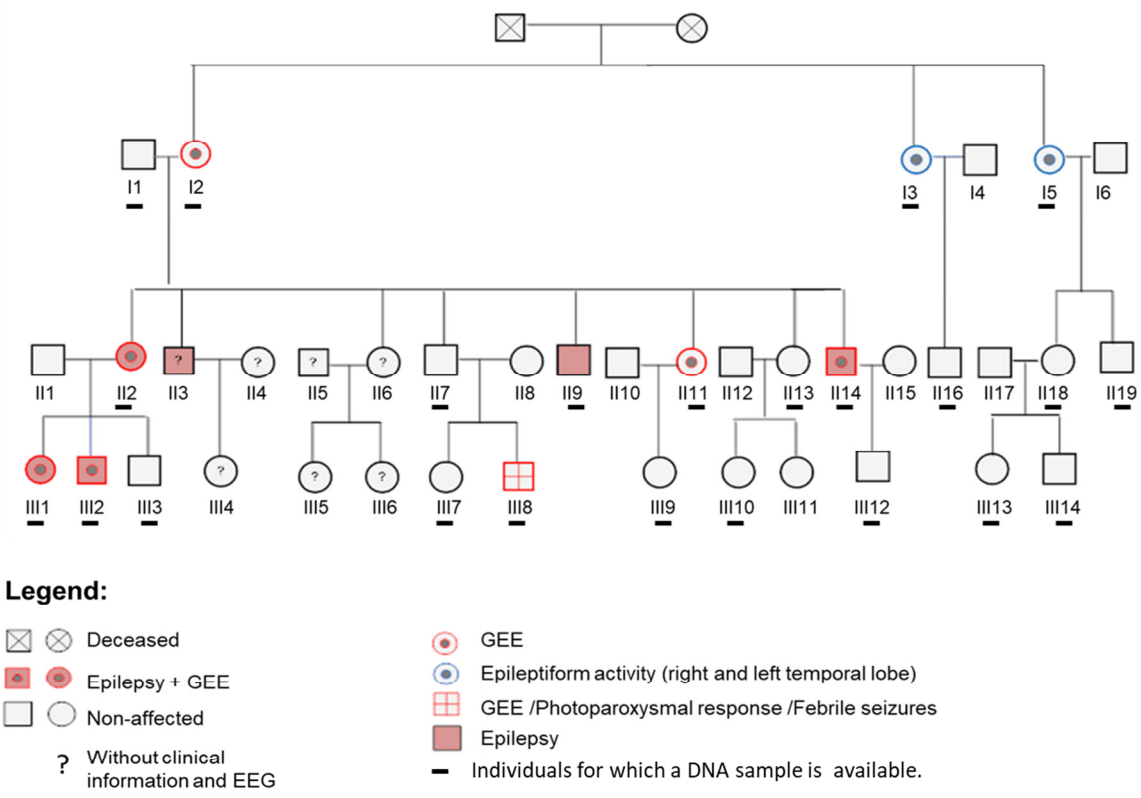


Figure 3.1: **Heredogram of family E1** – patients II-14, III-1 and III-2 were studied by WES.

Molecular Karyotyping

aCGH analysis was performed in nine members of this family (I1, I2, I3, II2, II9, II14, III1, III2 and III8) and allowed us to identify CNVs on cytobands 3q27, 3q29, 15q11, 15q13 and 16p13. Figures 3.2 to 3.6 present the results of the aCGH analysis performed in selected individuals of family E1 as well as the schematic representation of the cytobands region. Table 3.2 resumes the findings of the aCGH analysis.

Figure 3.2: **3q27dup;** A – Schematic representation of the cytoband region in which the CNV was detected. Highlighted rectangle marks the genes affected by the CNV. B – image of the 3q27 duplication.

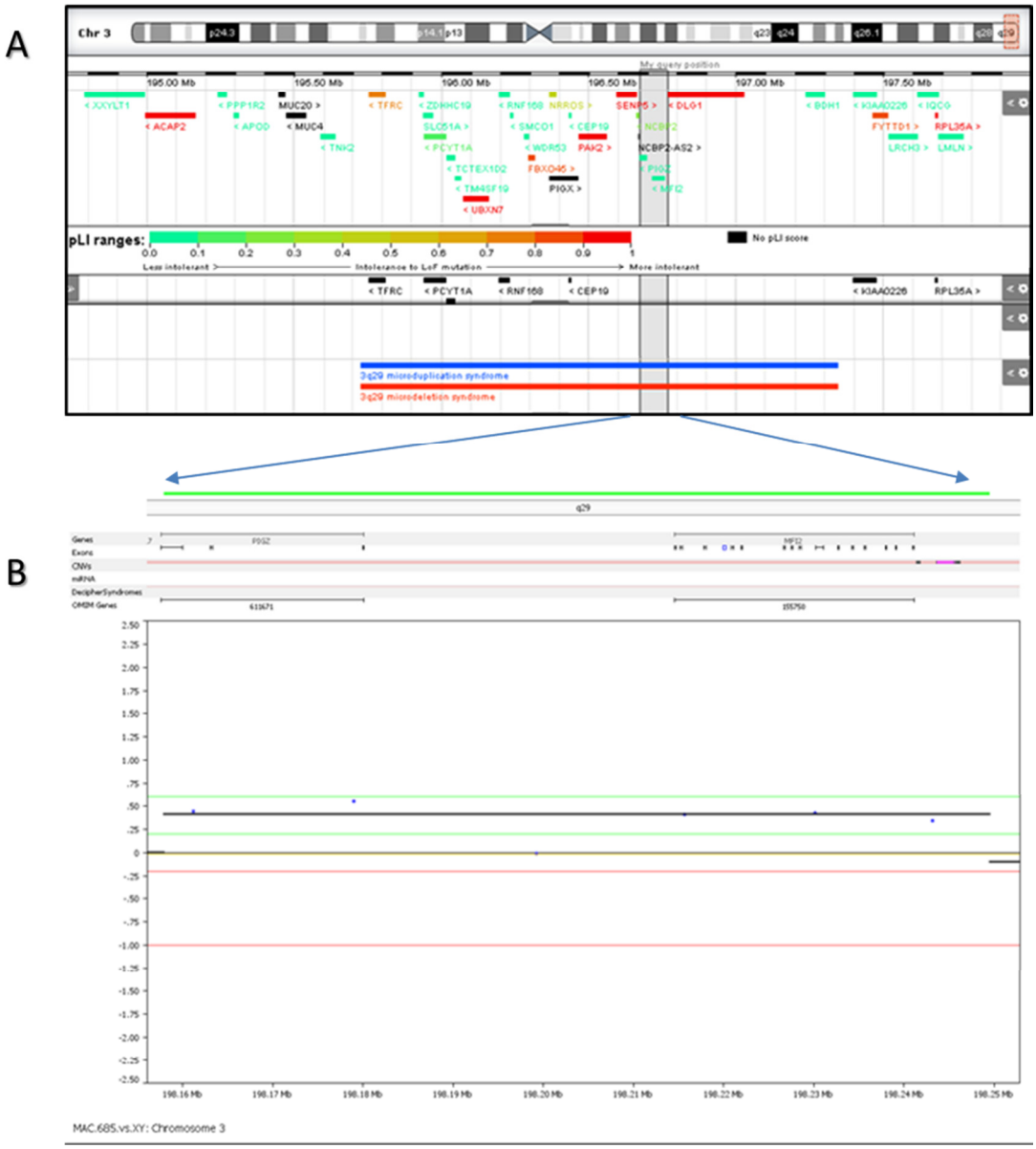


Figure 3.3: **3q29dup**; A – Schematic representation of the cytoband region in which the CNV was detected. Highlighted rectangle marks the genes affected by the CNV. B – image of the 3q29 duplication.

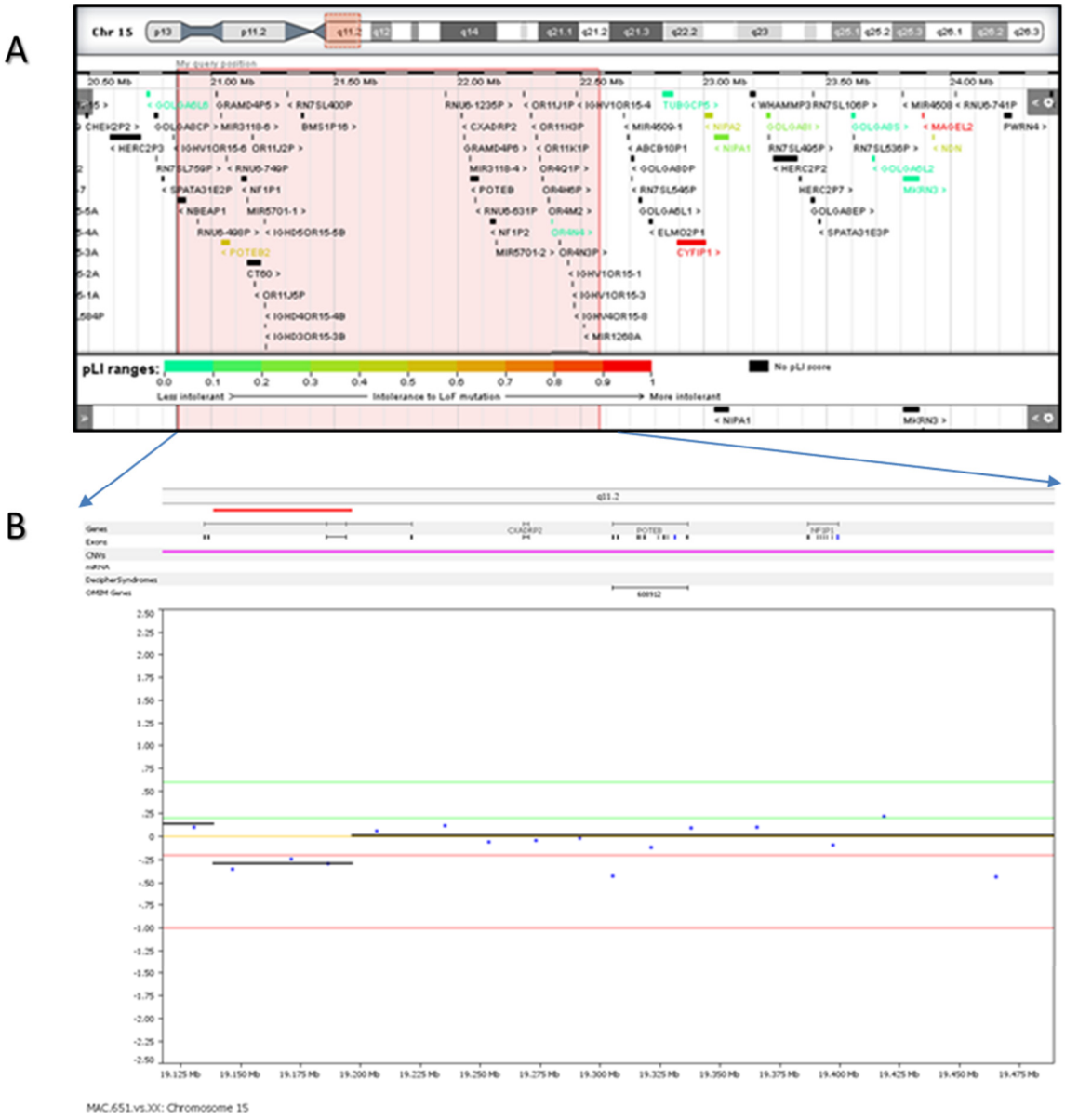


Figure 3.4: 15q11 CNVs; A – Schematic representation of the cytoband region in which the CNVs were detected. Highlighted rectangle marks the genes affected by the CNV. B – image of the 15q11 region.

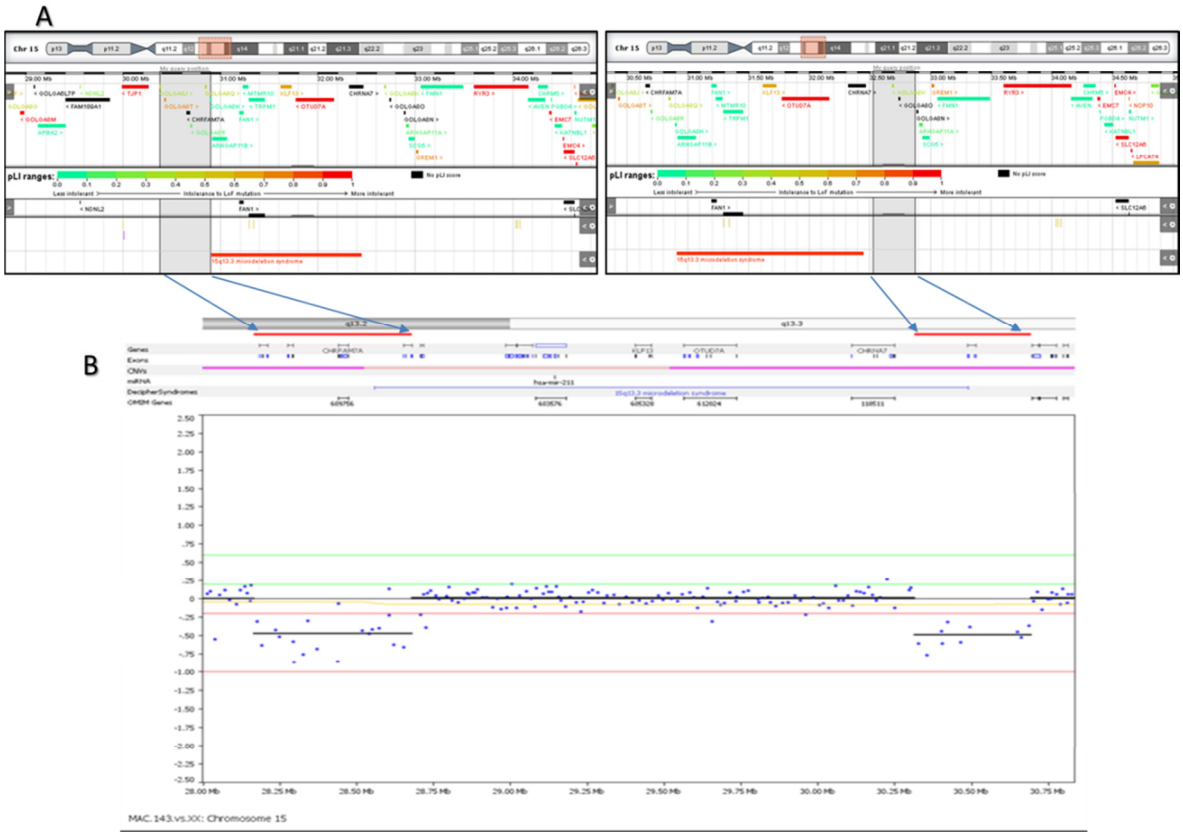


Figure 3.5: **15q13CNVs;** A – Schematic representation of the cytoband region in which the CNVs were detected. Highlighted rectangle marks the genes affected by the CNV. B – image of the 15q13 region.

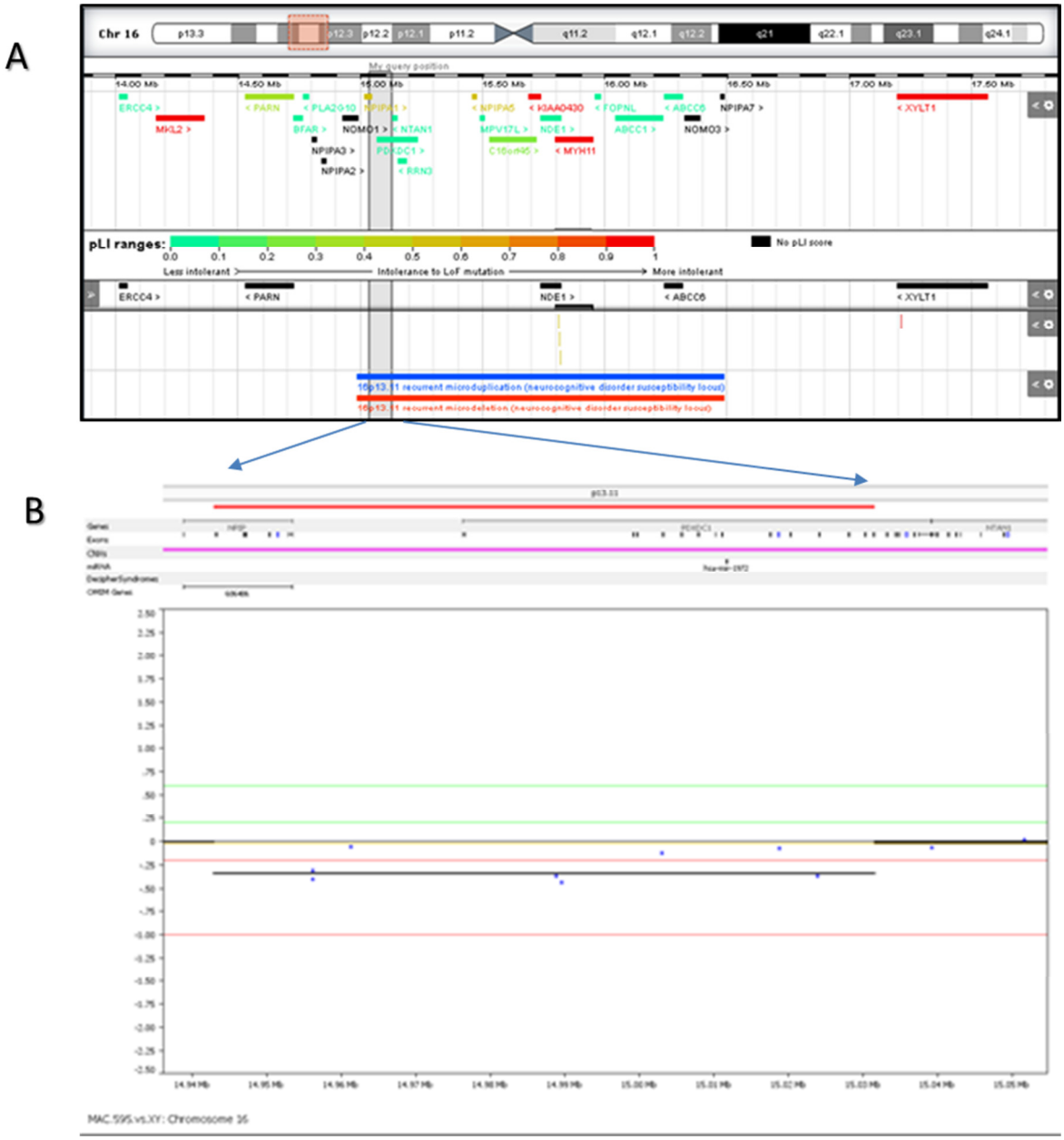


Figure 3.6: **16p13del;** A – Schematic representation of the cytoband region in which the CNVs were detected. Highlighted rectangle marks the genes affected by the CNV. B – image of the 16p13 region.

Table 3.2 – CNVs detected in the aCGH analysis

ISCN description (Hg19)	Type	Size (Kb)	Number of Genes	Genes (name)	DGV controls	Overlap with known CNV syndrome	Key genes	Subjects
3q27.2(184,777,835-184,837,824)X3	dup	59.99	1	*C3orf70*	1/2504 (larger)	No	NA	II2, II9, II14, III1, III2, I1
3q29(196,673,585-196,765,155)x3	dup	91.57	2	*PIGZ, MFI2*	1/2504; 1/29084	3q29 microduplication syndrome	No	III8
15q11.2(19,138,672-19,196,737)x1*	del	58.07	2	*NBEAP1 (BCL8), LOC646214*	Several	No	No	II2, II14, III1, III2, I1
15q11.1-q11.2(18,362,555.-19,922,429)x3*	dup	1560	11	*NBEAP1 (BCL8), LOC646214, CXADRP2, GOLGA6L6, GOLGA8C, LOC727924, NF1P1, OR4M2, OR4N3P, OR4N4, POTEB*	Several	No	No	II9, I2, I3, III8
15q13.2(30,379,275-30,893,521)x1	del	514.2	8	*GOLGA8J, ULK4P3, GOLGA8T, LINC02249, CHRFAM7A, GOLGA8R, ULK4P2, ULK4P1*	Several	near15q13 susceptibility *locus*	includes *CHRFAM7A* (negative regulator *of CHRNA7*)	II9, III1, III2, I3, I1
15q13.3(32,527,264-32,867,806)x1	del	340.5	7	*GOLGA8K, ULK4P3, ULK4P2, ULK4P1, GOLGA80, WHAMMP1, LOC100996255*	Several	No	No	II9, III1, III2, I3, I1
16p13.11(15,035,368-15,124,035)x1	del	88.67	2	*NPIPA1, PDXDC1*	Several	16p13.11 recurrent microdeletion (neurocognitive disorder susceptibility *locus*)	No	II14

Legend: * according to ISCN description (Hg18); NA – not applied

WES analysis

We next performed WES analysis in three affected members of this family (II14, III1 and III2). The average parameters obtained in the WES analysis were: a) mean depth of average coverage for targeted exons: 80X; b) mean quality threshold: coverage >10X in 95.1% of the targeted regions.

WES analysis of three affected members of family E1 (II14, III1 and III2) revealed 184963 coding variants which were filtered as outlined in Figure 3.7. Table 3.3 describe the final list of variants identified as present simultaneously in the three patients studied by WES, as well as bioinformatics predictions of their functional impact and likelihood of pathogenicity. Table 3.4 provides a description of the function, expression and existing cell/animal models concerning the genes with variants present in the three family members studied by WES, as well as their inclusion in classes of priority according to this integrated information. This analysis allowed to group the variants present in all three family members in 3 classes, from the strongest (class I) to the less strong (class III) candidates, according to the prioritization criteria described above. Finally, Table 3.5 presents the segregation of CNVs as well as of WES-identified class I variants, performed by Sanger sequencing in all the available members of the family. Class I variants were found in seven genes: *FERMT2*, *RCN2*, *BAIAP3*, *MAPK8IP3*, *DISC1*, *PARD3* and *CELSR1*.

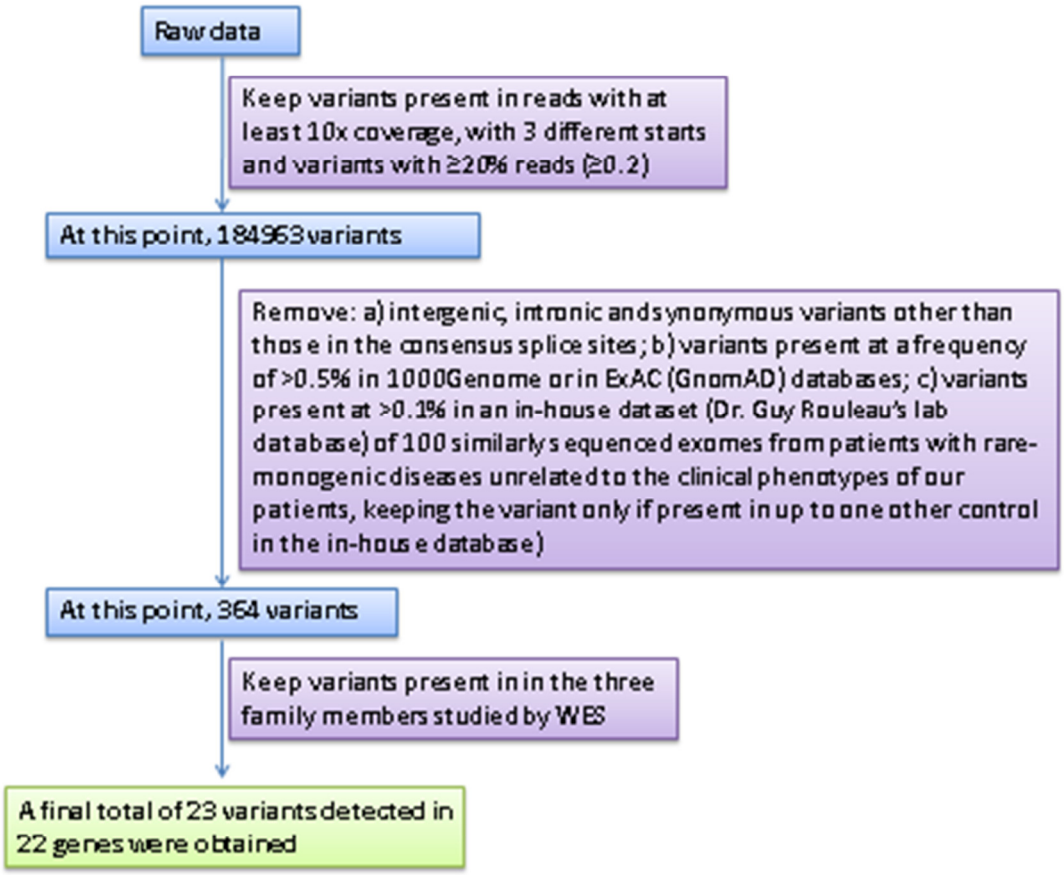


Figure 3.7 – Number of variants identified at different stages of the analysis pipeline performed in the WES study.

Table 3.3– List of variants present in the three family members studied by WES, II14, III1 and III2 and bioinformatic predictions of their impact

Chr	Cytoband	Refseq	Gene	gDNA	cDNA	Protein	Align GVGD	SIFT	Polyphen-2	MutationTaster	Splicing predictions	Nt conservation	Aa conservation	Protein Domain
1	1q42.2	NM_018662.2	*DISC1*	g.231902975C>T	c.1358C>T	p.(Thr453Met)	Class C15 (GV: 46.75 - GD: 62.23)	Deleterious (score: 0.01)	Possibly damaging (score: 0.663)	Polymorphism (prob: 1)	-	Weakly (phyloP: 1.50)	Highly	Prefoldin
10	10p11.22-p11.21	NM_019619.3	*PARD3*	g.34620077G>A	c.2810C>T	p.(Ala937Val)	Class C0 (GV: 250.76 - GD: 0.00)	Tolerated (score: 0.61)	Benign (score: 0.002)	Polymorphism (prob: 1)	-	Weakly (phyloP: 0.53)	Weakly	-
14	14q22.1	NM_006832.2	*FERMT2*	g.53331183G>A	c.1538C>T	p.(Thr513Met)	Class C0 (GV: 215.96 - GD: 58.08)	Deleterious (score: 0.03)	Possibly damaging (score: 0.735)	Disease causing (prob: 0.999)	-	Moderately (phyloP: 2.47)	Moderately	FERM central domain/Band 4.1 domain
15	15q24.3	NM_002902.2	*RCN2*	g.77224726C>A	c.169C>A	p.(Leu56Ile)	Class C0 (GV: 234.72 - GD: 4.86)	Deleterious (score: 0)	Probably damaging (score: 0.991)	Disease causing (prob: 1)	-	Moderately (phyloP: 3.76)	Highly	-
16	16p13.3	NM_003933.4	*BAIAP3*	g.1394822C>T	c.1886C>T	p.(Thr629Ile)	Class C0 (GV: 353.86 - GD: 0.00)	Deleterious (score: 0.05)	Possibly damaging (score: 0.690)	Disease causing (prob: 0.867)	-	Moderately (phyloP: 3.03)	Weakly	-
16	16p13.3	NM_015133.4	*MAPK8IP3*	g.1818553C>T	c.3815C>T	p.(Ser1272Leu)	Class C0 (GV: 236.96 - GD: 58.10)	Tolerated (score: 0.25)	Benign (score: 0.001)	Polymorphism (prob: 1)	-	Weakly (phyloP: 0.77)	Weakly	-
22	22q13.31	NM_014246.1	*CELSR1*	g.46932032T>A	c.1036A>T	p.(Thr346Ser)	Class C0 (GV: 353.86 - GD: 0.00)	Deleterious (score: 0)	Possibly damaging (score: 0.694)	Polymorphism (prob: 0.965)	-	Moderately (phyloP: 2.79)	Highly	Cadherin
1	1p35.2	NM_024522.2	*NKAIN1*	g.31660926C>T	c.163G>A	p.(Val55Met)	Class C0 (GV: 248.74 - GD: 0.00)	Tolerated (score: 0.05)	Possibly damaging (score: 0.467)	Disease causing (prob: 0.999)	-	Weakly (phyloP: 1.82)	Moderately	Na,K-Atpase Interacting protein
2	2q36.3	NM_030623.3	*SPHKAP*	g.228883801G>A	c.1769C>T	p.(Pro590Leu)	Class C0 (GV: 144.08 - GD: 0.00)	Tolerated (score: 0.77)	Benign (score: 0.000)	Polymorphism (prob: 1)	-	Weakly (phyloP: 0.45)	Weakly	-
2	2q36.3	NM_030623.3	*SPHKAP*	g.228884530G>A	c.1040C>T	p.(Ser347Phe)	Class C0 (GV: 111.88 - GD: 72.36)	Deleterious (score: 0.01)	Possibly damaging (score: 0.549)	Polymorphism (prob: 1)	-	Weakly (phyloP: 0.77)	Moderately	-

4	4p11	NM_015030.1	*FRYL*	g.48583610T>C	c.1999A>G	p.(Asn667Asp)	Class C0 (GV: 353.86 - GD: 0.00)	Tolerated (score: 0.06)	Benign (score: 0.028)	Disease causing (prob: 0.837)	Predicted change at acceptor site 13 bps upstream: -8.5% MaxEnt: 0.0% NNSPLICE: -17.0% SSF: 0.0%	Weakly (phyloP: 1.17)	Weakly	Armadillo-type fold
7	7q11.23	NM_001202560.2	*BUD23* (*WBSCR22*)	g.73111828G>A	c.734G>A	p.(Arg245Gln)	-	Tolerated (score: 0.50)	Benign (score: 0.006)	Polymorphism (prob: 1)	-	Not conserved (phyloP: -1.57)	No data available	Uncharacterised protein family, methyltransferase, Williams-Beuren syndrome
12	12p12.1	NM_003034.3	*ST8SIA1*	g.22440184T>C	c.280A>G	p.(Met94Val)	Class C0 (GV: 231.89 - GD: 0.00)	Tolerated (score: 0.61)	Benign (score: 0.010)	Disease causing (prob: 0.993)	-	Weakly (phyloP: 1.98)	Moderately	Glycosyl transferase, family 29 Sialyltransferase
14	14q21.2	NM_018353.4	*MIS18BP1*	g.45711988C>A	c.634G>T	p.(Ala212Ser)	Class C0 (GV: 353.86 - GD: 0.00)	Deleterious (score: 0.03)	Benign (score: 0.002)	Polymorphism (prob: 1)	-	Not conserved (phyloP: -1.73)	Moderately	-
16	16p13.3	NM_145294.4	*WDR90*	g.706846G>A	c.2317G>A	p.(Ala773Thr)	Class C0 (GV: 353.86 - GD: 0.00)	Tolerated (score: 0.08)	Benign (score: 0.068)	Polymorphism (prob: 1)	-	Not conserved (phyloP: -0.04)	Weakly	WD40 repeat/ WD40-repeat-containing domain
1	1p22.1	NM_000350.2	*ABCA4*	g.94467548C>G	c.6148G>C	p.(Val2050Leu)	Class C25 (GV: 0.00 - GD: 30.92)	Deleterious (score: 0)	Possibly damaging (score: 0.899)	Disease causing (prob: 1)	Predicted change at acceptor site 1 bps upstream: -12.5% MaxEnt: -24.5% NNSPLICE: -0.6% SSF: -6.7%	Highly (phyloP: 5.61)	Highly	ABC transporter-like/ Rim ABC transporter/ AAA+ ATPase domain/ P-loop containing nucleoside triphosphate hydrolase
2	2p24.1	NM_000384.2	*APOB*	g.21228339A>T	c.11401T>A	p.(Ser3801Thr)	Class C0 (GV: 353.86 - GD: 0.00)	Tolerated (score: 0.57)	Possibly damaging (score: 0.497)	Polymorphism (prob: 1)	-	Weakly (phyloP: 0.37)	Moderately	-
2	2q37.3	NM_022134.2	*GAL3ST2*	g.242743169C>T	c.785C>T	p.(Ala262Val)	Class C0 (GV: 118.37 - GD: 35.60)	Tolerated (score: 0.21)	Benign (score: 0.004)	Polymorphism (prob: 1)	-	Weakly (phyloP: 0.53)	Weakly	Galactose-3-O-sulfotransferase

6	6q23.3	NM_006290.3	*TNFAIP3*	g.138201240A>C	c.1939A>C	p.(Thr647Pro)	Class C0 (GV: 244.67 - GD: 0.00)	Tolerated (score: 0.17)	Benign (score: 0.001)	Polymorphism (prob: 1)	-	Not conserved (phyloP: -0.28)	Moderately	-
8	8q13.3	NM_016027.2	*LACTB2*	g.71556453_71556461del	c.431_439del	p.(Gly144_Thr146del)	-	-	-	Polymorphism (prob: 0.89)	Predicted change at acceptor site 18 bps upstream: -5.7% MaxEnt: 0.0% NNSPLICE: -11.5% SSF: 0.0%	-	-	-
11	11q13.1	NM_053054.3	*CATSPER1*	g.65793706C>T	c.145G>A	p.(Gly49Ser)	Class C0 (GV: 60.00 - GD: 46.82)	Deleterious (score: 0.05)	Probably damaging (score: 0.995)	Polymorphism (prob: 1)	-	Weakly (phyloP: 0.12)	Highly	-
16	16p13.13	NM_144674.1	*TEKT5*	g.10775949T>C	c.764A>G	p.(Asp255Gly)	Class C0 (GV: 171.76 - GD: 51.67)	Tolerated (score: 0.59)	Probably damaging (score: 0.993)	Disease causing (prob: 1)	-	Moderately (phyloP: 4.00)	Highly	Tektin
17	17q21.33	NM_175575.5	*WFIKKN2*	g.48917619C>T	c.970C>T	p.(Pro324Ser)	Class C0 (GV: 97.78 - GD: 73.35)	Tolerated (score: 0.16)	Benign (score: 0.032)	Disease causing (prob: 1)	-	Highly (phyloP: 6.02)	Highly	-

Legend: Nt – nucleotide; aa – amino acid

Table 3.4 – Description of the genes with variants present in the three family members studied by WES

Gene		Protein	Function[a]	Expression[b]	Constitutional disease association[c]	Cell/animal models
Class I	*FERMT2* (OMIM: 607746)	Fermitin family member 2 (also known as kindling 2)	FERMT2 or kindlin 2 is important in enhancing Wnt/β-catenin signaling by selectively binding to the active β-catenin (Yu et al. 2012).	Ubiquitous expression in fat (RPKM 84.8), endometrium (RPKM 56.1) and 22 other tissues (brain: RPKM 20.634)	-	The loss of kindlin 2 results in early embryonic lethality in mice (Montanez et al. 2008). The proteomic profiling of epileptogenesis in a rat model showed a strong over-expression of FERMT2, confirmed by immunohistochemistry, in the rat hippocampus ten days following status epilepticus (Keck et al. 2018)
	RCN2 (OMIM: 602584)	Reticulocalbin 2 (EF-hand calcium-binding protein)	It specifically localizes in the endoplasmic reticulum, and was identified as a binding partner for neuronal pentraxins (Kirkpatrick et al. 2000), which are thought to play a role in synaptic refinement and clustering of AMPA receptors (Bjartmar et al. 2006).	Ubiquitous expression in testis (RPKM 51.1), brain (RPKM 32.7) and 24 other tissues	Differential expression was observed for *RNC2* gene in lymphoblastoid cell lines (LCLs), obtained from both CAE and JAE patients, thus implicating dysregulation of *RCN2* in IAE (Helbig, Matigian, et al. 2008)	-
	BAIAP3 (OMIM: 604009)	BAI1 associated protein 3 (member of the mammalian uncoordinated 13 (Munc13) protein family of synaptic regulators of neurotransmitter exocytosis)	p53-target gene, encodes a seven-span transmembrane protein that interacts with the cytoplasmic region of brain-specific angiogenesis inhibitor-1 (BAI1) and controls the fate of dense-core vesicles in neuroendocrine cells (Zhang et al. 2017).	Broad expression in adrenal (RPKM 9.7), brain (RPKM 8.5) and 19 other tissues; with a striking expression pattern in amygdalae, hypothalamus and periaqueductal gray (Wojcik et al. 2013)	A *BAIAP3* missense variant was detected in a search for hypothalamic signalling genes related to extreme obesity (Mariman et al. 2015)	*BAIAP3* knockout mice display increased seizure propensity and anxiety behaviour (Wojcik et al. 2013).
	MAPK8IP3 (OMIM: 605431)	Mitogen-activated protein kinase 8 interacting protein 3, also known as c-Jun N-terminal kinase (JNK)-interacting protein 3 (JIP3), or JNK stress-activated protein kinase-associated protein 1 (JSAP1)	Scaffold protein mainly involved in the regulation of the pro-apoptotic signaling cascade mediated by c-Jun N-terminal kinase (JNK).	Ubiquitous expression in brain (RPKM 22.4), spleen (RPKM 12.7) and 24 other tissues	Two *de novo* missense variants identified by WES in a cohort of Smith-Magenis-like syndrome patients (Berger et al. 2017), and in a cohort of autism patients (Iossifov et al. 2014).	Increased expression of JIP3 in a kainic acid (KA)-induced mouse model of epileptic seizures; evidence that JIP3 is involved in epileptic seizures and the regulation of neuronal response to excitotoxic stress (Z. Wang et al. 2015). In *Caenorhabditis elegans* (*C. elegans*): UNC-16/JIP3 protein participates in the trafficking and biogenesis of synaptic vesicle proteins transport carriers (Choudhary et al. 2017).
	DISC1 (OMIM: 605210)	DISC1 scaffold protein	DISC1 is involved in the regulation of multiple aspects of embryonic and adult neurogenesis (J. Y. Kim et al. 2009); it plays a role in the microtubule network formation (Miyoshi et al. 2004).	Ubiquitous expression in placenta (RPKM 2.1), ovary (RPKM 1.4) and 24 other tissues (brain: RPKM 1.236)	-	-
	PARD3 (OMIM: 606745)	Adapter protein (Par-3 Family Cell Polarity Regulator) that belongs to the PARD protein family.	These proteins are essential for asymmetric cell division and polarized growth, having also a central role in the formation of epithelial tight junctions (Johansson, Driessens, and Aspenström	Ubiquitous expression in esophagus (RPKM 6.9), skin (RPKM 6.8) and 24 other tissues (brain: RPKM 1.367)	*PARD3* polymorphisms have been associated with SZ (S. K. Kim et al. 2012) while *PARD3* gene pathogenic variants have been	Dysfunction of *PARD3* gene in conjunction with dynamic HIPPO-signaling, causes severe malformations of the cortex, including increased volume, systematic and drastic changes in neuronal subtype composition, and

			2000; X. Chen et al. 2017), and the establishment of neuronal polarity and normal axon formation in cultured hippocampal neurons (Khazaei and Püschel 2009; X. Chen et al. 2017).		associated with neural tube defects (X. Chen et al. 2013, 2017; Gao et al. 2017).	massive ribbon-like heterotopia in mice, that also exhibit elevated seizure susceptibility (Liu et al. 2018).
	CELSR1 (OMIM: 604523)	Cadherin EGF LAG seven-pass G-type receptor 1 (non-classic type cadherin)	Developmentally regulated, neural-specific gene; member of the flamingo subfamily, proposed to have key roles in epithelial planar cell polarity (Goffinet and Tissir 2017).	Broad expression in kidney (RPKM 10.0), skin (RPKM 8.2) and 16 other tissues (brain: RPKM 0.146)	Pathogenic variants in *CELSR1* gene were associated with diverse neural tube defect phenotypes in humans (Z. Chen et al. 2018), such as craniorachischisis (A. Robinson et al. 2012; Allache et al. 2012) and *spina bifida* (Lei et al. 2014)	Severe neural tube defects were observed in homozygous mutant mice, with missense mutations within the coding region of *Celsr1* gene (Curtin et al. 2003). *Celsr1*-deficient mice show neural progenitor fate decision defects, leading to a reduced number of cortical neurons, abnormal brain architecture, microcephaly and behavioural impairment (Boucherie et al. 2018).
Class II	*SPHKAP* (OMIM: 611646)	SPHK1 (Sphingosine Kinase Type 1) interacting protein, AKAP domain containing	Anchoring protein that binds preferentially to the type I regulatory subunit of c-AMP-dependent protein kinase (PKA type I) and targets it to distinct subcellular compartments. May act as a converging factor linking cAMP and sphingosine signaling pathways.	Biased expression in heart (RPKM 11.8), brain (RPKM 8.0) and ovary (RPKM 1.304)	Has been recently described in a genome-wide association study as a susceptibility gene for SZ (Ikeda et al. 2018)	-
	ST8SIA1 (OMIM: 601123)	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 1	Type II membrane protein that catalyzes the transfer of sialic acid from CMP-sialic acid to GM3 to produce gangliosides GD3 and GT3; may be found in the Golgi apparatus and is a member of glycosyltransferase family 29.	Broad expression in brain (RPKM 1.5), adrenal (RPKM 0.7) and 19 other tissues	*ST8SIA1* variants were associated with thyroid-associated ophthalmopathy (Park et al. 2017)	-
	BUD23 (WBSCR22) (OMIM: 615733)	rRNA methyltransferase and ribosome maturation factor	Protein containing a nuclear localization signal and an S-adenosyl-L-methionine binding motif typical of methyltransferases, suggesting that it may act on DNA methylation.	Ubiquitous expression in testis (RPKM 63.4), thyroid (RPKM 24.2) and 25 other tissues (brain: RPKM 15.216)	This gene is 1 of at least 20 genes commonly deleted in Williams-Beuren syndrome (WBS; 194050) (Doll and Grzeschik 2001)	-
	NKAIN1 (OMIM: 612871)	Sodium/potassium transporting ATPase interacting 1	Member of a family of mammalian proteins with similarity to Drosophila Nkain that interacts with the beta subunit of Na,K-ATPase (ATP1B1; MIM 182330) (Gorokhova et al. 2007)	Biased expression in adrenal (RPKM 5.0), brain (RPKM 4.0) and 4 other tissues	Associated with significant risk for alcohol dependence in subjects of European descent in genome-wide association studies (GWASs) (Zuo et al. 2013)	-
	FRYL (OMIM: No entry)	FRY like transcription coactivator	Maintains the integrity of polarized cell extensions during morphogenesis, regulates the actin cytoskeleton and in patterning sensory neuron dendritic fields by promoting avoidance between homologous dendrites as well as by limiting dendritic branching (by similarity). May function as a	Ubiquitous expression in colon (RPKM 15.9), lymph node (RPKM 5.8) and 24 other tissues (brain: RPKM 5.557)	-	-

			transcriptional activator.			
	MIS18BP1 (OMIM: 618139)	MIS18 binding protein 1	Required for recruitment of CENPA to centromeres and normal chromosome segregation during mitosis.	Ubiquitous expression in bone marrow (RPKM 16.3), lymph node (RPKM 12.1) and 24 other tissues (brain: RPKM 1.968)	-	-
	WDR90 (OMIM: No entry)	WD repeat domain 90	WDR90 is required for ciliogenesis (Hamel et al. 2017)	Broad expression in testis (RPKM 12.0), bone marrow (RPKM 4.3) and 23 other tissues (brain: RPKM 0.896)	Two truncating variations in *WDR90* and *EFCAB5* genes were identified by WES in a family with a monozygotic twin pair concordant for ASD; additionally, they were not detected in 257 ASD patients, 677 SZ patients or 667 controls in a follow-up study (Egawa et al. 2015)	-
Class III	*ABCA4* (OMIM: 601691)	ATP binding cassette subfamily A member 4	Member of the superfamily of ATP-binding cassette (ABC) transporters; retina-specific ABC transporter with N-retinylidene-PE as a substrate	Expressed exclusively in retina; biased expression in kidney (RPKM 2.3), small intestine (RPKM 0.8)	Mutations in this gene are found in patients diagnosed with Stargardt disease, retinitis pigmentosa-19, cone-rod dystrophy type 3, early-onset severe retinal dystrophy, fundus flavimaculatus and macular degeneration age-related 2.	-
	CATSPER1 (OMIM: 606389)	Cation channel sperm associated 1	This gene belongs to a family of putative cation channels that are specific to spermatozoa and localize to the flagellum. The protein family features a single repeat with six membrane-spanning segments and a predicted calcium-selective pore region.	Restricted expression toward testis (RPKM 8.9)	Spermatogenic failure-7 (SPGF7, MIM 612997) is caused by homozygous mutation in the *CATSPER1* gene	-
	TEKT5 (OMIM: No entry)	Tektin 5	Probably a structural component of the sperm flagellum.	Restricted expression toward testis (RPKM 3.8)	-	.
	WFIKKN2 (OMIM: 610895)	WAP, follistatin/kazal, immunoglobulin, kunitz and netrin domain containing 2	This gene encodes a WFIKKN1-related protein which has the same domain organization as the WFIKKN1 protein. The WAP-type, follistatin type, Kunitz-type, and NTR-type protease inhibitory domains may control the action of multiple types of proteases. Probably has serine protease- and metalloprotease-inhibitor activity.	Biased expression in ovary (RPKM 30.8) and testis (RPKM 1.8)	-	-
	APOB (OMIM: 107730)	Apolipoprotein B	Major protein constituent of chylomicrons (apo B-48), LDL (apo B-100) and VLDL (apo B-100). Apo B-100	Biased expression in liver (RPKM 415.6), small intestine (RPKM 182.7)	Mutations in this gene or its regulatory region cause hypobetalipoproteinemia,	-

			functions as a recognition signal for the cellular binding and internalization of LDL particles by the apoB/E receptor.		normotriglyceridemic hypobetalipoproteinemia, and hypercholesterolemia due to ligand-defective apoB, diseases affecting plasma cholesterol and apoB levels.	
	GAL3ST2 (OMIM: 608234)	galactose-3-O-sulfotransferase 2	This enzyme catalyzes sulfonation by transferring a sulfate group to the hydroxyl at C-3 of nonreducing beta-galactosyl residues, and it can act on both type 1 and type 2 (Galbeta 1-3/1-4GlcNAc-R) oligosaccharides with similar efficiencies, and on core 1 glycans. This enzyme has been implicated in tumor metastasis processes.	Restricted expression toward colon (RPKM 8.9)	-	-
	TNFAIP3 (OMIM: 191163)	TNF alpha induced protein 3	Znc finger protein and ubiqitin-editing enzyme, has been shown to inhibit NF-kappa B activation as well as TNF-mediated apoptosis. The encoded protein, which has both ubiquitin ligase and deubiquitinase activities, is involved in the cytokine-mediated immune and inflammatory responses.	Broad expression in bone marrow (RPKM 156.2), appendix (RPKM 61.9)	Autoinflammatory syndrome, familial, Behcet-like; association with other auto-immune conditions, like rheumatoid arthritis and lupus erythematosus.	-
	LACTB2 (OMIM: No entry)	Lactamase beta 2	Endoribonuclease; cleaves preferentially 3 to purine-pyrimidine dinucleotide motifs in single-stranded RNA. The cleavage product contains a free 3 -OH group. Has no activity with double-stranded RNA or DNA. Required for normal mitochondrial function and cell viability.	Broad expression in testis (RPKM 25.7), kidney (RPKM 23.4) and 24 other tissues (brain: RPKM 3.728)	-	-

Legend: (a) data retrieved from NCBI, SwissProt and GeneCards databases;(b) expression data retrieved from BioProject: PRJEB4337 (Fagerberg et al. 2014). (c) data retrieved from Decipher (Firth et al. 2009), PubMed and HGMD.

Table 3.5 – Segregation analyses of aCGH findings and selected WES variants performed by Sanger sequencing in some members of the family

Genomic analysis		Family members																							
	Variant detected	Epilepsy					GEE/epileptiform activity/febrile seizures/photoparoxysmal response					Non-affected													
		II2	II9	II14	III1	III2	I2	I3	I5	II11	III8	I1	II7	II13	II16	II18	II19	III3	III7	III9	III10	III12	III13	III14
aCGH	3q27.2dup	■	■	■	■	■			X	X		■	X	X	X	X	X	X	X	X	X	X	X	X
	3q29dup								X	X	■		X	X	X	X	X	X	X	X	X	X	X	X
	15q11.2del	■		■	■	■			X	X		■	X	X	X	X	X	X	X	X	X	X	X	X
	15q11.2dup		■				■	■	X	X	■		X	X	X	X	X	X	X	X	X	X	X	X
	15q13.2del/15q13.3del		■		■	■		■	X	X		■	X	X	X	X	X	X	X	X	X	X	X	X
	16p13.11del			■					X	X			X	X	X	X	X	X	X	X	X	X	X	X
WES/Sanger	*FERMT2*: c.1538C>T p.(Thr513Met)	■	■	■	■	■	■		■		■		■											
	RCN2: c.169C>A p.(Leu56Ile)	■		■	■	■						■												
	BAIAP3: c.1886C>T p.(Thr629Ile)	■		■	■	■						■		■			■	■			■			
	MAPK8IP3: c.3815C>T p.(Ser1272Leu)	■		■	■	■						■		■				■			■			
	DISC1: c.1358C>T p.(Thr453Met)	■		■	■	■	■	■		■							■					■		
	PARD3: c.2810C>T p.(Ala937Val)	■	■	■	■	■						■	■					■	■			■		
	CELSR1: c.1036A>T p.(Thr346Ser)	■		■	■	■	■		■	■							■	■						

■	VARIANT PRESENT
X	aCGH NOT PERFORMED

Familial segregation analysis:

Familial segregation analysis was performed for class I genes and the results are presented in Table 3.5, together with the results of aCGH.

Affected individuals (II2, II9, II14, III1, III2):

In this group we could observe that: 1) the duplication in 3q27.2, the *FERMT2* and the *PARD3* heterozygous variants are present in all the individuals; 2) patient II9 does not carry the studied variants in the *RCN2*, *DISC1*, *BAIAP3*, *MAPK8IP3* and *CELSR1* genes; 3) CNVs other than the 3q27.2dup are randomly distributed in the affected group.

Individuals without epilepsy, but with GEE (I2, I3, I5, II11, III8):

In this group we could observe that: 1) individuals I2, I3 and III8 do not have the duplication in 3q27.2, CNVs other than the 3q27.2dup being randomly distributed among the individuals; 2) none has the variants in *RCN2*, *BAIAP3*, *PARD3* and *MAPK8IP3* genes; 3) the *FERMT2* variant is present in three individuals (I2, I5, III8); 4)the *DISC1* variant is present in three individuals (I2, I3, II11); 5) the *CELSR1* variant is present in three individuals (I2, I5, II11).

Non-affected (I1, II7, II13, II16, II18, II19, III3, III7, III9, III10, III12, III13, III14):

In this group we could observe that: 1) the *FERMT2* variant is only present in one individual (II7) (who also has the variant in *PARD3*, but not the 3q27.2dup); 2) the *RCN2* variant is only present in one individual (I1); 3) the *DISC1* and *CELSR1* variants are present in two individuals (II19, III12 and II19, III3, respectively); 4) the variants in *BAIAP3*, *MAPK8IP3* and *PARD3* genes are randomly distributed.

Discussion

Our genome wide exon sequencing and CNV analysis did not reveal an established or well-known genetic cause of epilepsy as the single cause of the phenotype segregating in this family. Importantly, no single nucleotide variants, deletions or duplications were identified in the 343 genes previously implied as causal for epilepsy (Helbig, Scheffer, et al. 2008; Helbig, Riggs, et al. 2018; Helbig, Heinzen, et al. 2018; Heyne, Singh, Stamberger, Abou Jamra, Caglayan, Lemke, et al. 2018). Therefore, we looked for novel causal variants in the rest of the exome as well as for dosage changes in susceptibility *loci*.

Considering the CNVs detected in this family, some of them lie within regions harbouring recurrent CNVs that have been uncovered as being risk factors for several NDs [reviewed in (Torres, Barbosa, and Maciel 2016)], such is the case of the CNVs in 3q29, 15q11.2, 15q13.3 and 16p13.1 cytobands, albeit not always overlapping the genes considered to be key within each region.

Recurrent 3q29 rearrangements are particularly rare, have approximately 1.5Mb and usually encompass 22 genes; however, the one detected in individual III8 is considerably smaller and only encompasses 2 genes of the critical region*, PIGZ* and *MFI2*, neither of which has been proposed to be the key gene for the clinical manifestations of 3q29 CNVs, such as ID, DD, BD, learning disability and SZ (Mulle et al. 2010; Levinson et al. 2011). Moreover, this CNV was detected only in patient III8 and, therefore, its contribution to phenotype segregation in this family can be ruled out.

The 15q11.2 microdeletions (15q11.2del) range from 253 Kb to 1.5 Mb and have been associated with ID, SZ, DD and ASD, as well as with epilepsy (De Kovel et al. 2010). This CNV encompasses four non-imprinted genes, *NIPA1*, *NIPA2*, *CYFIP1* and *TUBGCP5* (Burnside et al. 2011), neither of which is located in the 15q11.2 CNVs detected in this family; in fact, the CNVs that were detected include the pseudogene *NBEAP1* (*BCL8*) and several members of the Golgin family of proteins, which are localized to the Golgi apparatus, but not known to be associated with neurological phenotypes. Being so, it is not clear whether this microdeletion contributes at all to this family's phenotype.

Deletion at 15q13 was first described in patients with ID, epilepsy and variable facial and digital dysmorphisms (Sharp et al. 2008), and later on recognized as a major risk factor for IGE (Helbig et al. 2009). Its critical region encompasses at least seven genes, of which *CHRNA7* was soon considered the main candidate for some of the clinical findings associated with these CNVs, such as seizures (Hoppman-Chaney et al. 2013). The *CHRNA7* gene is not included in the CNV present in these family members; however, it includes the *CHRFAM7A* gene, which results from a partial duplication of *CHRNA7'* and is considered to be a negative antisense regulator of the *CHRNA7* expression. Being so, the number of copies and mutation status of *CHRFAM7A* can regulate *CHRNA7* function and so the ratio of the parent to the duplicated gene is important for evaluating the overall function of the α7 neuronal nicotinic acetylcholine receptor in human patients

(Araud et al. 2011; De Lucas-Cerrillo et al. 2011). Nevertheless, this CNV could not explain the disease in this family by itself since it is only present in 3/5 of the individuals with epilepsy.

As for the 16p13.11 deletion, it is significantly smaller than the rearrangements usually described in this cytoband, and contains only two genes, the *NPIPA1* and *PDXDC1*. Despite being a rearrangement significantly associated with IGE/GGE (De Kovel et al. 2010), a major contribution to the disease segregation in this family can be ruled out, since it was only observed in one individual (II14).

Finally, a duplication in 3q27.2 cytoband was also detected in several individuals of this family. It affects the *C3orf70* gene, which, according to the Bgee database (Bastian et al. 2008), has a broad expression in brain (RPKM 8.0), and particularly in the forebrain. Nevertheless, its function is completely unknown and so is its contribution to this family´s phenotype.

After prioritization of the variants identified by exome sequencing, according to: (I) the biological function according to information available in the literature; (II) brain expression (preference was given to genes expressed in the brain); (III) the existence of animal models (especially those predicting an epileptic phenotype); (IV) bioinformatics predictions of the effect of the variant, and (V) a positive-match in GeneMatcher software, the variants present in all three family members were grouped in 3 classes, from the strongest (class I) to the less strong (class III) candidates, for which a brief description is presented in Table 3.4. Of these, class I genes included *FERMT2*, *RCN2*, *BAIAP3, MAPK8IP3*, *DISC1, PARD3* and *CELSR1;* class II included the *SPHKAP, ST8SIA1*, *BUD23* (*WBSCR22*), *NKAIN1*, *FRYL*, *MIS18BP1*, and *WDR90* genes; and class III included *ABCA4*, *CATSPER1*, *TEKT5*, *WFIKKN2*, *APOB*, *GAL3ST2*, *TNFAIP3*, and *LACTB2*.

Given the results of the segregation analysis performed for selected gene variants (presented in Table 3.5), the one that better fits the segregation pattern of the disease in this family is the NM_006832.2: c.1538C>T p.(Thr513Met) variant detected in the *FERMT2* gene: it is present in all the affected individuals (i.e. those diagnosed with epilepsy); it is also present in three (I2, I5, III8) of five individuals with GEE/epileptiform activity/febrile seizures/photoparoxysmal response; moreover, it is not present in the non-affected individuals, with the exception of the obligate carrier II7, who is the son of patient I2 and father of patient III8. The presence of the variant in a non-affected individual could be explained by an

incomplete penetrance and/or the presence of additional factors that could contribute to modify the phenotype. In fact, all the individuals with epilepsy, besides having the heterozygous variant in the *FERMT2* also carry the duplication in 3q27.2 and the heterozygous c.2810C>T p.(Ala937Val) variant in the *PARD3* gene.

Consistent with the two latter variants not being causal *per se* is the fact that individual I1 (non-affected) is a carrier of the 3q27.2 duplication, but does not have the variants in the *FERMT2* and *PARD3* genes. Additionally, in the non-affected group a random pattern is observed for the variants detected in the *PARD3*, *BAIAP3* and *MAPK8IP3* genes, while the *DISC1* and *CELSR1* variants are present in two individuals. As for the variant detected in the *PARD3* gene, although present in all the individuals with epilepsy, it is also randomly present in several unaffected individuals, making its contribution to the family phenotype difficult to ascertain.

Considering the individuals who do not have epilepsy but present GEE/epileptiform activity/febrile seizures/photoparoxysmal response, none have the 3q27.2 duplication or the *PARD3* variant and, additionally, individuals I3 and II11 do not have the *FERMT2* variant. As for I3 patient, it could be speculated that other variants, such as the deletion in 15q13.2/15q13.3, could contribute to the epileptiform activity observed. However, for individual II11, this speculation cannot be made, as the only variants observed were those in *DISC1* and *CELSR1* genes, that appears to be randomly present in both affected and non-affected individuals; moreover, this individual was not initially studied by aCGH, being, therefore, necessary to assess the presence of any of the CNVs detected in the family that could explain the GEE activity observed (ongoing work).

The fermitin family homologue 2 (*FERMT2*, Kindlin 2) gene encodes for a protein of the Fermitin family, also called Kindlin family. Kindlins are a subclass of FERM-containing proteins that comprises structurally similar and evolutionarily conserved scaffold proteins that enhance integrin activation and are involved in the regulation of cell–matrix adhesion (Montanez et al. 2008). Although encoded by different genes, the three mammalian kindlins — kindlin-1 (also known as kindlerin and FERMT1), kindlin-2 (also known as MIG-2) and kindlin-3 (also known as URP2) — exhibit identical domain architecture and high sequence similarities (Larjava, Plow, and Wu 2008). FERMT2 or kindlin 2 is important in enhancing Wnt/β-catenin

signaling by selectively binding to the active β-catenin. On one hand, kindlin 2 stabilizes β-catenin by preventing glycogen synthase kinase 3 beta (GSK3β) binding to the active β-catenin. On the other hand, kindlin 2 promotes the transcription of Wnt target gene Axin2, mainly through the formation of a tripartite complex with β-catenin/TCF4 (Yu et al. 2012).

The malfunctioning of kindlins affects integrin signalling, cell–extracellular matrix (ECM) adhesion and cytoskeletal organization, playing an important role in the pathogenesis of certain diseases. Mutations in the kindlin 1 gene cause Kindler syndrome (KS), in which mainly skin and intestine are affected, whereas mutations in the kindlin 3 gene cause leukocyte adhesion deficiency type III (LAD III), which is characterized by impaired extravasation of blood effector cells and severe, spontaneous bleedings (Rognoni, Ruppert, and Fässler 2016). Also, aberrant expression of kindlins has been reported in various forms of cancer and in tissue fibrosis (Sin et al. 2011; Kato et al. 2004). The missense variant here identified occurs in the FERM domain of the protein, in band 4.1; FERM domains are compact clover-shaped structures typically involved in linking intracellular proteins to the cytoplasmic tails of transmembrane proteins (Meurice et al. 2010). Moreover, it affects a moderately conserved nucleotide as well as a moderately conserved amino acid, and according to the bioinformatic predictions it is most probably deleterious.

In mice, the loss of kindlin 2 results in early embryonic lethality (Montanez et al. 2008). More recently, Keck and colleagues have studied the proteomic profiling of epileptogenesis in a rat model: the study focused in the proteins functionally associated with cell stress, cell death, ECM remodeling, cell-ECM interaction, cell-cell interaction, angiogenesis, and blood-brain barrier function. The analysis revealed expression alterations of proteins which had not been discussed yet in the context of epileptogenesis; such was the case of FERMT2, which had a strong over-expression, confirmed by immunohistochemistry, in the rat hippocampus ten days following *status epilepticus* (Keck et al. 2018). Given these results, it would be particularly interesting to study a conditional knockout mouse model (Gondo 2008), by silencing the *FERMT2* gene in the brain and/or in specific regions of the brain, in an attempt to elucidate its contribution to disease.

Finally, a *de novo* likely pathogenic missense variant in the *FERMT2* gene was identified in a child with therapy refractory focal epilepsy by trio exome sequencing

(personal communication of Matias Wagner, Institute of Human Genetics, Munich, via GeneMatcher), adding further strength to a possible role of the *FERMT2* gene in epilepsy. This role, however, needs further confirmation, not only with functional studies, by means of patients' pluripotent stem cells and/or animal models, but also with the description of additional patients in order to confirm the *FERMT2* gene as a new epilepsy-causing gene.

In summary, our genomic analysis of this large family pinpoints the *FERMT2* gene as the most likely candidate gene to explain the disease in this family. Nevertheless, future work is needed to assess the functional impact of the *FERMT2* variants and ascertain the role of *FERMT2* in the context of epileptogenesis.

Supplementary data

Table S1 – Sequences of the primers used for Sanger validation and segregation analysis

Chr	RefSeq	Gene	Primer F	Location F	Primer R	Location R	Amplicon size (bp)
16	NM_003933.4	*BAIAP3*	CCAGCCTCTTCCACAGGT	ex19/int19	CACGTTCCAGGCTCACAG	int 21	514
15	NM_002902.2	*RCN2*	CGAAACCCTTTGGCAAGG	int1	CATTCAGCAGACACCGCTG	int2	282
10	NM_019619.3	*PARD3*	CTTGTAAACCTCCCAGGTTCTC	int18/ex19	CTCCCTGGAGCTCACTTTAAG	int19	314
1	NM_018662.2	*DISC1*	CTCCTAAGTATGAGAACAGATGGG	int4	GGGTTCACACCTGTAGATCCA	int5	429
16	NM_015133.4	*MAPK8IP3*	GATGCCGTGAAGTTCTTTGT	ex30/int30	CAGGTTGAGAGGTTAGAAGAGG	3' UTR	592
14	NM_006832.2	*FERMT2*	CTTAAATGAGTTGACCAGTCATCC	int11	GGATTTCAGAGGTTCAGATTAGG	int12	478
22	NM_014246.2	*CELSR1*	GCTATTACATGGAGGGGCTG	ex1	GTAGTTCTGCTCGCTGAACTG	ex1	521

Legend: Chr – chromosome; RefSeq – reference sequence; F – forward; R – reverse; bp – base pairs; ex – exon; int – intron; the following M13 tails were attached to: a) F-primers: TGTAAAACGACGGCCAGT; b) R-primers: CAGGAAACAGCTATGACC.

List of the genes usually included in epilepsy NGS based panels:

ABAT, ABCB1, ABCC8, ACY1, ADCK3, ADGRG1, ADGRV1, ADSL, AGA, AHI1, ALDH4A1, ALDH5A1, ALDH7A1, ALG1, ALG12, ALG2, ALG3, ALG6, ALG8, ALG9, AMT, APTX, ARFGEF2, ARG1, ARHGEF9, ARL13B, ARSA, ARSB, ARX, ASPA, ASPM, ATIC, ATP1A2, ATP2A2, ATP6AP2, ATP6V0A2, ATPAF2, ATR, ATRX, B4GALT1, BCS1L, BOLA3, BRAF, BTD, BUB1B, C12orf65, CACNA1A, CACNA1H, CACNB4, CASK, CASR, CBL, CC2D2A, CCL2, CDK5RAP2, CDKL5, CDON, CELSR1, CENPJ, CEP152, CEP290, CHRNA2, CHRNA4, CHRNB2, CLCN2, CLCNKA, CLCNKB, CLN3, CLN5, CLN6, CLN8, CNTNAP2, COG1, COG7, COG8, COL18A1, COL4A1, COQ2, COQ9, COX15, CPT2, CSTB, CTSA, CTSD, CUL4B, DCX, DLD, DOLK, DPAGT1, DPM1, DPM3, DPYD, EFHC1, EFHC2, EIF2B1, EIF2B2, EIF2B3, EIF2B4, EIF2B5, EMX2, EPM2A, ETFA, ETFB, ETFDH, FGD1, FGF8, FGFR3, FH, FKRP, FKTN, FLNA, FLVCR2, FOLR1, FOXG1, FUCA1, GABRA1, GABRB3, GABRD, GABRG2, GALC, GAMT, GATM, GCDH, GCSH, GFAP, GLB1, GLDC, GLI2, GLI3, GLRA1, GLRB, GNE, GNPTAB, GNPTG, GNS, GPC3, GPHN, GRIA3, GRIN1, GRIN2A, GRIN2B, GUSB, HCN1, HCN4, HEXA, HEXB, HGSNAT, HPD, HRAS, HSD17B10, IDS, IDUA, KAT6B, KCNA1, KCNJ1, KCNJ10, KCNJ11, KCNMA1, KCNQ2, KCNQ3, KCNV2, KCTD7, KDM5C, KIF1BP, KMT2D, KRAS, L2HGDH, LAMA2, LARGE, LBR, LGI1, LIG4, LRPPRC, MAGI2, MAP2K1, MAP2K2, MAPK10, MBD5, MCOLN1, MCPH1, MECP2, MED17, MEF2C, MFSD8, MGAT2, MLC1, MOCS1, MOCS2, MOGS, MPDU1, MPI, MTHFR, NAGLU, NDE1, NDUFA1, NDUFA2, NDUFS1, NDUFS3, NDUFS4, NDUFS7, NDUFS8, NDUFV1, NEU1, NF1, NHEJ1, NHLRC1, NIPBL, NODAL, NOTCH3, NPC1, NPC2, NPHP1, NRAS, NRXN1, OFD1, OPHN1, PAFAH1B1, PAK3, PANK2, PAX6, PC, PCDH19, PCNT, PDHA1, PDSS1, PDSS2, PEX1, PEX12, PEX14, PEX2, PEX26, PEX3, PEX5, PEX6, PEX7, PGK1, PHF6, PIGV, PLA2G6, PLCB1, PLP1, PMM2, PNKP, PNPO, POLG, POMGNT1, POMT1, POMT2, PPT1, PQBP1, PRICKLE1, PRICKLE2, PRODH, PRRT2, PSAP, PTCH1, PTPN11, QDPR, RAB39B, RAB3GAP1, RAF1, RAI1, RARS2, RELN, RFT1, RNASEH2A, RNASEH2B, RNASEH2C, RPGRIP1L, SAMHD1, SCARB2, SCN10A, SCN1A, SCN1B, SCN2A, SCN2B, SCN3A, SCN3B, SCN4A, SCN4B, SCN5A, SCN8A, SCN9A, SCO2, SDHA, SERPINI1, SETBP1, SGSH, SHH, SHOC2, SIX3, SLC17A5, SLC25A15, SLC25A19, SLC25A22, SLC2A1, SLC35A1, SLC35C1, SLC46A1, SLC6A5, SLC9A6, SMC1A, SMC3, SMPD1, SMS, SNAP29, SOS1, SPRED1, SPTAN1, SRPX2, STIL, STXBP1, SUMF1, SUOX, SYN1, SYNGAP1, SYP, TACO1, TBC1D24, TBX1, TCF4, TGIF1, TMEM216, TMEM67, TMEM70, TPP1, TREX1, TSC1, TSC2, TSEN2, TSEN34, TSEN54, TUBA1A, TUBA8, TUBB2B, UBE3A, VANGL1, VPS13A, VPS13B, VRK1, WDR62, ZEB2, ZIC2.

Bibliography

Allache, Redouane, Patrizia De Marco, Elisa Merello, Valeria Capra, and Zoha Kibar. 2012. “Role of the Planar Cell Polarity Gene CELSR1 in Neural Tube Defects and Caudal Agenesis.” *Birth Defects Research Part A - Clinical and Molecular Teratology* 94 (3):176–81. https://doi.org/10.1002/bdra.23002.

Arányi, Tamás, András Váradi, István Simon, and Gábor E. Tusnády. 2006. “The BiSearch Web Server.” *BMC Bioinformatics* 7:1–7. https://doi.org/10.1186/1471-2105-7-431.

Araud, Tanguy, Sharon Graw, Ralph Berger, Michael Lee, Estelle Neveu, Daniel Bertrand, and Sherry Leonard. 2011. “The Chimeric Gene CHRFAM7A, a Partial Duplication of the CHRNA7 Gene, Is a Dominant Negative Regulator of A7*nAChR Function.” *Biochemical Pharmacology* 82 (8):904–14. https://doi.org/10.1016/j.bcp.2011.06.018.

Bastian, Frederic, Gilles Parmentier, Julien Roux, Sebastien Moretti, Vincent Laudet, and Marc Robinson-Rechavi. 2008. “Bgee: Integrating and Comparing Heterogeneous Transcriptome Data among Species.” *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5109 LNBI:124–31. https://doi.org/10.1007/978-3-540-69828-9_12.

Berger, Seth I., Carla Ciccone, Karen L. Simon, May Christine Malicdan, Thierry Vilboux, Charles Billington, Roxanne Fischer, et al. 2017. “Exome Analysis of Smith–Magenis-like Syndrome Cohort Identifies de Novo Likely Pathogenic Variants.” *Human Genetics* 136 (4):409–20. https://doi.org/10.1007/s00439-017-1767-x.

Bjartmar, L., A. D. Huberman, E. M. Ullian, R. C. Renteria, X. Liu, W. Xu, J. Prezioso, et al. 2006. “Neuronal Pentraxins Mediate Synaptic Refinement in the Developing Visual System.” *Journal of Neuroscience* 26 (23):6269–81. https://doi.org/10.1523/JNEUROSCI.4212-05.2006.

Boucherie, C., C. Boutin, Y. Jossin, O. Schakman, A. M. Goffinet, L. Ris, P. Gailly, and F. Tissir. 2018. “Neural Progenitor Fate Decision Defects, Cortical Hypoplasia and Behavioral Impairment in Celsr1-Deficient Mice.” *Molecular Psychiatry* 23 (3):723–34. https://doi.org/10.1038/mp.2017.236.

Buffart, Tineke E., Daniëlle Israeli, Marianne Tijssen, Sjoerd J. Vosse, Alan Mrsić, Gerrit A. Meijer, and Bauke Ylstra. 2008. “Across Array Comparative Genomic Hybridization: A Strategy to Reduce Reference Channel Hybridizations.” *Genes, Chromosomes & Cancer* 47 (11):994–1004. https://doi.org/10.1002/gcc.20605.

Burnside, Rachel D, Romela Pasion, Fady M Mikhail, Andrew J Carroll, Nathaniel H Robin, Erin L Youngs, Inder K Gadi, et al. 2011. “Microdeletion/Microduplication of Proximal 15q11.2 between BP1 and BP2: A Susceptibility Region for Neurological Dysfunction Including Developmental and Language Delay.” *Human Geneticsenetics*

130 (4):517–28. https://doi.org/10.1007/s00439-011-0970-4.

Chen, Xiaoli, Yu An, Yonghui Gao, Liu Guo, Lei Rui, Hua Xie, Mei Sun, et al. 2017. “Rare Deleterious PARD3 Variants in the APKC-Binding Region Are Implicated in the Pathogenesis of Human Cranial Neural Tube Defects via Disrupting Apical Tight Junction Formation.” *Human Mutation* 38 (4):378–89. https://doi.org/10.1002/humu.23153.Rare.

Chen, Xiaoli, Yiping Shen, Yonghui Gao, Huizhi Zhao, Xiaoming Sheng, Jizhen Zou, Va Lip, et al. 2013. “Detection of Copy Number Variants Reveals Association of Cilia Genes with Neural Tube Defects.” *PLoS ONE* 8 (1):1–13. https://doi.org/10.1371/journal.pone.0054492.

Chen, Zhongzhong, Yunping Lei, Xuanye Cao, Yufang Zheng, Fang Wang, Yihua Bao, Rui Peng, Richard H. Finnell, Ting Zhang, and Hongyan Wang. 2018. “Genetic Analysis of Wnt/PCP Genes in Neural Tube Defects.” *BMC Medical Genomics* 11 (1). BMC Medical Genomics:1–9. https://doi.org/10.1186/s12920-018-0355-9.

Choudhary, Bikash, Madhushree Kamak, Neena Ratnakaran, Jitendra Kumar, Anjali Awasthi, Chun Li, Ken Nguyen, Kunihiro Matsumoto, Naoki Hisamoto, and Sandhya P Koushika. 2017. “UNC-16 / JIP3 Regulates Early Events in Synaptic Vesicle Protein Trafficking via LRK-1 / LRRK2 and AP Complexes,” 1–25. https://doi.org/10.1371/journal.pgen.1007100.

Curtin, John A, E Quint, V Tsipouri, AJ Arkell, DJ Henderson, N Spurr, P Stanier, et al. 2003. “Mutation of Celsr1 Disrupts Planar Polarity of Inner Ear Hair Cells and Causes Severe Neural Tube Defects in the Mouse.” *Current Biology* 13 (6):1129–33. https://doi.org/10.1016/S.

DePristo, M. a., E. Banks, R. E. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, a. a. Philippakis, et al. 2011. “A Framework for Variation Discovery and Genotyping Using Next- Generation DNA Sequencing Data.” *Nat Genet* 43 (5):491–98. https://doi.org/10.1038/ng.806.A.

Devinsky, Orrin, Annamaria Vezzani, Terence J O, Nathalie Jette, Ingrid E Scheffer, Marco de Curtis, and Piero Perucca. 2018. “Epilepsy.” *Nature Publishing Group* 4 (May). https://doi.org/10.1038/nrdp.2018.24.

Doll, A., and K.-H. Grzeschik. 2001. “Characterization of Two Novel Genes, WBSCR20 and WBSCR22, Deleted in Williams-Beuren Syndrome.” *Cytogenetic and Genome Research* 95 (1–2):20–27. https://doi.org/10.1159/000057012.

Egawa, Jun, Yuichiro Watanabe, Atsunori Sugimoto, Ayako Nunokawa, Masako Shibuya, Hirofumi Igeta, Emiko Inoue, et al. 2015. “Whole-Exome Sequencing in a Family with a Monozygotic Twin Pair Concordant for Autism Spectrum Disorder and a Follow-up Study.” *Psychiatry Research* 229 (1–2). Elsevier:599–601.

https://doi.org/10.1016/j.psychres.2015.07.018.

Fagerberg, Linn, Bjorn M Hallstromm, Per Oksvold, Caroline Kampf, Dijana Djureinovic, Jacob Odeberg, Masato Habuka, et al. 2014. “Synthesis and Coordination Chemistry of Thiophenol-Based Heterodonor Ligands Containing P,S, as,S and P,SAs Donor Atoms.” *Molecular & Cellular Proteomics* 13:397–406. https://doi.org/10.1074/mcp.M113.035600.

Firth, Helen V., Shola M. Richards, A. Paul Bevan, Stephen Clayton, Manuel Corpas, Diana Rajan, Steven Van Vooren, Yves Moreau, Roger M. Pettett, and Nigel P. Carter. 2009. “DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources.” *American Journal of Human Genetics* 84 (4). The American Society of Human Genetics:524–33. https://doi.org/10.1016/j.ajhg.2009.03.010.

Gao, Yonghui, Jianhua Wang, Shaofang Shangguan, Yihua Bao, Xiaoli Lu, Jizhen Zou, Yaohua Dai, Junling Liu, and Ting Zhang. 2017. “Quantitative Measurement of PARD3 Copy Number Variations in Human Neural Tube Defects.” *Cellular and Molecular Neurobiology*, no. 2. Springer US:1–10. https://doi.org/10.1007/s10571-017-0506-0.

Goffinet, Andre M., and Fadel Tissir. 2017. “Seven Pass Cadherins CELSR1-3.” *Seminars in Cell and Developmental Biology* 69. Elsevier Ltd:102–10. https://doi.org/10.1016/j.semcdb.2017.07.014.

Gondo, Yoichi. 2008. “Trends in Large-Scale Mouse Mutagenesis: From Genetics to Functional Genomics.” *Nature Reviews Genetics* 9 (10):803–10. https://doi.org/10.1038/nrg2431.

Gorokhova, Svetlana, Stéphanie Bibert, Käthi Geering, and Nathaniel Heintz. 2007. “A Novel Family of Transmembrane Proteins Interacting with β Subunits of the Na,K-ATPase.” *Human Molecular Genetics* 16 (20):2394–2410. https://doi.org/10.1093/hmg/ddm167.

Hamel, Virginie, Emmanuelle Steib, Romain Hamelin, Florence Armand, Susanne Borgers, Isabelle Flückiger, Coralie Busso, et al. 2017. “Identification of Chlamydomonas Central Core Centriolar Proteins Reveals a Role for Human WDR90 in Ciliogenesis.” *Current Biology* 27 (16):2486–2498.e6. https://doi.org/10.1016/j.cub.2017.07.011.

Helbig, Ingo, Erin L. Heinzen, Heather C. Mefford, Samuel F. Berkovic, Daniel H. Lowenstein, Mitsuhiro Kato, J. Helen Cross, et al. 2018. “Genetic Literacy Series: Primer Part 2—Paradigm Shifts in Epilepsy Genetics.” *Epilepsia* 59 (6):1138–47. https://doi.org/10.1111/epi.14193.

Helbig, Ingo, Nicholas A. Matigian, Lata Vadlamudi, Kate M. Lawrence, Marta A. Bayly,

Sharon M. Bain, Dileepa Diyagama, et al. 2008. “Gene Expression Analysis in Absence Epilepsy Using a Monozygotic Twin Design.” *Epilepsia* 49 (9):1546–54. https://doi.org/10.1111/j.1528-1167.2008.01630.x.

Helbig, Ingo, Heather C Mefford, Andrew J Sharp, Michel Guipponi, Marco Fichera, Andre Franke, Hiltrud Muhle, et al. 2009. “15q13.3 Microdeletions Increase Risk of Idiopathic Generalized Epilepsy.” *Nat Genet* 41 (2):160–62. https://doi.org/10.1038/ng.292.15q13.3.

Helbig, Ingo, Erin Rooney Riggs, Carrie-Anne Barry, Karl Martin Klein, David Dyment, Courtney Thaxton, Bekim Sadikovic, et al. 2018. “The ClinGen Epilepsy Gene Curation Expert Panel-Bridging the Divide between Clinical Domain Knowledge and Formal Gene Curation Criteria.” *Human Mutation* 39 (11):1476–84. https://doi.org/10.1002/humu.23632.

Helbig, Ingo, Ingrid E. Scheffer, John C. Mulley, and Samuel F. Berkovic. 2008. “Navigating the Channels and beyond: Unravelling the Genetics of the Epilepsies.” *The Lancet Neurology* 7 (3):231–45. https://doi.org/10.1016/S1474-4422(08)70039-5.

Heyne, Henrike O., Tarjinder Singh, Hannah Stamberger, Rami Abou Jamra, Hande Caglayan, Dana Craiu, Peter De Jonghe, et al. 2018. “De Novo Variants in Neurodevelopmental Disorders with Epilepsy.” *Nature Genetics*, 1. https://doi.org/10.1038/s41588-018-0143-7.

———. 2018. “De Novo Variants in Neurodevelopmental Disorders with Epilepsy.” *Nature Genetics* 50 (7):1048–53. https://doi.org/10.1038/s41588-018-0143-7.

Hildebrand, Michael S, Hans-Henrik M Dahl, John Anthony Damiano, Richard J H Smith, Ingrid E Scheffer, and Samuel F Berkovic. 2013. “Recent Advances in the Molecular Genetics of Epilepsy.” *Journal of Medical Genetics* 50 (5):271–79. https://doi.org/10.1136/jmedgenet-2012-101448.

Hoppman-Chaney, N, K Wain, P R Seger, D W Superneau, and J C Hodge. 2013. “Identification of Single Gene Deletions at 15q13.3: Further Evidence That CHRNA7 Causes the 15q13.3 Microdeletion Syndrome Phenotype.” *Clinical Genetics* 83 (4):345–51. https://doi.org/10.1111/j.1399-0004.2012.01925.x.

Ikeda, Masashi, Atsushi Takahashi, Yoichiro Kamatani, Yukihide Momozawa, Takeo Saito, Kenji Kondo, Ayu Shimasaki, et al. 2018. “Genome-Wide Association Study Detected Novel Susceptibility Genes for Schizophrenia and Shared Trans-Populations/Diseases Genetic Effect.” *Schizophrenia Bulletin*. https://doi.org/10.1093/schbul/sby140.

Iossifov, Ivan, Brian J. O’Roak, Stephan J. Sanders, Michael Ronemus, Niklas Krumm, Dan Levy, Holly a. Stessman, et al. 2014. “The Contribution of de Novo Coding Mutations to Autism Spectrum Disorder.” *Nature*, October. Nature Publishing Group.

https://doi.org/10.1038/nature13908.

Johansson, Ann-sofi, Mariette Driessens, and Pontus Aspenström. 2000. “The Mammalian Homologue of the Caenorhabditis Elegans Polarity Protein PAR-6 Is a Binding Partner for the Rho GTPases Cdc42 and Rac1.” *Journal of Cell Science* 113:3267–75. http://www.ncbi.nlm.nih.gov/pubmed/10954424.

Kato, Kiyoshi, Tanri Shiozawa, Junji Mitsushita, Ayaka Toda, Akiko Horiuchi, Toshio Nikaido, Shingo Fujii, and Ikuo Konishi. 2004. “Expression of the Mitogen-Inducible Gene-2 (Mig-2) Is Elevated in Human Uterine Leiomyomas but Not in Leiomyosarcomas.” *Human Pathology* 35 (1):55–60. https://doi.org/10.1016/j.humpath.2003.08.019.

Keck, Michael, Roelof Maarten van Dijk, Cornelia A. Deeg, Katharina Kistler, Andreas Walker, Eva Lotta von Rüden, Vera Russmann, Stefanie M. Hauck, and Heidrun Potschka. 2018. “Proteomic Profiling of Epileptogenesis in a Rat Model: Focus on Cell Stress, Extracellular Matrix and Angiogenesis.” *Neurobiology of Disease* 112 (August 2017):119–35. https://doi.org/10.1016/j.nbd.2018.01.013.

Khazaei, Mohammad R., and Andreas W. Püschel. 2009. “Phosphorylation of the Par Polarity Complex Protein Par3 at Serine 962 Is Mediated by Aurora A and Regulates Its Function in Neuronal Polarity.” *Journal of Biological Chemistry* 284 (48):33571–79. https://doi.org/10.1074/jbc.M109.055897.

Kim, Ju Young, Xin Duan, Cindy Y. Liu, Mi Hyeon Jang, Junjie U. Guo, Nattapol Pow-anpongkul, Eunchai Kang, Hongjun Song, and Guo li Ming. 2009. “DISC1 Regulates New Neuron Development in the Adult Brain via Modulation of AKT-MTOR Signaling through KIAA1212.” *Neuron* 63 (6). Elsevier Ltd:761–73. https://doi.org/10.1016/j.neuron.2009.08.008.

Kim, Su Kang, Jong Yoon Lee, Hae Jeong Park, Jong Woo Kim, and Joo Ho Chung. 2012. “Association Study between Polymorphisms of the PARD3 Gene and Schizophrenia.” *Experimental and Therapeutic Medicine* 3 (5):881–85. https://doi.org/10.3892/etm.2012.496.

Kirkpatrick, Laura L., Martin M. Matzuk, D. C. Dodds, and Mark S. Perin. 2000. “Biochemical Interactions of the Neuronal Pentraxins. Neuronal Pentraxin (NP) Receptor Binds to Taipoxin and Taipoxin-Associated Calcium-Binding Protein 49 via NP1 and NP2.” *Journal of Biological Chemistry* 275 (23):17786–92. https://doi.org/10.1074/jbc.M002254200.

Kovel, Carolien G F De, Holger Trucks, Ingo Helbig, Heather C Mefford, Carl Baker, Costin Leu, Christian Kluck, et al. 2010. “Recurrent Microdeletions at 15q11.2 and 16p13.11 Predispose to Idiopathic Generalized Epilepsies.” *Brain* 133 (1):23–32. https://doi.org/10.1093/brain/awp262.

Krijgsman, Oscar, Daniëlle Israeli, Hendrik F. van Essen, Paul P. Eijk, Michel L. M. Berens, Clemens H. M. Mellink, Aggie W. Nieuwint, et al. 2013. “Detection Limits of DNA Copy Number Alterations in Heterogeneous Cell Populations.” *Cellular Oncology (Dordrecht)* 36 (1):27–36. https://doi.org/10.1007/s13402-012-0108-2.

Landrum, Melissa J., Jennifer M. Lee, Mark Benson, Garth Brown, Chen Chao, Shanmuga Chitipiralla, Baoshan Gu, et al. 2016. “ClinVar: Public Archive of Interpretations of Clinically Relevant Variants.” *Nucleic Acids Research* 44 (D1):D862–68. https://doi.org/10.1093/nar/gkv1222.

Landrum, Melissa J., Jennifer M. Lee, George R. Riley, Wonhee Jang, Wendy S. Rubinstein, Deanna M. Church, and Donna R. Maglott. 2014. “ClinVar: Public Archive of Relationships among Sequence Variation and Human Phenotype.” *Nucleic Acids Research* 42 (D1):980–85. https://doi.org/10.1093/nar/gkt1113.

Larjava, Hannu, Edward F. Plow, and Chuanyue Wu. 2008. “Kindlins: Essential Regulators of Integrin Signalling and Cell-Matrix Adhesion.” *EMBO Reports* 9 (12):1203–8. https://doi.org/10.1038/embor.2008.202.

Lei, Yunping, Huiping Zhu, Wei Yang, M. Elizabeth Ross, Gary M. Shaw, and Richard H. Finnell. 2014. “Identification of Novel CELSR1 Mutations in Spina Bifida.” *PLoS ONE* 9 (3). https://doi.org/10.1371/journal.pone.0092207.

Levinson, Douglas F., Jubao Duan, Sang Oh, Kai Wang, Alan R Sanders, Jianxin Shi, Nancy Zhang, et al. 2011. “Copy Number Variants in Schizophrenia:Confirmation of Five Previous Findings and New Evidence for 3q29 Microdeletions and VIPR2 Duplications.” *Am J Psychiatry* 168 (March):302–16.

Liu, Wenying Angela, She Chen, Zhizhong Li, Choong Heon Lee, Ghayda Mirzaa, William B. Dobyns, M. Elizabeth Ross, Jiangyang Zhang, and Song Hai Shi. 2018. “PARD3 Dysfunction in Conjunction with Dynamic HIPPO Signaling Drives Cortical Enlargement with Massive Heterotopia.” *Genes and Development* 32 (11–12):763–80. https://doi.org/10.1101/gad.313171.118.

Lucas-Cerrillo, Ana M. De, M. Constanza Maldifassi, Francisco Arnalich, Jaime Renart, Gema Atienza, Rocío Serantes, Jesús S. Cruces, Aurora Sánchez-Pacheco, Eva Andrés-Mateos, and Carmen Montiel. 2011. “Function of Partially Duplicated Human A7 Nicotinic Receptor Subunit CHRFAM7A Gene: Potential Implications for the Cholinergic Anti-Inflammatory Response.” *Journal of Biological Chemistry* 286 (1):594–606. https://doi.org/10.1074/jbc.M110.180067.

MacDonald, Jeffrey R., Robert Ziman, Ryan K C Yuen, Lars Feuk, and Stephen W. Scherer. 2014. “The Database of Genomic Variants: A Curated Collection of Structural Variation in the Human Genome.” *Nucleic Acids Research* 42 (D1):986–92. https://doi.org/10.1093/nar/gkt958.

Mariman, E. C. M., F. G. Bouwman, E. E. J. G. Aller, M. A. van Baak, and P. Wang. 2015. “Extreme Obesity Is Associated with Variation in Genes Related to the Circadian Rhythm of Food Intake and Hypothalamic Signaling.” *Physiological Genomics* 47 (6):225–31. https://doi.org/10.1152/physiolgenomics.00006.2015.

Meurice, Nathalie, Lei Wang, CA Lipinski, Zhongbo Yang, Christopher Hulme, and Joseph C Loftus. 2010. “Structural Conservation in Band 4.1, Ezrin, Radixin, Moesin (FERM) Domains as a Guide To Identify Inhibitors of the Proline- Rich Tyrosine Kinase 2.” *Curr Opin Cell Biol.* 53 (2):669–77. https://doi.org/10.1038/jid.2014.371.

Miller, Chase A., Yi Qiao, Tonya Disera, Brian D’Astous, and Gabor T. Marth. 2014. “Bam.Iobio: A Web-Based, Real-Time, Sequence Alignment File Inspector.” *Nature Methods* 11 (12):1189. https://doi.org/10.1038/nmeth.3174.

Miyoshi, Ko, Masato Asanuma, Ikuko Miyazaki, Francisco J. Diaz-Corrales, Taiichi Katayama, Masaya Tohyama, and Norio Ogawa. 2004. “DISC1 Localizes to the Centrosome by Binding to Kendrin.” *Biochemical and Biophysical Research Communications* 317 (4):1195–99. https://doi.org/10.1016/j.bbrc.2004.03.163.

Montanez, Eloi, Siegfried Ussar, Martina Schifferer, Michael Bösl, Roy Zent, Markus Moser, and Reinhard Fässler. 2008. “Kindlin-2 Controls Bidirectional Signaling of Integrins.” *Genes and Development* 22 (10):1325–30. https://doi.org/10.1101/gad.469408.

Mulle, Jennifer Gladys, Anne F. Dodd, John A. McGrath, Paula S. Wolyniec, Adele A. Mitchell, Amol C. Shetty, Nara L. Sobreira, et al. 2010. “Microdeletions of 3q29 Confer High Risk for Schizophrenia.” *American Journal of Human Genetics* 87 (2). The American Society of Human Genetics:229–36. https://doi.org/10.1016/j.ajhg.2010.07.013.

Park, Hyo Jin, Ju Hee Kim, Jin Sook Yoon, Yang Ji Choi, Yoon Hee Choi, Koung Hoon Kook, and Ji Ha Choi. 2017. “Identification and Functional Characterization of ST3GAL5 and ST8SIA1 Variants in Patients with Thyroid-Associated Ophthalmopathy.” *Yonsei Medical Journal* 58 (6):1160–69. https://doi.org/10.3349/ymj.2017.58.6.1160.

Robinson, Alexis, Sarah Escuin, Kit Doudney Vekemans, Roger E. Stevenson, Nicholas D E Greene, Andrew J. Copp, and Philip Stanier. 2012. “Mutations in the Planar Cell Polarity Genes CELSR1 and SCRIB Are Associated with the Severe Neural Tube Defect Craniorachischisis.” *Human Mutation* 33 (2):440–47. https://doi.org/10.1002/humu.21662.

Robinson, James T, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S Lander, Gad Getz, and Jill P Mesirov. 2011. “Integrative Genome Viewer.” *Nature Biotechnology* 29 (1):24–26. https://doi.org/10.1038/nbt.1754.Integrative.

Rognoni, Emanuel, Raphael Ruppert, and Reinhard Fässler. 2016. “The Kindlin Family: Functions, Signaling Properties and Implications for Human Disease.” *Journal of Cell Science* 129 (1):17–27. https://doi.org/10.1242/jcs.161190.

Scheffer, Ingrid E., Samuel Berkovic, Giuseppe Capovilla, Mary B. Connolly, Jacqueline French, Laura Guilhoto, Edouard Hirsch, et al. 2017. “ILAE Classification of the Epilepsies: Position Paper of the ILAE Commission for Classification and Terminology.” *Epilepsia* 58 (4):512–21. https://doi.org/10.1111/epi.13709.

Sharp, Andrew J, Heather C Mefford, Kelly Li, Carl Baker, Cindy Skinner, Roger E Stevenson, Richard J Schroer, et al. 2008. “A Recurrent 15q13.3 Microdeletion Syndrome Associated with Mental Retardation and Seizures.” *Nat Genet* 40 (3):322–28.

Sin, Soraya, Florian Bonin, Valérie Petit, Didier Meseure, Franois Lallemand, Ivan Biche, Akeila Bellahcne, et al. 2011. “Role of the Focal Adhesion Protein Kindlin-1 in Breast Cancer Growth and Lung Metastasis.” *Journal of the National Cancer Institute* 103 (17):1323–37. https://doi.org/10.1093/jnci/djr290.

Sirven, Joseph I. 2015. “Epilepsy: A Spectrum Disorder.” *Cold Spring Harbor Perspectives in Biology* 7 (9):1–16. https://doi.org/10.1101/cshperspect.a022848.

Sobreira, Nara, François Schiettecatte, Corinne Boehm, David Valle, and Ada Hamosh. 2015. “New Tools for Mendelian Disease Gene Identification: PhenoDB Variant Analysis Module; and GeneMatcher, a Web-Based Tool for Linking Investigators with an Interest in the Same Gene.” *Human Mutation* 36 (4):425–31. https://doi.org/10.1002/humu.22769.New.

Sobreira, Nara, François Schiettecatte, David Valle, and Ada Hamosh. 2015. “GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene.” *Human Mutation* 36 (10):928–30. https://doi.org/10.1002/humu.22844.

Spillane, J., D. M. Kullmann, and M. G. Hanna. 2016. “Genetic Neurological Channelopathies: Molecular Genetics and Clinical Phenotypes.” *Journal of Neurology, Neurosurgery and Psychiatry* 87 (1):37–48. https://doi.org/10.1136/jnnp-2015-311233.

Stenson, Peter D., Matthew Mort, Edward V. Ball, Katy Shaw, Andrew D. Phillips, and David N. Cooper. 2014. “The Human Gene Mutation Database: Building a Comprehensive Mutation Repository for Clinical and Molecular Genetics, Diagnostic Testing and Personalized Genomic Medicine.” *Human Genetics* 133 (1):1–9. https://doi.org/10.1007/s00439-013-1358-4.

Torres, Fátima, Mafalda Barbosa, and Patrícia Maciel. 2016. “Recurrent Copy Number Variations as Risk Factors for Neurodevelopmental Disorders: Critical Overview and Analysis of Clinical Implications.” *Journal of Medical Genetics* 53 (2):73–90.

https://doi.org/10.1136/jmedgenet-2015-103366.

Tusnády, G.E., István Simon, András Váradi, and Tamás Arányi. 2005. “BiSearch: Primer-Design and Search Tool for PCR on Bisulfite-Treated Genomes.” *Nucleic Acids Research* 33 (1):e9–e9. https://doi.org/10.1093/nar/gni012.

Untergasser, Andreas, Harm Nijveen, Xiangyu Rao, Ton Bisseling, René Geurts, and Jack A.M. Leunissen. 2007. “Primer3Plus, an Enhanced Web Interface to Primer3.” *Nucleic Acids Research* 35 (SUPPL.2):71–74. https://doi.org/10.1093/nar/gkm306.

Wang, Kai, Mingyao Li, and Hakon Hakonarson. 2010. “ANNOVAR: Functional Annotation of Genetic Variants from High-Throughput Sequencing Data.” *Nucleic Acids Research* 38 (16):1–7. https://doi.org/10.1093/nar/gkq603.

Wang, Z., Y. Chen, Y. Lü, X. Chen, L. Cheng, X. Mi, X. Xu, et al. 2015. “Effects of JIP3 on Epileptic Seizures: Evidence from Temporal Lobe Epilepsy Patients, Kainic-Induced Acute Seizures and Pentylenetetrazole-Induced Kindled Seizures.” *Neuroscience* 300 (May):314–24. https://doi.org/10.1016/j.neuroscience.2015.05.008.

Wojcik, Sonja M, Martesa Tantra, Beata Stepniak, Kwun-Nok M Man, Katja Müller-Ribbe, Martin Begemann, Anes Ju, et al. 2013. “Genetic Markers of a Munc13 Protein Family Member, BAIAP3, Are Gender Specifically Associated with Anxiety and Benzodiazepine Abuse in Mice and Humans.” *Molecular Medicine (Cambridge, Mass.)* 19 (1):135–48. https://doi.org/10.2119/molmed.2013.00033.

Yu, Yu, Junzhou Wu, Yunling Wang, Ting Zhao, Bo Ma, Yuqing Liu, Weigang Fang, Wei Guo Zhu, and Hongquan Zhang. 2012. “Kindlin 2 Forms a Transcriptional Complex with β-Catenin and TCF4 to Enhance Wnt Signalling.” *EMBO Reports* 13 (8):750–58. https://doi.org/10.1038/embor.2012.88.

Zhang, Xingmin, Shan Jiang, Kelly A. Mitok, Lingjun Li, Alan D. Attie, and Thomas F.J. Martin. 2017. “BAIAP3, a C2 Domain–containing Munc13 Protein, Controls the Fate of Dense-Core Vesicles in Neuroendocrine Cells.” *The Journal of Cell Biology* 216 (7):2151–66. https://doi.org/10.1083/jcb.201702099.

Zuo, Lingjun, Ke-Sheng Wang, Xiang-Yang Zhang, John H Krystal, Chiang-Shan R Li, Fengyu Zhang, Heping Zhang, and Xingguang Luo. 2013. “NKAIN1-SERINC2 Is a Functional, Replicable and Genome-Wide Significant Risk Gene Region Specific for Alcohol Dependence in Subjects of European Descent *.” *Drug Alcohol Depend* 129 (3):254–64. https://doi.org/10.1016/j.drugalcdep.2013.02.006.NKAIN1-SERINC2.

Chapter 4

Discussion

Disclaimer:

This chapter summarizes the findings, strengths and limitations of this thesis, as well as future perspectives for the development of this work.

The aCGH and MPS were the techniques used to study a group of patients with NDs: a cohort of patients with idiopathic ID and a large family with IGE, in an attempt to define, in as much cases as possible, the genetic causes of disease, as well as identify novel genetic causes of NDs.

The importance of a diagnosis

Given the burden that NDs represent not only to the individual, but also in families and their community (Maulik et al. 2011), and considering the point of view of both care providers and families, there are specific benefits from establishing an etiologic diagnosis. The provision of condition-specific family support; and the opportunity for co-management of patients in the context of a medical home to ensure the best health, social, and health care services for the patient and family (Moeschler and Shevell 2014). However, finding a diagnosis for each individual may be a considerable challenge, due to the genetic and phenotypic variability associated with these diseases and our incomplete knowledge (Wright, FitzPatrick, and Firth 2018).

In fact, considering the cohort of ID patients of unknown etiology studied by aCGH in this thesis, it was only possible to identify a definitive cause of disease in 8% of patients, i.e., only 8% had variants associated with an established pathogenic effect. Adding to this group those patients in whom a likely pathogenic variant was found, the yield obtained (13.2%) 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; Sagoo et al. 2009). Nevertheless, the most significant group was constituted by rare CNVs, many restricted to one patient/family. Most of them were classified as VOUS (~17% of the cases), given the uncertainty of their clinical significance, others were considered likely pathogenic (those comprising newly described gene rich large CNVs with good candidates for disease association or gene(s) with a compelling function). To further address the possible contribution of these variants, a search was conducted, not only in the literature but in several databases (either public or laboratory databases), in order to find similar alterations, since the identification of patients sharing variants in a given *locus* and having common phenotypic features may confer a greater degree of certainty about the pathogenic nature of the variant, and might enable the definition of the role of novel genes in development and disease. Being so, 12 new candidate *loci* for ID were put forward: 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 (Lopes F, Torres F et al. submitted for publication).

As for the epilepsy patients studied by WES, the scenario was not very different. We have identified a set of variants present in the three patients, but none occurs

in a gene with a known and unequivocally established association with epilepsy (exception made to the *CELSR1* gene, included in the epilepsy panels due to its association with neural tube defects, but not showing segregation with the disease in this family). Being so, no established or well-known genetic cause previously implied as causal for epilepsy has been found. Therefore, WES analysis has retrieved a group of missense variants of unknown clinical significance, some of which affecting interesting candidate genes. Of those, the NM_006832.2: c.1538C>T p.(Thr513Met) variant observed in *FERMT2* gene appears to segregate with the disease in the family, taking into account two phenocopies attributable to a known risk *locus* for epilepsy, 15q13del.

Main limitations and most frequent hurdles to finding causative variants

Rare and many times private genomic variants are the main findings when working with techniques that retrieve large amounts of data, such as aCGH and MPS, namely if performing a WES and/or WGS analysis. Being so, sharing these rare findings within the scientific community is of utmost importance and hence the utility of patient-based databases such as Decipher (Firth et al. 2009) and GeneMatcher (Sobreira, Schiettecatte, Valle, et al. 2015; Sobreira, Schiettecatte, Boehm, et al. 2015). In fact, the finding of other patients that share both similar phenotypic characteristics as well as similar genetic lesions is a criterium that contributes to establish the pathogenicity of a given variant or gene. An example of this approach was the match obtained through GeneMatcher for the variant found in the *FERMT2* gene, in the WES analysis of a large family with epilepsy (chapter 3). Although further work is necessary, namely in the establishment of functional consequences of missense variants affecting the *FERMT2* gene, the fact is that it was possible to correlate two unrelated families that apparently share a similar genetic lesion and a background of epilepsy.

Another aspect that can strengthen the pathogenicity of a variant is related with the fact of it being inherited or resulting from a *de novo* event. The establishment of the inheritance pattern implies the analysis of the parents, but the availability of the parents' samples is not as straightforward as it would be desirable, the small number of the *de novo* events detected in the VOUS group of CNVs (sub-chapter 2.1) being a reflexion of this inaccessibility, particularly in the clinical diagnosis setting, since in most of the cases the parents were not available. When a CNV is

inherited from a similarly affected parent, it indicates that most probably the variant segregates with the disease in the family. However, when dealing with variants that could have incomplete penetrance and/or expressivity, it can be difficult to assess and/or collect, in the time of a consultation, an extensive clinical description that fully illustrate a particular phenotype and, thus, correlate it with a particular variant. Moreover, since most of the patients studied are children, followed by paediatricians (or were children when the follow-up began), it can also be difficult to fully collect the clinical data of the affected parent, since it would most probably imply certain approaches/tests that are out of the scope of a paediatric consultation.

In the specific context of clinical practice, when finding a genomic alteration very well described in many patients and with a clear and well-established pathogenic association with disease, the assignment of cause-effect can be easily done. However, when the clinical implication of the detected genetic alteration is not well established, there is the need for clinical re-evaluation of the patient and/or families in order to look for specific symptoms that could be previously disregarded. Being so, finding more patients with overlapping CNVs (either by means of personal peer communication and/or by searching in patient-databases), will always be useful to a better characterization and establishment of genotype-phenotype correlations. An example of this are the patients described in sub-chapters 2.2 (CNVs encompassing the *AKT3* gene), 2.3 (the microduplications in the 2p15) and 2.4 (the interstitial 7q33 CNVs). Furthermore, this is one aspect in which regular “diagnostic” labs may add a valuable contribution towards the refinement of genotype-phenotype correlations, not only of already known rearrangements/variants, but also in the characterization of new disease-associated variants/genes. Due to their privileged contact and daily access to such a number of cases, they could help providing the raw material to more differentiated research labs that could refer and/or integrate the findings in large series and/or projects, and complement them with functional studies using cell lines and/or animal models, in order to better characterize a particular variant and correlate it with a given phenotype. Being so, sharing the information concerning rare or atypical findings within the scientific community, through platforms such as Decipher database (Firth et al. 2009) or GeneMatcher software (Sobreira,

Schiettecatte, Valle, et al. 2015), is of utmost importance and can contribute to elucidate the pathogenicity of a given variant or gene.

Nevertheless, the need for clinical re-evaluation and continuous follow-up may represent, by itself, a difficult and problematic task for the family. In fact, due to social, economic and sometimes professional aspects/handicaps, the families don't always easily cooperate, by means of attending multiple consultations and assessment by multidisciplinary teams, or willing to participate in projects that surpass the primary goal of achieving a diagnosis for the affected individual. That was for sure the case of many VOUS referred in sub-chapter 2.1: many could possibly be re-grouped either to pathogenic or benign categories, providing that parents/family were available, not only for analysis, but also for the engagement with evaluation process.

These aspects have implications mainly in the clinical context, due to the inability to provide genetic counselling, but also in the research context, i.e. 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.

aCGH findings

When similar rearrangements were not found (i.e. similar-sized deletions or similar-sized duplications) our attention was focused in CNVs spanning the same region (either being deletions or duplications), and in the phenotypic presentation of the patients carrying those CNVs. Deletions and duplications of the same *locus* can present with identical or mirror features, an example of mirrored phenotypes being those observed in carriers of reciprocal 16p11.2 600 kb BP4-BP5 CNVs (Jacquemont et al. 2011) or distal 1q21.1 CNVs (Brunetti-Pierri et al. 2008). Using this approach, novel contributions were made to the definition of genotype-phenotype correlations in interstitial 7q33 CNVs, in 2p15 duplications and in CNVs encompassing the *AKT3* gene. Particular groups of patients could be described and, consequently, the following case-report articles were prepared:

<u>Interstitial 7q33 CNVs</u> (described in sub-chapter 2.4, (Lopes et al. 2018))

The description of a group of patients with rare interstitial 7q33 CNVs (Lopes et al. 2018); allowed us to pinpoint the most likely contributors for the ID and behavioral

phenotype presented by these patients (*CNOT4*, *AGBL3*, and *CALD1* genes). Moreover, we were also able to identify a chimeric *AGBL3/CALD1* transcript in the patients with the 7q33 duplication, to the best of our knowledge, described for the first time. Since no variants were identified by WES analysis in the index case of this family, this chimeric gene appeared as a likely contributor to the phenotype presented by the 7q33 duplication patients. Recurrent fusion genes with transforming potential have been recognized as drivers of cancer for decades, being estimated that these chimeras are responsible for 20% of global cancer morbidities (Dai et al. 2018). Outside the context of cancer, there are only a few reports describing chimeric genes, some of them in NDs such as ASD (Holt et al. 2012), SZ (Rippey et al. 2013) and ID (Mayo et al. 2017). Therefore, the way by which fusion genes may act as a driver of a germline condition is less clearly understood, although it could be by means of positional effect as demonstrated by Rippey and colleagues (Rippey et al. 2013). Being so, a possibility to further elucidate the way by which the chimeric *AGBL3/CALD1* gene could contribute to the phenotype of the 7q33 duplication patients, would be to replicate the experiment of Rippey and colleagues, i.e., to generate expression constructs of both the chimeric *AGBL3/CALD1* gene and parent genes and transfect them into cells in order to see eventual differences in protein subcellular positioning or biological activity.

2p15 duplications (described in sub-chapter 2.3)

Microduplications occurring in the 2p15p16.1 region are rare events presenting variability in breakpoints and duplication sizes. Nevertheless, and despite this variability, patients commonly have ID and variable morphological anomalies, namely digital anomalies. In sub-chapter 2.3 we present a family with three children and their mother, all with DD/learning difficulties and dysmorphic features, carrying a 150Kb duplication in the 2p15p16.1 region, and compare them with other published cases. This is the smallest duplication described so far in this region and includes 3 genes (*C2orf74*, *AHSA2* and *USP34*). *USP34* gene belongs to the critical region of 2p15p16 microdeletion syndrome and has been proposed, together with *BCL11A*, *XPO1* and *REL* genes, to play an important role in the phenotype presented by patients with microdeletions. Moreover, it encodes the ubiquitin specific peptidase 34 that regulates axin stability and Wnt/CTNNB

signaling, that influences different aspects of neuronal circuit (Salinas and Zou 2008). More recently, it was demonstrated that USP34 peptidase controls osteogenic differentiation and bone formation by regulating BMP2-signaling (Guo et al. 2018). The bone morphogenic proteins (BMP) are signaling molecules that belongs to the transforming growth factor-b (TGF-b) superfamily of proteins. BMPs play crucial roles in all organ systems, being important for embryogenesis and development, and also in maintenance of adult tissue homeostasis (R. N. Wang et al. 2014). Therefore, the *USP34* gene, which is disrupted by the CNV, appears as the most promising contributor to the phenotype presented by this family. However, all the tests conducted to date were not supportive of any role of the *USP34* gene in causing the phenotypes in the family: in fact, neither the protein expression from the subject-derived LCLs nor the topologically associated domains showed statistically significant changes compared to the normal controls. As for the *C2orf74* gene, also disrupted by the duplication, its function is currently unknown and therefore its contribution to the phenotype cannot be assessed. Being so, it could be particularly relevant to perform a WES analysis at least in the index case, in order to exclude other possible causes of disease in this family. However, if nothing relevant was found after the WES analysis, several possibilities could be further explored to address the role of this CNV in the phenotype of this family, such as:

a) analyzing the possibility of the occurrence of a *USP34-C2orf74* chimeric gene as the result of the duplication: in this case, as discussed previously, its contribution to the phenotype could occur by means of positional effect (Rippey et al. 2013);

b) generating and analyzing a *USP34* duplication model in *Drosophila*;

c) performing a transcriptomic analysis using induced pluripotent stem cells (iPSCs) derived from these patients, in order to analyze the transcriptional differences between 2p15 duplication patient iPSC-derived neurons and unaffected controls.

CNVs encompassing the *AKT3* gene (described in sub-chapter 2.2, Lopes et al. Frontiers Genetics, 2018)

Deletions occurring at 1q43-q44 are generally considered deleterious and are usually associated with a clinically recognizable phenotype of ID, facial

dysmorphisms and microcephaly. With the series of patients described in this thesis it was possible to highlight the not so straightforward and isolated implication of *AKT3* CNVs in the definition of human occipitofrontal circumference. Although it was possible to corroborate the evidence that *AKT3* is a key gene for microcephaly in patients with 1q43-q44 deletions, it was also clear that other factors, such as different genetic or epigenetic backgrounds of the individuals, must play a role in the arising of the phenotype, resulting in incomplete penetrance and variable expressivity. The same conclusion applies for *AKT3* copy number gains and its association with the mirror phenotype (macrocephaly): patient 4 in our series does present macrocephaly, supporting this model, previously described in the literature (Chung et al. 2014; Conti et al. 2015; Hemming et al. 2016; D. Wang et al. 2013). Nevertheless, his father is phenotypically normal and, to the best of our knowledge, becomes the first reported asymptomatic individual with a quadruplication affecting the *AKT3* gene (Lopes et al. 2018, in press). Since a trio WES analysis was performed in patient 4/parents without retrieving any causative variant, an additional possibility for elucidating this variation in phenotypic presentation would be to perform a methylome analysis. Studies of the brain and neurons have outlined an increasingly complex architecture of methylation patterns, and different modifications contribute differently to fine-tuning the regulation of gene expression and may play an important role in brain development, function and decline (Klein and De Jager 2016). Recognizing this level of heterogeneity and integrating it with histone modifications and other epigenetic markers are important steps to better understand the role of methylation in central nervous system (CNS). The epigenome reflects the conformational status of chromatin and, hence, the accessibility of a DNA segment to the transcriptional machinery; it may therefore represent a good substrate to identify disease genes, not because they are targeted by genetic variation, but because they may reflect the action of non-genetic risk factors (Klein and De Jager 2016).

Expression studies in the cohort of patients with idiopathic ID

In some selected cases (further described in sub-chapter 2.1), mRNA was extracted from peripheral blood of patients with copy number gains to perform expression analysis by quantitative real-time reverse transcription PCR (qRT-

PCR), for genes located either inside the duplicated/triplicated regions and/or at the breakpoints, in order to determine if there is an actual effect of gene dosage at the transcription level. We would expect *a priori* an increased expression in genes inside the duplication and a diminished expression in genes located at the breakpoints and apparently disrupted by the CNV. It was, in fact, possible to observe an increased expression in some genes located inside duplications, being the most interesting finding the increased expression of *CUL4B* gene observed in patient R22. Both point mutations and large deletions affecting *CUL4B* gene are described as causative of X-linked ID and cerebral malformations (Tarpey et al. 2007), and *CUL4B* is classified, accordingly, as a developmental disorder gene by the Developmental Disorders Genotype-Phenotype Database (DDG2P), from Decipher (Firth et al. 2009). Patient R22 presents an increased expression of *CUL4B* gene, which is consistent with the fact that the gene is completely inside the duplication and not disrupted by it, leading us to propose that the patient disorder could be driven by this increased expression of *CUL4B* gene. A positive correlation between expression levels and gene dosage for all genes within the 16p11.2 CNV region was observed by Jacquemont and colleagues; however, they also observed that while genes proximal (centromeric) to the rearrangement interval showed no significant variation in relative transcript levels between patients and controls, distal (telomeric) genes showed a significant alteration in relative expression, being similarly upregulated in both deletion and duplication carriers (Jacquemont et al. 2011). Therefore, when correlating expression data with chromosomal gains or losses, one must bear in mind that the presence/structure of the CNV itself, rather than the change in copy number, may affect transcription, and that CNVs can influence the transcriptome not merely by affecting the expression of strictly co-localizing genes, but by inducing alterations in chromatin architecture distal to the structural variant, that could modulate expression globally and modify the phenotype (Gheldof et al. 2013; Gamazon and Stranger 2015). Moreover, CNVs combined with epigenetic mechanisms can also influence transcription, such is the case of CNVs located on imprinted *loci* (Hogart et al. 2009). Being so, carefully designed, large-scale transcriptome studies across multiple tissues and cell types could help to elucidate the way by which a given CNV may affect expression networks and pathways and, thus, contribute to disease.

Rare de novo VOUS CNVs

As stated before, the most significant group of CNVs detected in the cohort of ID patients studied by aCGH in this thesis was constituted by rare CNVs, many restricted to one patient/family. A significant group of them (~17% of the cases) were classified as VOUS, being the full list in the sub-chapter 2.1. Despite the uncertainty of their clinical significance, there are some cases that deserve further attention, like those corresponding to *de novo* CNVs. *De novo* rearrangements are commonly associated with pathogenic CNVs and often affect dozens of genes. However, it was possible to identify seven cases in the VOUS group in which the CNV arised *de novo* in the patient, this number reflecting not the real number of *de novo* CNVs but instead the unavailability of parent's samples to test inheritance. Furthermore, these are relatively small CNVs, affecting a small number of genes, and no assigned *a priori* as pathogenic or likely pathogenic. For further discussion, these seven cases are presented in Table 4.1.

Table 4.1 - List of *de novo* copy number variants of unknown significance (CNVOUS).

Patients	Gender	ISCN description (Hg19)	Type	Size (Kb)	Number of Genes	Genes (name)	Confirmation	Inheritance	DGV controls	Similar case (Decipher)
R24	Male	1p13.2(112,243,130-112,331,235)X3	dup	88	7	*AK023457, RAP1A, INKA2, LOC100506343, BC041890, DDX20, KCND3*	qPCR	*de novo*	No	No
R28	Male	3q26.33(181,357,672-181,466,211)X1	del	108	2	*SOX2; SOX2OT*	qPCR	*de novo*	No	301549 (smaller dup)
C38	Fenale	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
R38	Male	17p13.1(6,955,115-7,409,331)X1	del	454	30	*DLG4, GABARAP, DULLARD, NEURL4, NLGN2, CHRNB1*	NP	*de novo*	3/19159 (0,000052)	260507, 289609, 258519 2346, 3474 (smaller del)
R43	Male	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 (bigger)	289634
R48	Male	3p21.31(48,464,967-48,574,235)X3	dup	109	7	*ATRIP, CCDC51, CCDC72, PFKFB4, PLXNB1, SHISA5, TREX1*	NP	Maternal¥	1/ 17421 (del) 1/2026 (del)	No
		7p22.3(1,565,982-1,701,871)X3	dup	136	5	*KIAA1908, MAFK, PSMG3, TFAMP1, TMEM184A*	NP	Paternal	1/270 (del)	No
		10p12.31(20,641,191-21,122,699)X3	dup	481	2	*NEBL, PLXDC2*	NP	Maternal¥	No	250836 (smaller)
		14q31.3(88,794,387-88,853,440)X3	dup	59	1	*SPATA7*	NP	*de novo*	No	No
R50	Male	18q12.1(29,316,291-29,569,853)X1	del	254	2	*TRAPPC8, SLC25A52*	NP	Paternal ¥	No	No
		20q11.23(36,531,120-36,618,758)X3	dup	88	2	*VSTM2L*	NP	*de novo*	1/29084 (del, bigger)	No

Two of these patients (R48 and R50), present not only a *de novo* CNV but also CNVs inherited from parents; in these cases, there is the question of whether these alterations together could lead to the disease in the child through epistatic effects.

The remaining five patients have the *de novo* CNV as the sole finding of the aCGH analysis, consistently with the increased prevalence of *de novo* mutations which has been observed in NDs (McRae et al. 2017). Some of the genes involved are already associated with human disease, such is the case of *KCND3*, associated with Brugada syndrome 9 (MIM 616399) and spinocerebellar ataxia 19 (MIM 607346), both with autosomal dominant (AD) inheritance; *SOX2*, associated with microphthalmia, syndromic 3 AD (MIM 206900); *GPT2*, associated with mental retardation, autosomal recessive (AR) 49 (MIM 616281); *CHRNB1*, with myasthenic syndrome, congenital (AD and AR) (MIM 616313 and 616314), and *SPATA7*, with Leber congenital amaurosis-3 and juvenile retinitis pigmentosa (AR) (MIM 604232). Some of these genes are associated with disorders with an autosomal recessive pattern of inheritance (*GPT2*), and therefore the disease in the patients bearing these variants could not be explained by a heterozygous CNV. Other genes are associated with disorders that are not associated *a priori* with neurodevelopmental conditions (*CHRNB1* and *SPATA7*). Nevertheless, many of the genes within these CNVs which have not yet been associated with human disease have, interestingly, a broad or even an ubiquitous expression in the brain (Bastian et al. 2008; Fagerberg et al. 2014), such is the case of *RAP1A*, *INKA2*, *DDX20*, *DNAJA2*, *ITFG1*, *NETO2*, *DLG4*, *GABARAP*, *CTDNEP1*, *NEURL4*, *NLGN2* and *ZSCAN18*. Of this group, there are some whose function led us to hypothesize for a possible involvement in neurodevelopment (data retrieved from Entrez Gene database, OMIM):

a. *NETO2* (*neuropilin and tolloid like 2*) (16q11.2q12.1dup, patient C38): this gene encodes a predicted transmembrane protein containing two extracellular CUB domains followed by a low-density lipoprotein class A (LDLa) domain. The orthologous in mice encodes a protein that modulates glutamate signalling in the brain by regulating kainate receptor function (Tang et al. 2012).

b. *NLGN2* (*neuroligin 2*) (17p13.1del, patient R38): this gene encodes a member of a family of neuronal cell surface proteins. Neuroligins are trans-synaptic adhesion molecules thought to function in synapse formation, specification, or both. They are ligands of neurexins, which in turn are synaptic cell adhesion molecules involved in synapse specification (Chubykin et al. 2007).
c. *DLG4* (*discs large MAGUK scaffold protein 4*) (17p13.1del, patient R38): this gene encodes a member of the discs large subfamily of the membrane-associated guanylate kinase (MAGUK) family. It heteromultimerizes with another MAGUK protein, DLG2, and is recruited into NMDA receptor and potassium channel clusters, playing an important role in the formation and maintenance of synaptic junctions (Zhou and Blumberg 2003).
d. *GABARAP* (*GABA type A receptor-associated protein*) (17p13.1del, patient R38): Gamma-aminobutyric acid A receptors [GABA(A) receptors] are ligand-gated chloride channels that mediate inhibitory neurotransmission. This gene encodes GABA(A) receptor-associated protein, that clusters neurotransmitter receptors by mediating interaction with the cytoskeleton. A knockdown model in zebrafish suggested that the *Gabarap* gene plays a role in brain development (Komoike et al. 2010).
e. *NEURL4* (*neuralized E3 ubiquitin protein ligase 4*) (17p13.1del, patient R38): this gene encodes a protein that interacts with the centrosomal capping protein CP110, promoting its ubiquitination and proteasome-dependent degradation; by counteracting accumulation of CP110, NEURL4 maintains normal centriolar homeostasis and prevents formation of ectopic microtubular organizing centers (Li et al. 2012).

Given the data above, the follow-up and further evaluation of these cases would be recommended, since, as long as new findings become available, a better assessment of a possible role of these genes in NDs could be either confirmed or ruled out.

MPS findings in a large family with epilepsy

The WES analysis performed in the three family members of a large epilepsy family did not retrieve a genetic variant with a well-known and established association with epilepsy which could unequivocal explain the disease in this family. The fact that we did not find any clearly causal variant in the WES performed is not surprising, *per se*. In fact, a substantial fraction of phenotypically and/or genetically heterogeneous conditions, such as NDs, and particularly ID and epilepsy, remain unexplained even after a WGS analysis: for example, the diagnostic yield in severe ID is ~42% for WGS and ~40% for WES (Wright, FitzPatrick, and Firth 2018).

Being so, several approaches could be employed in an attempt to find causative variants, being the most obvious the reanalysis of the WES data. In fact, several publications have stressed the importance of ongoing reanalysis of negative exome sequences, as advances in the knowledge of disease genes and their annotations will permit new diagnoses to be made. One of such examples, is a recent publication from the Epilepsy Genetics Initiative (EGI, http://www.cureepilepsy.org/egi/index.html), in which researchers reanalyzed 54 previously undiagnosed WES trios and reported the identification of novel disease-causing variants in alternative exon 5 A of *SCN8A* gene in three unrelated patients with epilepsy. These causing variants were initially missed since, at the time of analysis, exon 5 A was not recognized as protein coding in the consensus coding sequence database (EGI 2018).

In the present case, however, the familial segregation analysis performed in selected gene variants found in all the three family members studied by WES indicated *FERMT2* gene as the most likely candidate to explain the epilepsy phenotype in several individuals of this family. This likelihood was further supported by the finding of another *FERMT2* variant in an unrelated epilepsy patient by WES-trio analysis. Ideally, this association should be confirmed by the description of more unrelated patients/families with epilepsy-phenotypes sharing *FERMT2* variants, hence the relevance of patient-based databases, in addition to a peer inter-lab exchange of data. Yet, the possibility of these two being *FERMT2* "orphan" variants remains, and for that reason their contribution to epileptogenesis should be supported by functional evidence. One way of obtaining this evidence would be to establish iPSCs derived-neurons from patients of the two families, to

compare their neuronal differentiation, electrophysiological and transcriptional profiles, and additionally compare them with cells derived from unaffected controls. In fact, multi-omics analysis, such as transcriptomics, epigenomics, metabolomics and proteomics, although not yet routinely implemented in a diagnostic setting, can be used to investigate the functional impact of genetic variation on specific tissues (Wright, FitzPatrick, and Firth 2018). For example, transcriptomic data of muscle biopsy samples was used to guide genomic reanalysis in a cohort of undiagnosed patients with pediatric neuromuscular diseases, retrieving diagnosis for 21% of patients with no strong candidate genes previously identified by WGS or WES solely (Cummings et al. 2017). Therefore, a reanalysis of the data, eventually complemented with additional transcriptomic and/or methylome analysis, could help in the elucidation of the etiological basis of this family's disease.

Additionally, exploring the relationship between sequence variation and epigenetic state could also provide novel insights of the mechanisms of disease. Rare epigenetic changes were identified in ~20% of a large cohort of patients with diverse NDs and/or congenital anomalies, without any putative pathogenic mutation identified by microarray testing and/or WES (WGS was only performed in some patients); these *de novo* epivariations were significantly enriched in cases and often had an impact on gene expression comparable to loss-of-function mutations, having, most likely, diagnostic relevance (Barbosa et al. 2018).

Patient-based databases, rare variants, rare diseases

Patient-based databases are an utmost importance since they provide a collection of both clinical and genetic data concerning a particular condition, and are, nowadays, essential tools both for research and clinical diagnosis. The UK's Deciphering Developmental Disorders (DDD, http://www.ddduk.org/intro.html) project was launched in order to undertake systematic phenotyping and detailed genomic analysis for 12.000 children with severe diagnosed developmental disorders, combining the clinical expertise of all the UK NHS Regional Genetics Services with the research and bioinformatics expertise of the Wellcome Trust Sanger Institute (Firth and Wright 2011). The results of this project are easily found on PubMed: a search using DDD Study [Corporate Author] retrieved nearly 80 articles, most of them published between 2016 and 2018. Moreover, this

project paved the way for the 100,000 Genomes Project, i.e., the sequencing of 100,000 genomes from around 70,000 people from UK's NHS, being the participants, patients with a rare disease, plus their families, and patients with cancer (https://www.genomicsengland.co.uk/the-100000-genomes-project/). A similar initiative was launched in Japan, with the Initiative on Rare and Undiagnosed Diseases (IRUD) (Adachi et al. 2017). Studying these large cohorts is extremely important, not only due to the acquisition of both categorical and quantitative phenotypic data, but also since they facilitate the development of statistically robust methods for establishing pathogenicity solely on the basis of genomic data (McRae et al. 2017).

Last but not least, a global cooperation and collaboration in order to capitalize and maximize the output of rare diseases research efforts around the world is the commitment of the International Rare Diseases Research Consortium (IRDiRC). The establishment of global networks of clinical and research laboratories in order to enable all people living with a rare disease to receive an accurate diagnosis, care and available therapy within one year of coming to medical attention are their major goals for the next decade (Austin et al. 2017), and hopefully will contribute to our understanding of the etiology of several disorders and for the ever-improving diagnosis of patients.

Main findings and conclusions of this thesis

1. We identified pathogenic or likely pathogenic variants in 13.2% of patients with ID.

2. Most of the likely pathogenic aCGH findings were rare CNVs, many restricted to one patient/family, whose clinical significance needs to be carefully addressed; nevertheless, 12 new candidate *loci* for ID were put forward: 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.

3. We confirmed that *AKT3* is a key gene for microcephaly in patients with 1q43-q44 deletions, although other factors, such as different genetic or epigenetic backgrounds of the individuals, must play a role in the arising of the phenotype, resulting in incomplete penetrance and variable expressivity.

4. We showed that microduplications in the 2p15p16.1 region are rare events presenting variability in breakpoints and duplication sizes. Patients commonly

have ID and variable morphological anomalies, including digital anomalies; however, the underlying genetic mechanism causing this abnormal phenotype remains puzzling and need to be further addressed.

5. We showed that small CNVs at 7q33 region cause ID and behavioral alterations, namely aggressiveness and disinhibition, the *CALD1* gene being the most likely candidate for the core phenotype.

6. We observed an increased expression, consistent with the change in copy number, for several genes within duplications; one of these genes was *CUL4B*, the increased expression being observed in a patient carrying a Xq24 duplication, and proposed to be the most likely disease contributor for patient´s phenotype.

7. We pinpointed the *FERMT2* gene as the most likely candidate to explain the epilepsy phenotype observed in several individuals of a large Portuguese family.

Future perspectives

Although it was possible to identify several genetic alterations as the likely candidates to explain the disease in particular patients and/or group of patients, additional work should be done to gather further support to the evidence collected. Being so, it would be desirable to:

a. Establish iPSCs derived-neurons from 2p15 dup patients, in order to define the genetic mechanism associated with their abnormal phenotype (ID and variable morphological anomalies);
b. Analyze a *USP34* duplication model in a model organism (for example, *Drosophila melanogaster*);
c. Perform a database search (in exome databases) for *USP34* variants;
d. Perform a linkage analysis in the large family with epilepsy to obtain further evidence that the heterozygous *FERMT2* variant identified maps with in a chromosomal region that is linked to the disease;
e. Establish iPSCs and iPSCs-derived neurons from epilepsy patients with heterozygous variants in the *FERMT2* gene, in order to investigate the functional impact of the detected variant;
f. Analyze a mouse model of *fermt2* heterozygous mutation;
g. Perform a methylome analysis in patients with *AKT3* CNVs, to ascertain if the phenotypic differenced observed in patients sharing similar CNVs result from epigenetic differences;

h. Review the cases of CNVs classified as VOUS, namely those with *de novo* CNVs affecting *NLGN2*, *NETO2*, *DLG4*, *GABARAP* and *NEURL4* genes; not only they are highly expressed in the brain, hence interesting candidates *a priori*, but also due to the prevalence of *de novo* mutations observed in NDs.

Bibliography

Adachi, Takeya, Kazuo Kawamura, Yoshihiko Furusawa, Yuji Nishizaki, Noriaki Imanishi, Senkei Umehara, Kazuo Izumi, and Makoto Suematsu. 2017. “Japan’s Initiative on Rare and Undiagnosed Diseases (IRUD): Towards an End to the Diagnostic Odyssey.” *European Journal of Human Genetics* 25 (9). Nature Publishing Group:1025–28. https://doi.org/10.1038/ejhg.2017.106.

Austin, Christopher P, Christine M Cutillo, Lilian PL Lau, Anneliene H Jonker, Ana Rath, Daria Julkowska, David Thomson, et al. 2017. “Future of Rare Diseases Research 2017-2027: An IRDiRC Perspective.” *Clinical and Translational Science*, no. October 2017:21–27. https://doi.org/10.1111/cts.12500.

Barbosa, Mafalda, Ricky S. Joshi, Paras Garg, Alejandro Martin-Trujillo, Nihir Patel, Bharati Jadhav, Corey T. Watson, et al. 2018. “Identification of Rare de Novo Epigenetic Variations in Congenital Disorders.” *Nature Communications* 9 (1). Springer US. https://doi.org/10.1038/s41467-018-04540-x.

Bastian, Frederic, Gilles Parmentier, Julien Roux, Sebastien Moretti, Vincent Laudet, and Marc Robinson-Rechavi. 2008. “Bgee: Integrating and Comparing Heterogeneous Transcriptome Data among Species.” *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5109 LNBI:124–31. https://doi.org/10.1007/978-3-540-69828-9_12.

Brunetti-Pierri, Nicola, Jonathan S Berg, Fernando Scaglia, John Belmont, Carlos A Bacino, Trilochan Sahoo, Seema R Lalani, et al. 2008. “Recurrent Reciprocal 1q21.1 Deletions and Duplications Associated with Microcephaly or Macrocephaly and Developmental and Behavioral Abnormalities.” *Nat Genet* 40 (12):1466–71. https://doi.org/10.1038/ng.279.Recurrent.

Chubykin, Alexander A., Deniz Atasoy, Mark R. Etherton, Nils Brose, Ege T. Kavalali, Jay R. Gibson, and Thomas C. Südhof. 2007. “Activity-Dependent Validation of Excitatory versus Inhibitory Synapses by Neuroligin-1 versus Neuroligin-2.” *Neuron* 54 (6):919–31. https://doi.org/10.1016/j.neuron.2007.05.029.

Chung, Brian K., Patrice Eydoux, Clara D. An Karnebeek, and William T. Gibson. 2014. “Duplication of AKT3 Is Associated with Macrocephaly and Speech Delay.” *American Journal of Medical Genetics. Part A* 164 (7):1868–69. https://doi.org/10.1002/ajmg.a.36521.

Conti, V., M. Pantaleo, C. Barba, G. Baroni, D. Mei, A. M. Buccoliero, S. Giglio, et al. 2015. “Focal Dysplasia of the Cerebral Cortex and Infantile Spasms Associated with Somatic 1q21.1-Q44 Duplication Including the AKT3 Gene.” *Clinical Genetics* 88 (3):241–47. https://doi.org/10.1111/cge.12476.

Cummings, Beryl B, Jamie L Marshall, Taru Tukiainen, Monkol Lek, A Reghan Foley,

Veronique Bolduc, Leigh B Waddell, et al. 2017. “IMproving Genetic Diagnosis in Mendelian Disease with Transcriptome Sequencing.” *Sci Transl Med* 9 (386):1–25. https://doi.org/10.1126/scitranslmed.aal5209.Improving.

Dai, Xiaofeng, Rutaganda Theobard, Hongye Cheng, Mengtao Xing, and Jianying Zhang. 2018. “Fusion Genes: A Promising Tool Combating against Cancer.” *Biochimica et Biophysica Acta - Reviews on Cancer* 1869 (2). Elsevier B.V:149–60. https://doi.org/10.1016/j.bbcan.2017.12.003.

EGI. 2018. “De Novo Variants in the Alternative Exon 5 of SCN8A Cause Epileptic Encephalopathy.” *Genetics in Medicine* 20 (2):275–81. https://doi.org/10.1038/gim.2017.100.

Fagerberg, Linn, Bjorn M Hallstromm, Per Oksvold, Caroline Kampf, Dijana Djureinovic, Jacob Odeberg, Masato Habuka, et al. 2014. “Synthesis and Coordination Chemistry of Thiophenol-Based Heterodonor Ligands Containing P,S, as,S and P,SAs Donor Atoms.” *Molecular & Cellular Proteomics* 13:397–406. https://doi.org/10.1074/mcp.M113.035600.

Firth, Helen V., Shola M. Richards, A. Paul Bevan, Stephen Clayton, Manuel Corpas, Diana Rajan, Steven Van Vooren, Yves Moreau, Roger M. Pettett, and Nigel P. Carter. 2009. “DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources.” *American Journal of Human Genetics* 84 (4). The American Society of Human Genetics:524–33. https://doi.org/10.1016/j.ajhg.2009.03.010.

Firth, Helen V., and Caroline F. Wright. 2011. “The Deciphering Developmental Disorders (DDD) Study.” *Developmental Medicine and Child Neurology* 53 (8):702–3. https://doi.org/10.1111/j.1469-8749.2011.04032.x.

Gamazon, Eric R., and Barbara E. Stranger. 2015. “The Impact of Human Copy Number Variation on Gene Expression.” *Briefings in Functional Genomics* 14 (5):352–57. https://doi.org/10.1093/bfgp/elv017.

Gheldof, Nele, Robert M. Witwicki, Eugenia Migliavacca, Marion Leleu, Gérard Didelot, Louise Harewood, Jacques Rougemont, and Alexandre Reymond. 2013. “Structural Variation-Associated Expression Changes Are Paralleled by Chromatin Architecture Modifications.” *PLoS ONE* 8 (11). https://doi.org/10.1371/journal.pone.0079973.

Guo, Yu-chen, Meng-yuan Wang, Shi-wen Zhang, Yun-shu Wu, Chen-chen Zhou, Ri-xin Zheng, Bin Shao, et al. 2018. “Ubiquitin-specific Protease USP34 Controls Osteogenic Differentiation and Bone Formation by Regulating BMP2 Signaling.” *The EMBO Journal* e99398:1–17. https://doi.org/10.15252/embj.201899398.

Hemming, Isabel A., Alistair R.R. Forrest, Peter Shipman, Karen J. Woodward, Peter Walsh, David G. Ravine, and Julian Ik Tsen Heng. 2016. “Reinforcing the Association

between Distal 1q CNVs and Structural Brain Disorder: A Case of a Complex 1q43-Q44 CNV and a Review of the Literature." *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics* 171 (3):458–67. https://doi.org/10.1002/ajmg.b.32427.

Hogart, Amber, Karen N Leung, Nicholas J Wang, David J Wu, Jennette Driscoll, Roxanne O Vallero, N Carolyn Schanen, and Janine M Lasalle. 2009. "Chromosome 15q11-13 Duplication Syndrome Brain Reveals Epigenetic Alterations in Gene Expression Not Predicted from Copy Number." *J Med Genet* 46 (2):86–93. https://doi.org/10.1136/jmg.2008.061580.Chromosome.

Holt, Richard, Nuala H Sykes, Inês C Conceição, Jean-Baptiste Cazier, Richard JL Anney, Guiomar Oliveira, Louise Gallagher, Astrid Vicente, Anthony P Monaco, and Alistair T Pagnamenta. 2012. "CNVs Leading to Fusion Transcripts in Individuals with Autism Spectrum Disorder." *European Journal of Human Genetics* 20 (11):1141–47. https://doi.org/10.1038/ejhg.2012.73.

Jacquemont, Sébastien, Alexandre Reymond, Flore Zufferey, Louise Harewood, Robin G Walters, Zoltán Kutalik, Danielle Martinet, et al. 2011. "Mirror Extreme BMI Phenotypes Associated with Gene Dosage at the Chromosome 16p11.2 Locus." *Nature* 478 (7367). Nature Publishing Group:97–102. https://doi.org/10.1038/nature10406.

Klein, Hans Ulrich, and Philip L. De Jager. 2016. "Uncovering the Role of the Methylome in Dementia and Neurodegeneration." *Trends in Molecular Medicine* 22 (8). Elsevier Ltd:687–700. https://doi.org/10.1016/j.molmed.2016.06.008.

Komoike, Yuta, Keiko Shimojima, Jao Shwann Liang, Hiroshi Fujii, Yoshihiro Maegaki, Makiko Osawa, Sakiko Fujii, Toru Higashinakagawa, and Toshiyuki Yamamoto. 2010. "A Functional Analysis of GABARAP on 17p13.1 by Knockdown Zebrafish." *Journal of Human Genetics* 55 (3). Nature Publishing Group:155–62. https://doi.org/10.1038/jhg.2010.1.

Li, Ji, Sehyun Kim, Tetsuo Kobayashi, Feng Xia Liang, Nina Korzeniewski, Stefan Duensing, and Brian D. Dynlacht. 2012. "Neurl4, a Novel Daughter Centriole Protein, Prevents Formation of Ectopic Microtubule Organizing Centres." *EMBO Reports* 13 (6):547–53. https://doi.org/10.1038/embor.2012.40.

Lopes, Fátima, Fátima Torres, Sally Ann Lynch, Arminda Jorge, Susana Sousa, João Silva, Paula Rendeiro, Purificação Tavares, Ana Maria Fortuna, and Patrícia Maciel. 2018. "The Contribution of 7q33 Copy Number Variations for Intellectual Disability." *Neurogenetics* 19 (1). neurogenetics:27–40. https://doi.org/10.1007/s10048-017-0533-5.

Lu, Xinyan, Chad A. Shaw, Ankita Patel, Jiangzhen Li, M. Lance Cooper, William R.

Wells, Cathy M. Sullivan, et al. 2007. “Clinical Implementation of Chromosomal Microarray Analysis: Summary of 2513 Postnatal Cases.” *PloS One* 2 (3):e327. https://doi.org/10.1371/journal.pone.0000327.

Maulik, Pallab K, Maya N Mascarenhas, Colin D Mathers, Tarun Dua, and Shekhar Saxena. 2011. “Prevalence of Intellectual Disability: A Meta-Analysis of Population-Based Studies.” *Research in Developmental Disabilities* 32:419–36. https://doi.org/10.1016/j.ridd.2010.12.018.

Mayo, Sonia, Sandra Monfort, Mónica Roselló, Carmen Orellana, Silvestre Oltra, Alfonso Caro-Llopis, and Francisco Martínez. 2017. “Chimeric Genes in Deletions and Duplications Associated to Intellectual Disability.” *International Journal of Genomics*, 1–11. https://doi.org/10.1155/2017/4798474.

McRae, Jeremy F., Stephen Clayton, Tomas W. Fitzgerald, Joanna Kaplanis, Elena Prigmore, Diana Rajan, Alejandro Sifrim, et al. 2017. “Prevalence and Architecture of de Novo Mutations in Developmental Disorders.” *Nature* 542 (7642):433–38. https://doi.org/10.1038/nature21062.

Moeschler, J. B., and M. Shevell. 2014. “Comprehensive Evaluation of the Child With Intellectual Disability or Global Developmental Delays.” *Pediatrics* 134 (3):e903–18. https://doi.org/10.1542/peds.2014-1839.

Rippey, Caitlin, Tom Walsh, Suleyman Gulsuner, Matt Brodsky, Alex S. Nord, Molly Gasperini, Sarah Pierce, et al. 2013. “Formation of Chimeric Genes by Copy-Number Variation as a Mutational Mechanism in Schizophrenia.” *American Journal of Human Genetics* 93 (4). The American Society of Human Genetics:697–710. https://doi.org/10.1016/j.ajhg.2013.09.004.

Rosenberg, C., J. Knijnenburg, E. Bakker, A. M. Vianna-Morgante, W. Sloos, P. A. Otto, M. Kriek, et al. 2006. “Array-CGH Detection of Micro Rearrangements in Mentally Retarded Individuals: Clinical Significance of Imbalances Present Both in Affected Children and Normal Parents.” *Journal of Medical Genetics* 43 (2):180–86. https://doi.org/10.1136/jmg.2005.032268.

Sagoo, Gurdeep S., Adam S. Butterworth, Simon Sanderson, Charles Shaw-Smith, Julian P. T. Higgins, and Hilary Burton. 2009. “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.” *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 11 (3):139–46. https://doi.org/10.1097/GIM.0b013e318194ee8f.

Salinas, Patricia C., and Yimin Zou. 2008. “Wnt Signaling in Neural Circuit Assembly.” *Annual Review of Neuroscience* 31 (1):339–58. https://doi.org/10.1146/annurev.neuro.31.060407.125649.

Sobreira, Nara, François Schiettecatte, Corinne Boehm, David Valle, and Ada Hamosh. 2015. “New Tools for Mendelian Disease Gene Identification: PhenoDB Variant Analysis Module; and GeneMatcher, a Web-Based Tool for Linking Investigators with an Interest in the Same Gene.” *Human Mutation* 36 (4):425–31. https://doi.org/10.1002/humu.22769.New.

Sobreira, Nara, François Schiettecatte, David Valle, and Ada Hamosh. 2015. “GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene.” *Human Mutation* 36 (10):928–30. https://doi.org/10.1002/humu.22844.

Tang, Man, Evgueni Ivakine, Vivek Mahadevan, Michael W Salter, and Roderick R McInnes. 2012. “Neto2 Interacts with the Scaffolding Protein GRIP and Regulates Synaptic Abundance of Kainate Receptors.” *PloS One* 7 (12):e51433. https://doi.org/10.1371/journal.pone.0051433.

Tarpey, Patrick S, F Lucy Raymond, Sarah O Meara, Sarah Edkins, Jon Teague, Adam Butler, Ed Dicks, et al. 2007. “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.” *American Journal of Human Genetics* 80:345–52.

Wang, David, Susan Zeesman, Mark A. Tarnopolsky, and Małgorzata J. M. Nowaczyk. 2013. “Duplication of AKT3 as a Cause of Macrocephaly in Duplication 1q43q44.” *American Journal of Medical Genetics. Part A* 161A (8):2016–19. https://doi.org/10.1002/ajmg.a.35999.

Wang, Richard N., Jordan Green, Zhongliang Wang, Youlin Deng, Min Qiao, Michael Peabody, Qian Zhang, et al. 2014. “Bone Morphogenetic Protein (BMP) Signaling in Development and Human Diseases.” *Genes and Diseases* 1 (1). Elsevier Ltd:87–105. https://doi.org/10.1016/j.gendis.2014.07.005.

Wright, Caroline F., David R. FitzPatrick, and Helen V. Firth. 2018. “Paediatric Genomics: Diagnosing Rare Disease in Children.” *Nature Reviews Genetics* 19 (5). Nature Publishing Group:253–68. https://doi.org/10.1038/nrg.2017.116.

Zhou, Changcheng, and Bruce Blumberg. 2003. “Overlapping Gene Structure of Human VLCAD and DLG4.” *Gene* 305 (2):161–66. https://doi.org/10.1016/S0378-1119(02)01235-0.