UNIVERSITÀ DEGLI STUDI DI SASSARI

CORSO DI DOTTORATO DI RICERCA IN SCIENZE BIOMEDICHE

Coordinatore del Corso: Prof. Andrea Fausto Piana

CURRICULUM IN GENETICA MEDICA

Responsabile di Curriculum: Prof.ssa Maioli Margherita

XXXII CICLO

How exome sequencing is shedding light on the complexity of Mendelian disorders: some examples from Sardinia

Coordinatore:

Prof. Andrea Fausto Piana

Tutor:

Prof. Francesco Cucca

Tesi di dottorato di:

Dott.ssa Laura Crisponi

Index

Abstract *2*

Introduction *4*

Rationale of the work *8*

Mendelian disorders under study *10*

Crisponi Syndrome/Cold-induced sweating syndrome type 1 (CS/CISS1)-like phenotypes **10**

Epileptic Encephalopathies (EE) **11**

Syndromic Intellectual Disability (SID) **13**

Methods *15*

Patients' selection, samples acquisition and DNA extraction **15**

Whole-Exome Sequencing workflow **18**

Sequence analysis and variant calling pipeline **19**

Biological interpretation and identification of causal alleles **20**

Genetic validation of sequence variants **22**

Results *22*

Crisponi syndrome/Cold-induced sweating syndrome type 1 (CS/CISS1)-like phenotypes **22**

Epileptic Encephalopathies (EE) **23**

Syndromic Intellectual Disability (SID) **25**

Discussion *27*

From simple beginnings to complex endings **28**

"Phenotype-driven" to "genotype-driven" **29**

WES complementary approaches **30**

Clinical and functional interpretation of variants of uncertain significance **31**

Conclusions *32*

Take-home messages from Mendelian disorders research **33**

References *34*

WEB resources *44*

Acknowledgments *45*

Abstract

The total number of identified rare Mendelian disorders is currently estimated to be around 7,000, and while each is individually rare, together, these genetic conditions contribute significantly to morbidity, mortality, and healthcare costs.

In the last decade, there has been a paradigm shift in the investigation of Mendelian disorders, due to the development of powerful new DNA sequencing technologies, commonly referred to as Next-Generation Sequencing (NGS). Whole exome sequencing (WES) is an NGS method that analyzes all protein-coding regions (exons) of genes in the genome. When compared with other methods WES has the critical advantages that it does not require any *a priori* hypothesis about the nature of the defective genes and that it offers a high coverage of the exons. Considering that exonic mutations cause the vast bulk of Mendelian disorders, currently WES is thus the most powerful approach for Mendelian gene identification. Still, despite the recent progress in knowledge of the diversity of Mendelian phenotypes largely due to WES, substantial gaps remain. Up to 50% of patients affected by a rare genetic disorder never receive a diagnosis.

Diagnosis depends on gene discovery, and early diagnosis is crucial to establish an adequate treatment and to identify novel knowledge-based therapeutic targets to improve the overall outcome.

We focused our attention on such rare Mendelian disorders, and in a collaborative effort, we studied by WES a cohort of heterogeneous samples affected by the following disorders: Crisponi syndrome/Cold-induced sweating syndrome type 1 (CS/CSS1)-like, syndromic Intellectual Disabilities and Epileptic Encephalopathies (EE).

The results of our work led to the finding of both recessive and dominant pathogenic variants of human NALCN associated complex neurodevelopmental syndromes. Recessive pathogenic variants of *NALCN* are linked to the infantile hypotonia with psychomotor retardation, and characteristic facies type 1 syndrome (IHPRF), while dominant pathogenic variants of *NALCN* are responsible for the congenital contractures of limbs and face, hypotonia and developmental delay syndrome (CLIFAHDD). The discoveries of pathogenic variants in *NALCN, MAGEL2* and *SCN2A*, allowed us to reassess the clinical phenotype of patients with a suspect of CS/CISS1-like syndrome, respectively into CLIFAHDD, SHFYNG or EIEE11 syndromes, and although they present with overlapping features especially in infancy, such as feeding problems, contractures and temperature instability, they must be considered as distinct entities. These results showing a high neonatal phenotypic overlap among

CS/CISS1 with the other known syndromes will be very helpful for clinicians. Although clinical expertise allows in some cases to make a confident clinical diagnosis, adequate counselling should only be possible after the diagnosis has been confirmed by molecular genetic testing. Furthermore, the finding of variants in the imprinted *MAGEL2* locus shows the importance to consider also complex genetic mechanisms, such as genomic imprinting in search of pathogenic variants underlying Mendelian disorders.

Our finding of a pathogenic *de novo* variant in a particular region of the *CREBBP* gene, implicated in the Rubinstein-Taybi syndrome-1 (RSTS1) but associated to a different clinical phenotype, helped to establish genotype-phenotype correlations in this disorder, and to define it as a distinct clinical entity distinct from RSTS1, called Menke-Hennekam syndrome (MKHK1), to provide an adequate name for its identification by parents and specialists, thus facilitating appropriate clinical management and research.

Furthermore, the discovery of a novel *de novo* variant in *ANKRD11* causing a mild phenotype of KGB syndrome associated to a *SCN9A* variant causing GEFS+, offers an example of how WES has been instrumental allowing us to better dissect the clinical phenotype under study, which is a multilocus variation aggregating in one proband, rather than a phenotypic expansion associated with a single genomic locus, underscoring the role of multiple rare variants at different loci in the aetiology of clinical phenotypes making problematic the diagnostic path. The successful identification of the causal variant in a gene may not be sufficient, making it necessary to identify other variants that help fully explain the clinical picture. The prevalence of blended phenotypes from multiple monogenic disorders is currently unknown and will require a systematic re-analysis of large WES datasets for proper diagnosis in daily practice.

These results along with others reported in the literature, are contributing to reveal the extensive clinical variability and genetic complexity underlying Mendelian phenotypes and inheritance, to provide insight into study design and approach and analytical strategies and to identify novel mechanisms.

Our increasing knowledge on the genetic basis of rare disorders is shedding light on the “complex” nature of the “simple” Mendelian disorders and that “true monogenic” disorders are very rare. These complex relationships between Mendelian conditions and their associated genes and phenotypes underscore the current challenges of clinical diagnostics and discovery. The goal of identifying and characterizing the molecular architecture of Mendelian conditions would rely on the ability to detect with sufficient sensitivity and specificity the relevant types of variants and their

functional interpretation and will transform medical care for rare disorder patients and their families.

Introduction

Mendelian or monogenic disorders are the kind of genetic disorders in humans that arise from a disease-associated variant in a single gene. The term “Mendelian” is used because single-gene disorders run in families according to Mendel’s laws and can be dominant or recessive, and autosomal or sex-linked (Chial et al., 2008).

Variants causing Mendelian disorders are typically rare (minor allele frequency or MAF<1% in the population) since they tend to be negatively selected from the population owing to their extremely deleterious effects, and highly penetrant because nearly all individuals who carry a particular variant also express the associated phenotype. Therefore, Mendelian disorders are rare, with a prevalence, stated as the number of cases present in a population, that in the European Community has been defined as 5 cases per 10,000 people (https://www.eurordis.org).

The prevalence of genetic disorders varies widely between diverse geographical areas due to different population organization and structure, reproductive practices and environmental factors. According to a recently published research, rare disorders currently affect around 4% of the worldwide population, equivalent to a conservative estimate of 300 million people worldwide (Nguengang Wakap et al., 2019). They are characterized by a wide variety of illnesses and symptoms. Rare disorders are severe, often chronic, essentially disabling and the patients’ quality of life is affected by the lack or loss of autonomy due to the chronic, progressive, degenerative, and frequently life-threatening aspects of the disorder (https://www.rarediseaseday.org). The difficulties in access to treatment and care often results in heavy emotional, social and financial burdens on patients and their families (https://www.eurordis.org).

Rare disorders often go undiagnosed because patients, families and physicians have limited awareness of the condition, and rare disorder symptoms may not always be evident to healthcare providers who have never encountered it (https://www.globalrarediseasecommission.com). Besides, clinical variability and genetic heterogeneity can complicate the diagnosis, and common related symptoms can hide rare disorders, leading to misdiagnosis. This often results in the so-called “diagnostic odyssey”, a succession of steps, starting from the onset of symptoms to diagnosis of the

disorder which every person affected by a rare disorder goes through. According to an international online survey, approximately 40% of rare disorder patients received a misdiagnosis at least once, and the average time to an accurate diagnosis after symptoms have appeared is approximately 4.8 years (https://www.globalraredisordercommission.com; Engel et al., 2013).

Delays in diagnosis can lead to improper disorder management as well as disorder progression, and a misdiagnosis can lead to additional interventions later deemed to be inappropriate given the underlying disorder (https://www.globalraredisordercommission.com).

The field of rare disorders suffers from a deficit of medical and scientific knowledge. For a long time, doctors, researchers and policymakers were unaware of rare disorders, and until very recently, there was no real research or public health policy concerning issues related to the field (https://www.orpha.net).

Currently, two international databases curate clinical and genetic data for the community: Online Mendelian Inheritance in Man (https://www.omim.org/) and Orphanet (https://www.orpha.net/).

Online Mendelian Inheritance in Man® (OMIM®), is the primary repository of comprehensive, curated information on human genes and genetic phenotypes and the relationships between them, freely available and updated daily. This database was firstly published in 1996 by Dr Victor A. McKusick as a catalogue of Mendelian traits and disorders, entitled Mendelian Inheritance in Man (MIM). The online version, OMIM, was created in 1985 by a collaboration between the National Library of Medicine and the William H. Welch Medical Library at Johns Hopkins. At present (as October 25th, 2019), it has over 25,000 entries describing over 16,000 genes (of approximately ~19,000 protein-coding genes predicted to exist in the human genome) and 8,900 phenotypes. All entries in OMIM are given unique and stable MIM numbers. Genes and phenotypes are described in separate entries. Analysis of the discovery metrics at present shows that variants causing 5,470 Mendelian phenotypes have been identified in 3,790 genes (~20%). Genes underlying 1,555 (~8%) Mendelian phenotypes have been mapped but not identified. Genes underlying 1,749 (~9%) suspected Mendelian phenotypes had been not yet identified either mapped. Genes whose impact in humans has not yet been determined are ~12,000. Based on the analysis of knockout mouse models, loss of function variants in up to ~30% of genes (~6,000) could result in embryonic lethality in humans. Collectively, ~16,000 genes remain candidates for Mendelian phenotypes, with a rate of new genes associated with Mendelian conditions per year of about 300 (Chong et al., 2015).

Orphanet has maintained a register of both genetic and other rare disorders since 1997 and aims to collect epidemiological data on every rare clinical entity included in its nomenclature

(https://www.orpha.net/). While OMIM categorizes rare disorder on the basis of genetic etiology, Orphanet groups by clinically recognizable disorders. Furthermore, the Orphanet nomenclature of rare disorders is being annotated with the Human Phenotype Ontology (HPO) terms in order to allow for deep phenotyping of rare disorders in health records and registries. The HPO was initially published in 2008 (http://www.human-phenotype-ontology.org; Köhler, 2018) with the goal to provide a standardized vocabulary of phenotypic abnormalities encountered in human disorder, so enabling the integration of phenotype information across scientific fields and databases, providing comprehensive bioinformatic resources for the analysis of human disorders and phenotypes, offering a computational bridge between genome biology and clinical medicine.

Several genes responsible for rare Mendelian disorders have been identified beginning in the mid-1980s, and for the following two decades, by an approach to gene discovery that was a combination of linkage analysis, positional cloning and Sanger sequencing of candidate or mapped genes, most of which was hypothesis-driven (Boycott et al., 2017). However, factors such as the reduced availability of affected subjects and families to study, the heterogeneity of locus and the reduced reproductive fitness often limit the power of these conventional approaches (Ng et al., 2010). The subsequent introduction in 2009 (Ng et al., 2009) of hypothesis-free next-generation DNA sequencing (NGS), primarily whole-exome sequencing (WES) strategies provided an alternative method to identify genes associated with the disorder through the study of information contained within the human genome. This has resulted in increased identification of rare disorder genes, many of which were previously intractable to conventional gene discovery approaches.

The first monogenic disorder to be solved by WES was Miller syndrome, a multiple malformation disorder (Ng et al., 2010). Since then, WES became the primary technological approach for Mendelian disorder gene identification.

WES has the critical advantage of permitting the finding of both known and novel disorder-causing genes at the same time and considering that, in 85% of cases variants causing Mendelian disorders map within protein-coding sequences or in canonical splicing sites, currently it is the best approach for Mendelian gene identification (Choi et al., 2009).

For the first time, this technology has been proved to be successful in the identification of rare and extremely rare variants causative of the clinical phenotype under study (minor allele frequency; MAF <1% and < 0.5% respectively) also with small numbers of informative cases (one trio or few unrelated individuals with the same phenotype), which have been refractory to classical approaches. It is possible to study different patterns of inheritance (recessive, dominant/de novo,

X-linked) with the appropriated strategy of variant filtering (Gilissen et al., 2012). We can assume to look for bi-allelic variants for a recessive model (homozygous or compound heterozygous and the parents to be obligate carriers for the variant); heterozygous variant for a dominant model (only one parent obligate carrier for the variant or none in case of a *de novo* event); homozygous or hemizygous variant for an X-linked model. Genes for recessive disorders are, in general, more accessible to find than genes for dominant disorders because fewer genes in each exome are homozygous or compound heterozygous for rare non-synonymous variants (Gilissen et al., 2012).

The benefits of exome sequencing studies are clear (Bamshad et al., 2011):

- cost-effective and rapid strategy for analyzing nearly all the coding regions of the genome;
- unbiased sequencing approach, because it does not limit the analysis to a predefined region, so it is possible to identify genes acting in unexpected pathways or cellular processes;
- sample size required for a sufficiently powered study dramatically reduced (few phenotypically unrelated individuals, a single extended family, or a single parent and child);
- provides new insights into human gene pathways;
- crucial for drawing accurate genotype/phenotype correlations,
- facilitate clinical diagnostic, both in cases that present with atypical manifestations, or that require extensive or costly evaluation (high genetic heterogeneity), opening the way to a specific choice of treatment.

Although the advantages of sequencing the exome are evident, it also has some limitations. WES was designed to find variants in coding regions and UTR's such as single nucleotide polymorphisms (SNPs), small insertions and deletions (indels). Therefore, most of the non-coding variants escape the technology itself, along with others such as trinucleotide repeats and chromosomal rearrangements. Large deletions and duplications (Copy Number Variants or CNVs) are also challenging to be detected from exome data, although several tools have been developed. The four well known WES-based CNVs detection tools HXMM, CoNIFER, ExomeDepth, and CONTRA, use each an algorithm that has its strengths and weaknesses, different minimum and maximum CNV length detection, accuracy. Usually, the majority of the software is based on the analysis of the read depth detecting the number of fragments that mapped in defined genomic regions, and on the comparison between patient and controls data. However, these measures could be affected by errors caused by inaccurate capture of the region for capture kit problem or alignment problems so that they should always be confirmed by another technique (de Ligt et al., 2013).

The more significant challenge, however, is data interpretation. On average, exome sequencing identifies about 20,000 – 25,000 single nucleotide variants (SNVs), a number that can change in consideration of various factors such as different methods used, and ethnicity of samples. More than 95% of these variants are known polymorphisms in human populations. (Bamshad et al., 2011; 1000 Genomes Project Consortium, 2010).

In this context, the strategies to search causal alleles are different and take into account several factors (Bamshad et al., 2011):

- mode of inheritance of the trait;
- pedigree or population structure;
- *de novo* or inherited variant;
- extent of locus heterogeneity;
- allele frequency;
- biological function of each variant;
- analysis of network-disorders.

Another challenge is also to get a comprehensive clinical phenotype that can help to address the search towards a particular gene/pathway. Sometimes WES reveals mutations that help the clinicians to make the proper diagnosis, and this type of approach, genotype-driven is called "reverse phenotyping" (Schulze and McMahon 2004).

While taking into account the complexity of interpretation, the enormous potential of WES explains why it is becoming the first-choice method in laboratories dealing with molecular diagnostics, with an overall diagnostic rate in unselected, consecutive patients reported to be∼25%-40% (Dragojlovic et al. 2018). This variable range across different studies probably is due to several factors including sequencing approaches, analysis strategies, family history, type of disorder, age at presentation, variant interpretation (Ji et al., 2019).

Rationale of the work

Although substantial progress has been made toward identifying the genetic basis of Mendelian disorders, more than 50% of patients affected by a rare genetic disorder never receive a diagnosis, crucial to establish an adequate treatment and to identify novel knowledge-based therapeutic

targets to improve the overall outcome (Boycott et al., 2017). Diagnosis depends on gene discovery, and for Mendelian disorders, science can provide some answers.

The accurate understanding of the genetic bases of Mendelian disorders is hugely relevant for various reasons concerning 1) to formulate a precise diagnosis and therefore to provide the parents of the affected subjects with an explanation of the severe pathology of their children; 2) to offer, through an adequate genetic consultation, exact estimates of the risk of recurrence allowing the parents of these subjects to plan, in an informed manner future reproductive strategies; 3) to provide molecular tools for a possible prenatal diagnosis in future pregnancies where the parents would freely use this option; 4) to indicate potential therapeutic targets useful for the development of new drugs based on a full understanding of the pathophysiological mechanisms of each form of disorder and to create the conditions for approaches of personalized medicine.

For this purpose, multiple and integrated research programs are necessary, including the collection of clinically well-characterized patients and their families for genotype/phenotype correlations, genetic studies covering the whole genome/exome, genetic and biological validation of the potentially disorder-associated variants identified with functional characterization of the transcriptional and protein products of the candidate genes. Only through this primary research activity will it be possible to arrive at the subsequent transfer of the acquired knowledge to the diagnostic and therapeutic field. Since the first successful application for gene discovering in rare Mendelian disorders of unknown cause (Ng et al., 2010), WES turned out to be a powerful approach to detect high penetrance disorder-associated variants.

The rationale of this proposal is to contribute to the identification of the disorder causing genes by WES in molecularly unsolved cases with a family history of a rare disorder.

We focused our attention on a cohort of heterogeneous samples affected by the following disorders: Crisponi syndrome/Cold-induced sweating syndrome type 1 (CS/CSS1)-like phenotype, Epileptic Encephalopathies (EE) and syndromic Intellectual Disabilities (SID), Some samples are of Sardinian origin. Sardinia is a founder population in which the frequency of particular alleles can be higher than that in large populations, a cause of genetic drift, natural selection and bottleneck. Its population has already proved extremely useful for mapping genes implicated in rare disorders. Furthermore, the availability of a reference Sardinian genome database generated at the CNR Institute of Genetic and Biomedical Research (IRGB), from the sequencing of the entire genome of 3500 Sardinian individuals from different areas of the island (Sidore et al., 2015) can help to 1) discriminate specific Sardinian variants 2) make a more accurate estimate of the allele frequency;

3) identify sub-regions within the island where the frequency of carriers for a particular causal variant is higher, suggesting a possible founder effect.

The centres involved in the study have been the following:

1. Department of General Pediatrics, Muenster University Children's Hospital (Germany) for the collection and clinical phenotyping of CS/CISS1-like phenotype individuals.
2. Epilepsy Unit, Department of Neurology and Paediatrics Psychiatry, A. Cao Hospital, AOB, Cagliari (Italy) and Research group Epi-Genétic, Salvador, Bahia (Brazil), for the collection and clinical phenotyping of individuals with Epileptic Encephalopathies.
3. Institute of Child Neuropsychiatry, University of Sassari and AOS (Italy) for the collection and clinical phenotyping of individuals who have syndromic Intellectual Disability.
4. Institute for Genetic and Biomedical Research, National Research Council, Monserrato (Italy) for the generation of WES data and genetic analysis.
5. Research, development and higher education centre in Sardinia (CRS4), Pula (Italy) for the generation of WES data and bioinformatics analysis.

Mendelian disorders under study

Crisponi Syndrome/Cold-induced sweating syndrome type 1 (CS/CISS1)-like phenotypes

Text extract from Buers et al., 2019: "*Crisponi syndrome (CS, MIM#601378) was firstly described in Sardinian families and although some phenotypic overlap existed with other phenotypes, the syndrome was hypothesized to be a unique entity and a new syndrome.*[1] *Later on, Accorsi et al*[2] *and Nannenberg et al*[3] *reported further patients with an identical phenotype, and suggested to name the entity Crisponi syndrome. CS is a rare recessively inherited disorder with a high neonatal lethality. The principal clinical symptoms are recurrent periods of hyperthermia, camptodactyly, feeding and respiratory difficulties induced by abnormal paroxysmal contractions of the facial and oropharyngeal muscles. Some CS individuals survive the first year of life and often display a spontaneous improvement of feeding difficulties and hyperthermia.*[1,4] *However, these individuals develop scoliosis, sometimes mild psychomotor retardation as well as cold-induced sweating combined with massively elevated plasma noradrenaline (NA) levels in early childhood.*[5] *Cold induced sweating*

syndrome type 1 (CISS1) was initially described in two siblings at the age of 20 and 21 years.[6] They showed cold induced sweating at ambient temperatures lower than 20°C. A retrospective study revealed that most of the patients showed crises of hyperthermia, respiratory difficulties or contractions of facial and oropharyngeal muscles in infancy, typical symptoms of CS.[7] Scoliosis as well as other dysmorphic features are additional symptoms of this syndrome. Both, CS and CISS1 are caused by variants in the cytokine receptor like factor 1 (CRLF1) gene, coding for a ligand of the ciliary neurotrophic factor receptor (CNTFR).[4,8,9] Genetic and functional studies on altered CRLF1 indicated that CS and CISS1 are two occurrences of the same disorder. Thus, CS is considered as the infantile presentation of CISS1.[10,11] CS/CISS1 belongs to the family of "CNTFR-related disorders" with overlapping clinical features including cold-induced sweating syndrome type 2 (CISS2, MIM#610313) and Stüve-Wiedemann Syndrome (STWS, MIM#601559). CS/CISS2 is caused by variants in cardiotrophin-like cytokine factor 1 (CLCF1),[7,12,13] and STWS by variants in leukemia inhibitory factor receptor (LIFR).[14]

The combination of typical clinical CS/CISS symptoms and the detection of disorder-associated variants in CRLF1 or CLCF1 are the prerequisites for diagnosis. Alterations in CRLF1 and CLCF1 can be found in more than 60% of individuals with clinical diagnosis of CS/CISS, and a subset of CS/CISS cases (about 40%) remains yet genetically unexplained (CS/CISS-like phenotypes).

The differential diagnosis in CS/CISS has rapidly changed with the use of next generation-based sequencing techniques. Variants in genes others than those related to the family of "CNTFR-related disorders" have been recently found to be associated with a CS/CISS-like phenotype. In 2016, disorder causing variants in kelch like family member 7 (KLHL7) were detected in individuals with a CS/CISS-like phenotype associated with retinitis pigmentosa (RP).[15]"

Epileptic Encephalopathies (EE)

Epileptic encephalopathies (EE) include a large and heterogeneous group of severe epilepsies with genetic aetiology. In the last two decades, the so-called molecular revolution in medicine has also had an essential impact in the diagnosis and treatment of EE, and genetic research in this field is currently a highly suggestive and promising study frontier (Gobbi et al. 2014).

Epileptic Encephalopathies (EE) is a group of epilepsies typically at pediatric-onset and often severe, characterized by frequent and intractable seizures, severe disruption of brain electrical activity and cognitive arrest or regression, in which epilepsy in itself contributes to aggravate the cognitive and

behavioural deterioration of the subject. This category includes about 40% of epilepsies in the first three years of life (Nieh and Sherr, 2014).

The first genetic cause for a form of EE was first recognized in 2001, with the discovery of a *de novo* mutation in the *SCN1A* gene associated with Dravet's syndrome (Claes et al., 2001). Subsequently, with the advent of new genomic technologies, such as the CGH array (Comparative Genomic Hybridization), the NGS and in particular the WES, the number of mutated genes involved in EE has grown rapidly leading to the identification to date of more than 70 genes (Helbig and Tayoun, 2016). Recent studies show that the application of NGS techniques through the sequencing of specific multigenic panels for epilepsy and/or exomic sequencing allows to reach a diagnostic yield of 20-40% for EE (Chambers et al., 2016; EuroEPINOMICS-RES Consortium, 2017; Helbig et al., 2016; Kwong et al., 2015; Trump et al., 2016), increasing cost-effectiveness and significantly reducing the time of the diagnostic odyssey that families with EE encounter before receiving a definite diagnosis. Genetic mechanisms underlying epilepsies known to date include genomic rearrangements (ring chromosomes, translocations, monosomies and trisomies), copy number variants (CNVs; submicroscopic rearrangements, deletions or duplications involving one or more genes) and alterations of individual nucleotides that result in missense, frameshift or nonsense mutations.

However, despite the progress made in recent years, the aetiology of about 80% of EE forms remains unexplained, probably due to the high genetic and phenotypic heterogeneity. The emerging genetic architecture has led to a better understanding of the pathophysiology of EE syndromes, but still very limited. In the last decade, several pathophysiological mechanisms involved have emerged, including dysfunction of the ion channels, deficit in synaptic transmission, alterations of the mTOR signal pathway, chromatin remodelling and transcription regulation (von Deimling et al., 2017; Myers and Mefford, 2016; Epi4K Consortium et al., 2013). This led to the identification of different pharmacological targets for the various types of epilepsies. Therefore, it is critical to know the genetic cause in the patient in order to find the optimal treatment. Genetic mutations causing EE are mostly sporadic, i.e. resulting from *de novo* mutations in a single autosomal dominant expression gene, but there are also autosomal recessive EE forms and X-linked forms (Gobbi et al., 2014). Moreover, it has already been demonstrated that: similar mutations can be associated with different epileptic syndromes (variable expressivity), two individuals carrying the same genetic mutation have a different probability of developing epilepsy (incomplete penetration), and the same monogenic epileptic syndrome can be caused by mutations in different genes (genetic heterogeneity, Gardiner, 2006). These aspects explain why the relationship between genotype and

epileptic phenotype is not always linear and why within EE similar syndromic pictures may be associated with different gene mutations and, on the contrary, why pictures are not always linear.

Syndromic Intellectual Disability (SID)

Intellectual Disability (ID) until recently defined as Mental Retardation (MR), is a condition defined by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), as characterized by significant limitations in intellectual functioning and adaptive behavior, which include conceptual, social, and practical skills, arising "prior to age 18" (Chiurazzi and Pirozzi, 2016). It is not a single disorder with defined signs and symptoms, but it represents, especially in its syndromic forms, a wide variability of clinical phenotypes who have, as a common trait, a compromised intellectual capacity. People affected by ID need social support and specific educational services in order to achieve the skills necessary for the autonomy of daily life. Clinically it is defined by three main criteria (Van Bokhoven, 2011):

1. an intelligence quotient (IQ) below 70;
2. limitations in two or more adaptive behaviours, such as communication, self-care, and social skills;
3. age at onset of mental manifestations before 18 years old.

ID is the most common developmental disorder and clinically can be distinguished as syndromic or not in case is associated with other clinical findings such as dysmorphic features, multiple congenital abnormalities or metabolic defects as part of a syndrome. Together, the syndromic and non-syndromic forms of ID affect about 1% of the population (Maulik et al., 2011), with life-long care costs that can be extremely high along with dramatic repercussions at the emotional and social level of the affected and their families.

Although the development and functioning of the nervous system, prenatal, perinatal and postnatal, may be influenced by non-genetic factors (such as infections, trauma, exposure to neurotoxic substances), resulting in cognitive impairment, most severe forms of ID have a genetic origin. Although the percentage varies in different studies, it was found that hereditary forms can account for up to 65% of moderate and severe forms of ID (Van Bokhoven, 2011).

Based on our current knowledge, the genetic causes of ID include:

- chromosome rearrangements that result mainly in deleterious effects of the dosage gene (about 30% of cases, Topper et al., 2011);

- mutations and deregulation of genes or genomic regions subject to imprinting for the transcriptionally active allele variant depending on its parental origin (rare cases, like Prader Willi and Angelman syndromes);
- dysfunction of individual genes (monogenic causes of ID), which are individually needed for development of cognitive functions (about 10% of these are X-linked; Topper et al., 2011, Chelly et al., 2006).

So far, mutations have been identified in about 700-800 genes causing both non-syndromic and syndromic ID (Vissers et al., 2016, Chiurazzi and Pirozzi, 2016) and the functions of their respective proteins are very different. Although the number of involved genes has rapidly increased with the introduction of NGS technologies, still many ID genes are to be discovered, and although it is difficult to predict the number, an estimation of about 1,000 is likely (Vissers et al., 2016).

Recent studies confirm the importance of de novo mutations in larger cohorts with unexplained severe ID; therefore, the WES in this context is a necessary diagnostic test (Vissers et al., 2016).

The increase in the number of ID genes being identified, show extended and complex functional interaction networks converging into common molecular and cellular pathways, including neurogenesis, neuronal migration, synaptic functions, transcription and translation.

Additionally, Gene Ontology (GO)-based annotations of multiple biological processes in several studies revealed that ID risk genes are significantly associated with nervous system development, RNA metabolism, and transcription, presenting convergent functional features in specific biological pathways that reflect the complexity underlying the development of a functional brain (Kochinke et al., 2016, Liu et al., 2018).

There is growing evidence that mutations in many of the known ID genes give rise to a high degree of clinical variability, that can be a reflection of the type of mutation or other genetic and environmental factors, and to the association with a variety of different Cognitive Disorders (CDs). There is a high co-morbidity commonly observed between ID and other CDs, such as autism (ASD), attention deficit hyperactivity disorder (ADHD), depression and other behavioural problems. This phenotypic overlap between ID and various CDs is mirrored at the genetic level, all supporting the notion that they share a common molecular aetiology.

Genetic studies are showing the increasing inheritance complexity in ID, and genotype-phenotype correlation studies indicate that phenotypes are only rarely explained entirely by a mutation in a single gene, and other variations may affect the expressivity and penetrance of the disorder.

To date, more than 800 genes are known to be involved in the pathogenesis of syndromic and non-syndromic conditions with ID (Chiurazzi and Pirozzi, 2016), and the functions of their respective proteins are very different. Although the number of ID genes is increasing rapidly, currently the genetic aetiology of 60% of ID remains unexplained (Topper et al., 2011), and the majority of patients remains without a molecular diagnosis.

Methods

Patients' selection, samples acquisition and DNA extraction

We have selected seven families with different rare disorders to be enrolled in this project, aiming to discover new genes involved in their pathogenesis by applying the WES approach. The selected families were affected by:

- Crisponi syndrome/Cold-induced sweating syndrome type 1 (CS/CISS1)-like phenotypes; five families of various ethnicities, were recruited in collaboration with Prof. Frank Rutsch from the Department of General Pediatrics, Muenster University Children's Hospital, Muenster (Germany).

Text extract from Angius et al., 2019: "*Five individuals with the initial diagnosis of CS/CISS but negative for mutations in CRLF1 and CLCF1 were clinically assessed by the respective clinical geneticists and selected for WES analysis. Clinical findings of these cases are listed in Table 1. Our inclusion criteria comprised features related to the neonatal phenotype and especially to the four main typical CS criteria: hyperthermia in the first months of life, feeding difficulties, contraction of oropharyngeal muscles and camptodactyly. We assigned the five cases to rank 1 (very likely) and 2 (questionable) based on their phenotype similarity to CS/CISS. Rank 1 individuals (CS_239 and CS_306) presented in the neonatal period the four main typical CS criteria, while rank 2 (CS_125, CS_141 and CS_207) individuals fulfilled two or three of them (See clinical details and WES analysis, Supporting Information, Table S1).*"

- Epileptic Encephalophaties (EE); one Sardinian and one Brazilian family, recruited in collaboration with Dr. Pruna from the Epilepsy Unit, Department of Neurology and Paediatrics Psychiatry, A. Cao Hospital, AOB, Cagliari (Italy) and Dr. Alves from the Research group Epi-Genétic, Salvador, Bahia (Brazil).

Sardinian cases - Text extract from Angius et al., 2018: "*Two siblings of a Sardinian family presenting with dysmorphic facies, hypotonia, psychomotor retardation, epilepsy, absent speech, sleep disturbance, hyperkinetic movement disorder, cachexia and chronic constipation. Both siblings experienced generalized seizures with age at the onset of 4 and 6 years. They present a disorganized electroencephalogram (EEG), and bi-frontal and diffuse high amplitude sharp/slow waves are evident. Particularly, male proband EEG has a high frequency of epileptic discharges. Despite a recent moderate improvement, his electric activity is much more impaired than that in his sister. Furthermore, he has weekly seizures, mainly during sleep, and no response to anti-epileptic drugs (AEDs), while AEDs treatment was efficient in the female, with lack of seizures. Clinical and EEG pictures of our cases are very similar to those described for UNC80- mutated cases (Figure 1). The dysmorphic features are triangular face, bi-temporal narrowing, high nasal bridge, downslanting palpebral fissures, posterior rotated low set-ears, large and persistently opened mouth, and long fingers. Since the first months of life, they showed severe generalized hypotonia, but they could maintain in sitting position without support. They are unable to develop speech and present severe intellectual disability and sleep disturbances. Hyperkinetic movement and autistic spectrum disorder are evident*".

Brazilian case - Text extract from Alves et al., 2019: "*Female 14.5 years old. Delivered full term, by emergency C-section, due to lack of fetal movement, weighing 2.971 Kg, 48 cm long and 33 cm head circumference. Neonatal period had no complications. From the neonatal period onward, she showed difficulty to breastfeed, with low weight gain. At 9 months old, she had mioclonus-atonic type seizures with sudden falling of the head and trunk. Initially precipitated by fever, these seizures became afebrile and daily, several times a day, and were controlled after substituting phenobarbital for sodiumvalproate (VPA), in low doses. The EEG tests initially showed focal spikes (centro-temporal regions) and only at 4 age, one EEG test showed a theta rhythm (4-5 Hz) in the temporo-occipital regions (T5-O1; T6-O2). At the age of 4 years and 8 months, after remission of seizures for 3 years, and normal EEG tests, VPA was suspended. Starting from 6 years of age, the EEG tests showed persistence of several bursts of irregular generalized polyspike-wave (PSW) and spike-wave discharge (SW), lasting 1–3 s. (Fig. 1a-p). Despite persisting abnormal EEGs, patient has not presented relapse of seizures and is not on medication. Patient presented with recurrent otitis episodes and developed conductive hearing loss in left ear. A computed tomography scan of the mastoid showed signs of otomastoiditis in the left ear with obliteration of Prussak's space and cholesteatomatous process. Orthodontic evaluation conducted at 8 years of age showed*

dolichofacial pattern, maxillary protrusion, absence of lip seal, delayed eruption of permanent teeth, besides size increase of upper central incisors, with extra mamelar structures and whitish material of incisors and other teeth, compatible with hypoplasia (Fig. 2a, b). Cone-beam computed tomography of right oral lower-posterior region at 14.5 years of age, revealed dental units partially erupting and the presence of mixed-aspect images located between the dental roots, suggesting bone dysplasia (Fig. 2). The skeletal X-ray assessment showed inversion of physiological cervical lordosis (Fig. 2c); deviation of left dorsal axis, accen- tuated thoracic and lumbar lordosis and concealed spina bifida at L5/S1 (Fig. 2f). The proband has also shortening of the distal phalanx of the 5th finger, clinodactyly of the 2th and 5th (Fig. 2d, e); myopia; bifid uvula with submucous cleft palate; weight and height growth curve below percentile <5. Neuro- psychological analysis at age 8 showed IQ of 73."

- Syndromic Intellectual Disability (SID); one Sardinian family, recruited in collaboration with Prof. Sotgiu and Dr. Serra from the Institute of Child Neuropsychiatry, University of Sassari and AOS (Italy).

Text extract from Angius et al., 2019: "*The proband II.4 is 17-year-old boy born to a Sardinian family with unrelated parents, and two older healthy siblings (Figures 1 and 2a). The mother had one miscarriage at the fourth month of pregnancy (II.1). None of the family members has an overweight. Pregnancy was complicated by threatened miscarriage at 16 weeks. He was born at full term, Apgar scores were 4 and 8 after 1 and 5 minutes, and soon after birth started to show hypoglycemia, hypotonia, bradypnea, and bradycardia. Weight at birth was 3,395 g (50thP), length was 49 cm (50thP), and head circumference 35 cm (50thP). He had a cleft palate that was surgically corrected at 5 months of age. Growth parameters at 6 months were below 3thP. Several dysmorphisms were noticed (Figure 1): a round face, narrow forehead, strabismus, epicanthi, hypertelorism, curly eyelashes, sparse eyebrow, broad nasal bridge, small mouth, small chin, and low-set ears. No dental anomalies were present. His hands were small and broad with short and proximally implanted thumbs and bilateral 5th finger clinodactyly. His genitalia were underdeveloped. He was hypotonic and had a diffuse joint laxity and bilateral clubfeet. Skeletal radiographs showed normal bone age, a pectus excavatum, lumbar lordosis, hip dysplasia, valgus knee with a left discoid meniscus, and bilateral sandal gap. His initial development was normal, language development was regular, and he was able to stand up and walk independently at 13 months, although with some clumsiness and frequent falls. At 18 months, he started to gain weight and develop an over-eating behavior. At evaluation at 17 years, his weight was 120 kg (>97thP), height 160 cm (<5thP), and head circum-*

ference 60 cm (>97thP). Formal evaluation of cognitive functioning showed mild intellectual disability (Wechsler Intelligence Scale for Children—III: QIT 70; QIV 74; QIP 72) and learning difficulties. MRI showed enlargement of the cisterna magna and ventricular system, sub-ependymal areas of nodular heteroto- pia, a wavy-look corpus callosum, and a mild contrast enhancement of the pituitary stalk (Figure 1c). EEG showed no signs of electrical seizures".

All the individuals enrolled in the study (probands and parents where available) signed an informed consent approved by the Ethics Committee of the referent clinical centres drawn up according to current legislation.

Blood samples (or already extracted DNA) from these individuals were pseudonymized and sent to the IRGB-CNR Institute in Cagliari (Italy) and processed for DNA extraction using a standard salting-out protocol. Red blood cells lysis and separation of a white pellet of white blood cells were obtained by osmotic lysis. The cell pellet was resuspended in a lysis buffer, SDS 10% and protein kinase K. At the end of incubation, NaCl 6M was added and vortexed to precipitate the proteins. The supernatant was transferred into a new Eppendorf tube containing isopropanol and DNA was harvested with a pipette tip and dipped in a 70% ethanol Eppendorf tube. Finally, the tubes were centrifuged to pellet down the DNA. Supernatant was discarded and DNA air-dried. All the samples were resuspended in TE 1-0,1 mM. Samples underwent electrophoresis in 1% agarose gel to evaluated DNA degradation, and concentration/purity of double-stranded DNA was measured in Qubit 2.0 fluorometer using BR dsDNA Assay Kit (Invitrogen by life technology).

Whole-Exome Sequencing workflow

Initially we exome sequenced the samples using the TruSeq Exome Library Prep Kit (Illumina) and later with the Nextera Expanded Exome Enrichment Kit (Illumina) when this replaced the previous one. Both contain 62 Mb of genomic content, including exons, UTRs and miRNA (> 340,000 95mer probes to enrich about 200,000 exons spanning 20,794 genes). The Nextera kit allowed the preparation of the genomic DNA library and the exome capture by hybridization in just three days instead of 1 week, with an input quantity of DNA of 50 ng instead of 1 µg.

In general, with all the kits in commerce, the exome library preparation consists of: fragmentation, repair of the blunt-ended, adenylation, ligation of the adapters, a first step of PCR, hybridization and capture with biotinylate exome probes created in target regions, precipitation with the use of streptavidin beads, and a second step of PCR; all interspersed by various purification steps. Adapters contain unique index sequences ligated to sample fragments so allowing the pool of more samples

and the unique identification of the sequence reads during downstream analysis. Furthermore, with the Nextera kit, the fragmentation step is enzymatic rather than mechanic, so being more time-saving. The last step is the validation of the enriched library performed by the Agilent 2100 BioAnalyzer with DNA 1000 Kit or High Sensitivity DNA Kit, and Qubit 2.0 using BR dsDNA Assay Kit. Libraries were loaded on the cBot System (Illumina) to create clonal clusters on the flow cell and then sequenced on the HiSeq 2000 Instrument and analyzed with the Illumina extraction pipeline, that consists in a base-calling and imaging analysis performed locally by RTA software. The HiSeq 2000 uses the Sequencing By Synthesis (SBS) method: a massively parallel sequencing of short reads using solid-phase sequencing by reversible terminators, able to generate about 600-700 Gb of sequence from over 1 billion clusters in the form of 2×100 base reads (paired-end sequencing) from 2 flow cells in about 11 days with meagre error rates (< 1%). We obtained a medium coverage of about 80X for each sample, and we were able to run 24 samples in a single flow cell.

We carried out the WES sequencing at the NGS Facility in Pula, directly connected with a high-performance computing centre and a fully automated infrastructure to support the bioinformatics analysis of sequencing data. Furthermore, an optimized variant calling pipeline for WES was developed in collaboration with the Center for Advanced Studies, Research and Development in Sardinia (CRS4). These analyses have been conducted in collaboration with Dr Fotia's team at the CRS4 Biosciences research and development sector.

Sequence analysis and variant calling pipeline

Demultiplexing analysis and generation of FastQ files for the variant calling analysis were implemented in Orione, a Galaxy-based framework consisting of publicly available research software and specially designed pipelines to build complex, reproducible workflows for next-generation sequencing data analysis (Cuccuru et al., 2014). It is part of an ongoing project at CRS4 to integrate Galaxy with Hadoop-based tools to provide scalable computing and a specialized version of OMERO to model biomedical data and the chain of actions that connect them. Galaxy is used to manage all workflow-based operations, while a custom "Automator" daemon is used to execute and monitor workflow progress from the HiSeq instrument to the FastQ files and to link workflows to each other. One of the main advantages of this approach is that Galaxy tracks all operations with its histories.

In details, the paired-end sequence reads were aligned to the human genome (hg19) with the BWA-MEM aligner (BWA v.0.7.5) (Li, 2013). SAMtools were used for converting (SAM/BAM), sorting and

indexing alignments. Quality metrics and coverage statistics were calculated with FASTQC and Picard tools. First mappings were processed using the GATK framework (GATK v.2.8.1) (McKenna et al., 2010) according to their Best Practices recommendations. Briefly, reads were locally realigned around known insertion and deletion (INDELs) sites. Those zones are hard to mapping correctly so is necessary this step to correct errors made from the first alignment. Then, PCR and optical duplicates were marked with Picard tools (v.1.119) and excluded from downstream analysis. Finally, raw base quality scores were recalibrated to adjust for quality variation due to machine cycle and sequence context, and Single Nucleotide Polymorphisms (SNPs) and insertions/deletions (INDELs) were identified using GATK Unified Genotyper. Highly accurate variant calls are identified with GATK hard filters. Variants were classified as known or novel variants based on dbSNP146. Different tools [GATK Variant Annotator, SnpSift and SnpEff (Cingolani et al., 2012), KGGSeq (Li et al., 2012), SilVA (Buske et al., 2013)] were used to annotate variants with position in UCSC, RefGene, Ensembl and GENCODE transcripts, OMIM, ClinVar and DDD annotations, potential false positive signals as described in Fuentes Fajardo et al., 2012, for maternally or paternally imprinted genes (Joshi et al., 2016), presence of pseudogenes, allele frequency in dbSNP141, 1000 Genome Project, ESP6500, ExAC, Evade v.5 (an internal exome database of ~500 samples) and in a Sardinian population database derived from whole genome sequencing (WGS) of 3,500 samples (Sidore et al., 2015); sequence conservation, predicted pathogenicity according to different models [non-synonymous variants: CADD, SIFT, Polyphen2, LRT, MutationTaster, MutationAssessor and FATHMM retrieved from dbNSFP v. 2.9 (Liu et al., 2013); synonymous variants: SilVA], and evidences from literature automatically retrieved from PubMed by co-citation of keywords of interest and the genes in which the variants are located.

Sequence data were analyzed with CoNIFER software (Copy Number Inference From Exome Reads; Krumm et al., 2012) for the discovery of CNVs in all sequenced patients. CNV calls were annotated with CLINGEN (https://www.clinicalgenome.org/) and DGV databases (http://dgv.tcag.ca/dgv/app/home).

These analyses have been conducted in collaboration with Dr Fotia's team at CRS4.

Biological interpretation and identification of causal alleles

On average, exome sequencing identifies ∼24,000 single nucleotide variants (SNVs) in African American samples and ∼20,000 in European American samples of whom more than 95% are known polymorphisms in human populations (Bamshad et al., 2011). Prioritizing variants in WES studies of

Mendelian disorders to finally identify the causal allele can be done at different levels, and it is very challenging.

1. Variant-gene level:

1.1 Frequency: in rare severe Mendelian disorders, causal variants generally show complete penetrance and are rare, and therefore not present in public databases. Ideally, individuals in the reference population should have the same origin as the sample under investigation. However, because local population frequencies are not always available, public databases are used. Common variants deposited in public databases (1000 Genomes Project, ESP6500, ExAC, and dbSNP) are filtered out according to an adjustable allele frequency threshold. As most of the disorders under study are relatively uncommon and ideally caused by rare pathogenic variants, a filtering exclusion based on a minor allele frequency (MAF)>1% is well powered (Bamshad et al., 2011). However, caution is necessary given that some of the databases also include patients. Furthermore, an in-house database of about 500 exomes (Evade v.5) available at CRS4 provides a useful filter to remove additional variants, with a 75% increase of the filtering power, for frequency or platform/laboratory artefacts.

1.2 - Gene features: as severe Mendelian disorders are more likely to be caused by non-synonymous mutations (missense, nonsense, splice-site defects, frameshift and non-frameshift indels), focusing only on this functional class of variants narrows down the number of candidates. Because not all non-synonymous variants contribute equally to affect function of coded proteins, a predicted effect on protein function and on whether it is potentially disorder-casual is challenging especially for missense variants.

2. Genetic level:

2.1 Pedigree information: affected family members usually share the genomic segment harbouring the causal mutation(s).

2.2 Mode of inheritance of disorder can also be used to exclude impossible disorder-causal variants effectively. We can assume that there are bi-allelic variants for a recessive model (homozygous or compound heterozygous with the parents obligate carriers for the variant); heterozygous variant for a dominant model (only one parent obligate carrier for the variant or none in case of a *de novo* event); or hemizygous variant for an X-linked model.

3. Knowledge level: disorder-network analysis added a complementary layer of information to the sequencing data. It predicted role in a biological pathway and interaction with genes or proteins that were known to cause a similar phenotype and could be used at different steps of analysis. A

primary detailed network analysis of the gene(s) harbouring variants genetically and biologically validated facilitated the clarification of their putative role in the pathogenesis of the syndrome. It could be expected that disruption of the same molecular network could have a similar phenotypic effect in general.

Genetic validation of sequence variants

Each sequence variant identified during the screening phase as a potential candidate in the pathogenesis of the disorders understudy, has been genetically validated by traditional Sanger sequencing and disorder-segregation analyses in the family. All family members were sequenced by using the sequencer ABI Applied Biosystem 3730 DNA Analyzer, and all the primers were designed manually or with the program Primer3. PCR was performed in 25µl of the total volume and 10 to 50 ng of DNA for problematic samples (mainly degraded samples), 1X buffer, MgCl2 2 mM, 0,12 µM of primers, 0,4 U of Taq Polymerase and 0,2 mM of each dNTPs were used. Five µl of Betaine 5M were added to the mix to amplify problematic GC abundant sequences. PCR purification was done using Exosap according to manufacturer's protocol. Big Dye v1.1 and relative Buffer 5X were used for sequencing the PCR fragment. Then the sequence reaction was precipitated with 95% ethanol, sodium acetate 3M pH 5.4.

Results

Crisponi syndrome/Cold-induced sweating syndrome type 1 (CS/CISS1)-like phenotypes

Text extract from Angius et al., 2019: "*Here, we identified two de novo missense variants in NALCN NM_052867.3: c.1800C>A: p.(Asp600Glu) and c.1571G>A: p.(Ser524Asn) in two unrelated CS patients (Family A and C). These novel variants predicted as pathogenic by LRT, MutationTaster, Polyphen2 and FATHMM, occurred at highly conserved amino acids within NALCN from different organisms, even those very distinct from vertebrates. NALCN forms a voltage-independent, nonselective, non-inactivating cation channel permeable to Na+, K+, and Ca (2+). It is responsible for the neuronal background sodium leak conductance.*[6] *Functional testing of some human variants in Caenorhabditis elegans demonstrated that CLIFAHDD can be caused by dominant loss or gain of*

*function mutations in ion channel function. In 2016, Bend et al[7] conducted functional studies on C. elegans NCA-1 mutants carrying the p.(Asp647Glu) mutation, which is the orthologous position of p.(Asp600Glu) in human NALCN. They described this de novo variant as a gain of function, which gives very severe effects. In animal model, this mutation displays dramatically reduced locomotion, small body size and curly posture, consistent with neuronal dysfunction. In family B and E, we identified two MAGEL2 mutations of paternal origin: a de novo NM_019066.4: c.2056_2066del: p.(Trp686Alafs*23), already reported, and an insertion c.1996dupC: p.(Gln666Profs*47): rs770374710 which represents a mutational hotspot.[8] In family D, we identified a de novo mutation in the SCN2A gene, NM_001040143.1: c.2567G>A: p.(Arg856Gln): rs797045942, which has been previously reported as associated to Ohthara syndrome. This substitution results in the conversion of an arginine to glutamine. This arginine is highly conserved within SCN2A from different organisms, even those very distinct from vertebrates. It is predicted to be pathogenic by five popular algorithms, SIFT, Polyphen2, LRT, Mutation Assessor, and FATHMM.*

In summary, heterozygous mutations in NALCN, MAGEL2 and SCN2A can result in a CS/CISS-overlapping phenotype, which seems to be most similar in the neonatal period. Although CS/CISS, CLIFAHDD, SHFYNG and EIEE11 syndromes show overlapping features such as feeding problems in infancy, contractures and temperature instability, they must be considered as distinct entities. These observations suggest that similar pathophysiological mechanisms may lead to such clinically overlapping phenotypes.

Based on the considerable phenotypic overlap between CS/CISS and other syndromes described in this article, the clinical geneticist should employ the current clinical diagnostic criteria for CS/CISS with caution, especially in the neonatal period. Adequate counseling should only be possible after the diagnosis has been confirmed by molecular genetic testing. Therefore, sequencing analysis of these genes has to be considered for those cases with a suspected CS/CISS during neonatal period who were tested as mutation negative in the known genes, because an expedited and corrected diagnosis can improve patient management and can provide a specific clinical follow-up".

Epileptic Encephalopathies (EE)

Text extract from Angius et al., 2018: "*Infantile hypotonia with psychomotor retardation and characteristic facies-1 (IHPRF1, MIM#615419) is a severe autosomal recessive neurologic disorder with onset at birth or in early infancy. It has recently been found to be caused by mutations in the NALCN gene that encodes a voltage-independent, cation channel permeable to Na+, K+ and Ca2+.1*

NALCN forms a channel complex with 2 other proteins, UNC80 and UNC79, involved in the folding, stabilization, cellular localization and activation of NALCN. This channel complex is mainly expressed in the central nervous system (CNS) and it plays a crucial role in regulating the resting membrane potentials and neuronal excitability.[2] Recessive mutations in the UNC80 gene are associated to the IHPRF2 syndrome (MIM#616801) with a clinical phenotype very similar to IHPRF1.[3] UNC79 so far has no corresponding human phenotype.

Only 11 cases from 4 families have been reported in literature as affected by IHPRF1 with homozygous mutations in NALCN.[1,4,5.] We report on 2 additional cases, siblings of a Sardinian family presenting with dysmorphic facies, hypotonia, psychomotor retardation, epilepsy, absent speech, sleep disturbance, hyperkinetic movement disorder, cachexia and chronic constipation"…"Cytogenetic analysis showed a normal karyotype for both patients. Array-based comparative genomic hybridization (CGH) analysis, made using commercially available oligonucleotide microarrays containing about 60-mer probes (Human Genome CGH Microarray 244A Kit; Agilent Technologies Santa Clara, USA), and subsequent next generation sequencing (NGS) analysis using an Early Infantile Epileptic Encephalopathy panel (ALDH7A1, ARHGEF9, ARX, ATP1A2, CACNA1A, CDKL5, CNTNAP2, EFHC1, FOXG1, GABRG2, GRIN2A, GRIN2B, KCNJ10, KCNJ16, KCNK18, KCNQ2, MAGI2, MECP2, MEF2C, NRXN1, PCDH19, PLCB1, PNKP, PNPO, POLG, SCN1A, SCN1B, SCN2A, SCN9A, SLC25A22, SLC2A1, SLC9A6, SPTAN1, STXBP1, UBE3A) performed according to the Roche FLX Titanium protocols, failed to identify any pathogenic variant.

Whole-exome sequencing (WES) was performed for all family. Initially, the father was not available for testing and he was recruited subsequently after the WES findings. We identified 2 truncating mutations in the NALCN gene (NM_052867); chr13:101733940G/A c.3823C>T p.(Arg1275) (rs569371758), maternally derived with a gnomAD reported frequency of 0.000008134 and chr13:101759921_101759922insTATGA c.2495_2496insTCATA p.(Phe833Hisfs*40), hypothetically paternally derived, never reported so far. Sanger sequencing confirmed both variants in probands and parents allowing assessing the clinical phenotype to IHPRF1. This is the fifth family reported worldwide, and the first European case with IHPRF1 syndrome. Among all cases, epilepsy is reported in 7/11, but EEG characteristics have never been described.*

Although the number of cases with NALCN recessive mutations reported so far is small, we propose the inclusion of such gene in the targeted re-sequencing gene panels usually employed in the diagnostic setting to identify gene defects associated to epilepsy, also in light of the relevant implications in diagnostic workup and therapeutic opportunities.".

Text extract from Alves et al., 2019**: "***KBG syndrome is a very rare autosomal dominant disorder, characterized by macrodontia, distinctive craniofacial findings, skeletal findings, post-natal short stature, and developmental delays, sometimes associated with seizures and EEG abnormalities. So far, there have been over 100 cases of KBG syndrome reported.*

Here, we describe two sisters of a non-consanguineous family, both presenting generalized epilepsy with febrile seizures (GEFS+), and one with a more complex phenotype associated with mild intellectual disability, skeletal and dental anomalies. Whole exome sequencing (WES) analysis in all the family members revealed a heterozygous SCN9A mutation, p.(Lys655Arg), shared among the father and the two probands, and a novel de novo loss of function mutation in the ANKRD11 gene, p.(Tyr1715), in the proband with the more complex phenotype. The reassessment of the phenotypic features confirmed that the patient fulfilled the proposed diagnostic criteria for KBG syndrome, although complicated by early-onset isolated febrile seizures. EEG abnormalities with or without seizures have been reported previously in some KBG cases. The shared variant, occurring in SCN9A, has been previously found in several individuals with GEFS+ and Dravet syndrome.*

This report describes a novel de novo variant in ANKRD11 causing a mild phenotype of KGB syndrome and further supports the association of monogenic pattern of SCN9A mutations with GEFS+. Our data expand the allelic spectrum of ANKRD11 mutations, providing the first Brazilian case of KBG syndrome. Furthermore, this study offers an example of how WES has been instrumental allowing us to better dissect the clinical phenotype under study, which is a multilocus variation aggregating in one proband, rather than a phenotypic expansion associated with a single genomic locus, underscoring the role of multiple rare variants at different loci in the etiology of clinical phenotypes making problematic the diagnostic path. The successful identification of the causal variant in a gene may not be sufficient, making it necessary to identify other variants that fully explain the clinical picture. The prevalence of blended phenotypes from multiple monogenic disorders is currently unknown and will require a systematic re-analysis of large WES datasets for proper diagnosis in daily practice".

Syndromic Intellectual Disability (SID)

Text extract from Angius et al., 2019: "*Analysis of WES data yielded the de novo CREBBP variant [NM_004380, c.5170G>A; p.(Glu1724Lys)] in the pro- band, confirmed by Sanger sequencing (Figure 2b and Supporting Information Table S1). No putative causal variants were found in EP300. The*

Glutamic acid (E) residue at position p.1724 of CBP is located in a conserved residue in the zinc finger 2 domain (ZNF2, ZZ- type, aa 1,701–1,744) which contains important cysteine residues that mediate Zn2+ binding (Figure 2c). Molecular dynamics analyses demonstrated that, within the 3D structures of ZZ-domain of CBP, the mutated amino acid (Glu1724Lys) is exposed to the solvent and it is in a flexible loop, in particular, it is adjacent to the second conserved, invariable cysteine residue (Cys1723) in the Cys-X-X-Cys motif of the second cluster of ZNF2, indispensable for coordination of one of the two zinc ions (Figure 2c,d). The change to Lysine (A) is predicted to be "damaging," "deleterious," or "disease causing" in all functional prediction models and is associated with a very high CADD score of deleterious mutation prediction of 54 (Supporting Information Table 2). This variant affects the last amino acid of exon 30 where few pathogenic missense mutations have been observed (Menke et al., 2018, 1998). The present report adds a further patient to the list of the 22 reported so far with variants in the last part of exon 30 and the beginning of exon 31, not associated with a classical RSTS phenotype. The 25 cases with the new entity do not show the striking facial dysmorphisms not the characteristic hand and foot findings of RSTS, although some clinical features are shared such as intellectual disability and short stature."…"The previous reported cluster of variants associated to the new phenotype discussed here, is located into the ZNF2 and ZNF3 domains, containing cysteine residues that mediated ion zinc binding to stabilize helical folding and mediate interactions with transcriptional regulatory proteins (Menke et al., 2018, 1998). Our molecular dynamics analyses demonstrated that the mutated amino acid (Glu1724Lys) is exposed to the solvent and it is in a flexible loop, in particular, it is adjacent to the second invariably conserved cysteine residue (Cys1723) in the Cys-X-X-Cys motif of the second cluster of ZNF2, indispensable for coordination of one of the two zinc ions. This suggests that the difference in the electro- static potential (polar acidic vs. polar basic) caused by this amino acid replacement could actually play an important role in modifying protein–protein interactions and explain the observed phenotype.

The phenotype of the here presented patient is unusual because of his remarkable obesity which is not segregating in this family. Three of the 22 reported patients show obesity as well (Menke et al., 2018, 1998). More individuals need to be reported to determine whether marked obesity can be part of the phenotype. The present brain findings are also unusual, only the enlarged ventricles have been reported before (Fergelot et al., 2016).

In the 2016 article of Menke et al. (Fergelot et al., 2016), it is suggested that the region between the last part of exon 30 and the beginning of exon 31 of CREBBP is associated with two different phenotypes: the proximal one (as the patient reported here) between bp 5,128 and 5,594 (aa 1,710–

1864) of CREBBP, in which the present patient's variant is located, shows a less specific phenotype, and a more distal one, between 5,595 and 5,614 (aa 1865–1872), in which variants cause a more specific and clinically recognizable phenotype. The phenotype of the present patient is in accordance with this suggested subdivision. He does share the typical clinical characteristics of this new phenotypic entity, such as intellectual disability, short stature, feeding problems, and other infrequently findings such as cleft palate, hip dysplasia, sandal gaps, and cryptorchidism. The present patient adds as unusual manifestations marked overweight and cerebral heterotopias. However, the full variability of the phenotype(s) will only become evident when more patients will have been reported in international literature."

Recently, after the publication of our work, Banka et al., 2019 proposed to call this disorder "Menke-Hennekam syndrome" (MKHK1; MIM#618332) to establish it as a clinical entity distinct from RTS and to provide a satisfactory name for adoption by parents and professionals, thus facilitating appropriate clinical management and research.

Discussion

The results reported here, along with others reported in the literature, are contributing to reveal the extensive clinical variability and genetic complexity underlying Mendelian phenotypes and inheritance, as well as to provide insight into study design and approach and analytical strategies and to identify novel mechanisms (Chong et al., 2015).

The discoveries of pathogenic variants in *NALCN, MAGEL2* and *SCN2A,* allowed us to reassess the clinical phenotype of patients with a suspect of CS/CISS1-like phenotype respectively into CLIFAHDD, SHFYNG or EIEE11 syndromes, and although they present with overlapping features especially in infancy, such as feeding problems, contractures and temperature instability, they must be considered as distinct entities. In particular, our results disclosed both recessive and dominant pathogenic variants of human NALCN associated complex neurodevelopmental syndromes. Recessive variants of *NALCN* are linked to the IHPRF1 (MIM#615419) syndrome, while dominant variants of *NALCN* are responsible for the CLIFAHDD (MIM#616266) syndrome. Both IHPRF1 and CLIFAHDD patients exhibit complex clinical traits of variable severity that may cause premature death in some cases. Notably IHPRF1 and CLIFAHDD patients present with shared symptoms such

as facial dysmorphisms, hypotonia, global developmental delay, constipation, and respiratory defects (Al Sayed et al., 2013; Chong et al., 2015).

All these results showing a high neonatal phenotypic overlap among CS/CISS1 with the other known syndromes will be very helpful for clinicians. Although clinical expertise allows in some cases to make a confident clinical diagnosis, adequate counselling should only be possible after the diagnosis has been confirmed by molecular genetic testing.

The finding of variants in the imprinted *MAGEL2* locus shows the importance to consider also complex genetic mechanisms, such as genomic imprinting in search of Mendelian disorder-associated variants.

Our discovery of a disorder-associated variant in a particular region of the *CREBBP* gene, implicated in the Rubinstein-Taybi syndrome-1 (RSTS1; MIM#180849) but associated to a different clinical phenotype, contributed to establish genotype-phenotype correlations in this disorder, and to define it as a distinct clinical entity distinct from RSTS1, called Menke-Hennekam syndrome, (MKHK1; MIM#618332) to provide a suitable name for adoption by parents and professionals, thus facilitating appropriate clinical management and research. So far, only 26 cases have been reported in the literature (Menke et al., 2016, 2018; Angius et al., 2019; Banka et al., 2019).

Furthermore, the discoveries of a novel *de novo* variant in *ANKRD11* causing a mild phenotype of KGB syndrome associated to a *SCN9A* variant causing GEFS+, offers an example of how WES has been instrumental allowing us to better dissect the clinical phenotype under study, which is a multilocus variation aggregating in one proband, rather than a phenotypic expansion associated with a single genomic locus, underscoring the role of multiple rare variants at different loci in the aetiology of clinical phenotypes making problematic the diagnostic path.

From simple beginnings to complex endings

The pervasive use of WES and approaches using NGS in general, to identify the genetic basis of Mendelian conditions, has 1) accelerated the rate of novel Mendelian disorders delineation (known gene associated with a novel or different phenotype, or novel gene associated with a novel or known phenotype; 2) allowed identification of >1,000 new Mendelian disorders; 3) replaced ‘‘phenotype-driven’’ with ‘‘genotype-driven’’ syndrome delineation; 4) expanded our understanding about the complex relationships between Mendelian phenotypes and their associated genes and genotypes (Bamshad et al., 2019).

Classically, Mendelian conditions have been categorized as having a dominant or recessive, and autosomal or sex-linked pattern of inheritance. However, in the last years, their study has revealed the extent to which many rare disorders depart from classical genetic expectations, showing complex modes of inheritance (Posey, 2019) instead. It has become clear that conditions whose inheritance strictly conforms to Mendelian principles are relatively rare. The growing genome complexity is changing the fundamentals based on the paradigm of Mendelian genetics "one gene = one disorder" moving far beyond this concept (Katsanis, 2016).

This extent can be characterized by the occurrence of both dominant and recessive inheritance associated with a single locus (i.e. *NALCN*), the observation of more than one Mendelian condition associated with a single locus (i.e. *CREBBP*), and that a number of Mendelian disorders are not the product of a single gene mutation, but of digenic inheritance or of the co-occurrence of two or more genetic disorders ((i.e. *ANKRD11* and *SCN9A*; Deltas, 2017; Fernández-Marmiesse et al., 2018).

It is becoming clear that clinical expertise allows to make a confident clinical diagnosis in a limited number of cases and that in most of them, it is necessary to integrate and correlate both molecular genetic or genomic data (genotype) with the clinical features (phenotype) of the patient (Wright et al., 2018). Clinicians need to consider that rarely a single variant acts in isolation, therefore the clinical features presented by an individual may be the result of a mutational burden and compound inheritance of rare and common variants that can help to explain the clinical heterogeneity in penetrance and expressivity seen between individuals with the same genetic diagnosis, as well as the phenotype expansion, that is the expansion of the spectrum of clinical characteristics of an explained known Mendelian phenotype (Chong et al., 2015; Wright et al., 2018). Therefore, such variation in penetrance and expressivity may suggest that a given trait could be monogenic in one individual, and complex in another (Fournier and Schacherer, 2017).

Inherent to the goal of identifying and characterizing the molecular architecture of Mendelian conditions, is the ability to detect with sufficient sensitivity and specificity the relevant types of variants (Posey, 2019).

"Phenotype-driven" to "genotype-driven"

The definition of a phenotype is an artificial construct useful for clinicians to characterize a particular group of patients with particular clinical and physiological measures, and is based on the combination of an organism's observable characteristics or traits, such as its morphology, development, biochemical or physiological properties, phenology, behavior, and products of

behavior (Chung and Adcock, 2013). Phenotypes depend on both gene inheritance and gene expression. So far, delineation of new Mendelian disorders has been mostly done with a phenotype-driven approach, that is proceeding by collecting multiple individuals with overlapping clinical findings and then identifying the underlying gene. Introduction of NGS-based approaches has impacted the rate of syndrome delineation, switching from the traditional phenotype-driven to genotype-driven method, that is persons with overlapping clinical findings identified only after discovery that they share pathogenic variants in the same candidate gene. In the long run, most Mendelian conditions will be ascertained via genotype-driven delineation (Bamshad et al., 2019). Although our understanding of the variety of Mendelian phenotypes is increasing, significant lacunas remain, in particular, as discussed by Chong et al., 2015, in certifying the number of Mendelian phenotypes that exist, in defining new Mendelian phenotypes, in discriminating new from known Mendelian phenotypes, in delineating what constitutes expansion of a known phenotype, in developing measures to define similarities and diversity of phenotypes caused by variants in the same gene (Chong et al., 2015). This is in particular due to the variability of "normal" human morphology and physiology and to the difficulty in setting limits to define normal versus abnormal (Chong et al., 2015).

Therefore, a key point to achieve the goal of understanding genome function and more importantly, its relations to human disorder, is the need for deep-clinical phenotyping in families along with the use of WES- and NGS-based strategies (Bamshad et al., 2019).

WES complementary approaches

Although current WES- and NGS-based approaches are accelerating the speed of gene discovery for Mendelian conditions, the underlying etiologies for approximately half remain unknown (Esplin et al., 2014). The strength of the WES to be one of the most cost-efficient sequencing approaches for investigating the exome considering that 85% of all mutations that have been identified in Mendelian disorders are in coding regions, is also its limit since a substantial fraction of rare disorder mechanisms will be intractable to WES. Other limits include technical limitations and more complex genetic mechanisms (for example, structural variation, tissue-specific somatic mosaicism, splicing and regulatory mutations, genomic imprinting, oligogenic and digenic inheritance, and gene-environment interactions). Therefore, innovative approaches to address rare disorders caused by these mechanisms must be developed to maintain, or even accelerate, the current pace of discoveries. Whole-genome sequencing (WGS), in particular, has been considered the logical

approach to complement WES for Mendelian gene discovery, especially for the opportunity it provides to uncover non-exomic variants and its ability to identify structural variation, but to date, WGS has produced few discoveries of novel genes or loci underlying Mendelian disorders. (Bamshad et al., 2019; Smedley et al., 2016).

As with any study, detection rates for both WES and WGS range from 25 to 57% varying with the inclusion criteria (Thiffault et al., 2019). Considering that WES would yield the same usable information for current clinical analysis, the cost of WGS is presently difficult to justify (Thiffault et al., 2019). However, as its cost decreases and the ability to interpret non-coding variants improves, the use of WGS over WES will be meaningful (Thiffault et al., 2019). Therefore, until then, periodic reanalysis of WES data would be recommended before performing WGS (Costain et al., 2018).

In addition, WGS presents significant challenges to data interpretation for more than 3 million SNVs per sample, and the validation of non-coding variants (as well as coding changes that impact RNA expression and splicing) usually requires additional functional studies at the transcriptional level (Gonorazky et al., 2019). Recently, the complementation of genetic sequencing with transcriptome sequencing (RNA-seq) has successfully improved the diagnostic yield in Mendelian disorders through the identification of non-coding variants, deep intronic splice variants and synonymous variants with unexpected effects on splicing (Cummings et al., 2017). However, its use is currently limited by the restricted access to tissues where the disorder-gene is expresses and by the cost of creating transdifferentiated cell lines from affected individuals and controls tissues (Bamshad et al., 2019).

Clinical and functional interpretation of variants of uncertain significance

As the technical limitations of WES are in part overcome by WGS and RNA-seq, the significant challenge in Mendelian conditions diagnosis remains the clinical and functional interpretation of variants of uncertain significance. The task of assigning pathogenicity to identified variants is critical to make a definite diagnosis in the patient (Boycott et al., 2017).

It is therefore fundamental to improve both computational and statistical models for variant identification, annotation, functional prediction, and prioritization—particularly for variants in non-coding regions, and platforms that store clinical genetic and phenotype data necessary to share genomic and phenotypic data openly and at scale worldwide (hundreds of putative gene discoveries are unreported buried in medical records, proprietary or restricted-access databases, and scientific papers, and most are difficult to access; Bamshad et al., 2017). In particular the availability of

population-specific disorder and control databases, the development of *in vitro* or cell-based functional assays applicable to gene regulation or protein function, and simultaneous analysis with other broad-based "-omics" approaches such as metabolomics, transcriptomics, or proteomics would be critical to clarify the functional effect of the candidate variant (Posey, 2019).

The possible therapeutic implications are obviously difficult to predict as they presuppose a series of elements that are not known *a priori*, however, the understanding of pathogenetic mechanisms is a fundamental prerequisite for adopting specific preventive and therapeutic actions. As sequence data becomes available in laboratories, it may be possible to classify patients based on mutated genes, identify common phenotypes, or even obtain a sufficient number of affected patients to allow for a search for possible modifier genes.

Despite all these challenges, it is clear that WES has had and will continue to have an increasingly profound influence on the genetics and diagnosis of Mendelian disorders, providing more excellent timeliness in the diagnosis and administration of appropriate therapies and treatments with a high impact in scientific, social, technological and economic terms.

Conclusions

The study of human genetic diseases has traditionally followed a simplistic categorization. Disorders are classified as simple (or Mendelian) vs complex based on family pedigree and recurrence risks in individuals with different degrees of genetic relatedness. Likewise, variant forms of the DNA sequence of the genes are classified as necessary and sufficient to cause a disease or just predisposing based on their alleged phenotypic effect. The recent advent of next generation sequencing approaches with the dramatic advance in knowledge of the consequences of genetic variation on a large number of increasingly better classified phenotypes is, however, beginning to challenge classic definitions of genetic causality.

In my doctoral work, I provide some further examples showing that beyond the apparent simplicity of Mendelian inheritance, there are different and often cryptic levels of difficulty in defining a Mendelian phenotype. This emphasizes, the continuous level of the genetic complexity underlying phenotypes and the hidden complexity of how genetic variants exert a functional impact (Schacherer, 2016).

The rare genetic conditions for which the genetic mechanism has yet to be identified are likely enriched with those that will not be solved easily by existing WES based approaches (Boycott et al., 2017). Identifying the molecular basis of conditions intractable to existing approaches requires broader and innovative application of existing discovery strategies (e.g., WGS, RNA sequencing), improvement of computational and statistical models for variant identification, annotation, functional prediction, and prioritization-particularly for variants in non-coding regions; and development of strategies for discovering causal genetic mechanisms.

As more and more genes are discovered to be associated with Mendelian disorders and appropriate diagnostic tests are established, a significant challenge in their diagnosis will remain the interpretation of a growing number of variants of uncertain significance. This challenge can be addressed by the integration of clinical genetic and phenotype data stored in different platforms (e.g., ClinVar, Leiden Open Variation Database [LOVD], and DECIPHER), by the availability of population- specific disorder and control databases, by the use of broad-based "-omics" approaches such as metabolomics, transcriptomics, or proteomics to clarify functional effect (Boycott et al., 2017). Clearly, the task of assigning pathogenicity to individual variants remains the key to diagnostic precision, critical for patient care and must be curated by pertinent experts.

Moreover, data accumulated from NGS studies supports the awareness that Mendelian and complex disorders can now be considered part of one continuum, whereby common and rare variants in the context of environmental influences result in perturbation of the biological balance of a restricted set of networks activating final common pathways that ultimately cause disorder (Lupski et al., 2011)

Take-home messages from Mendelian disorders research

- "True Mendelian" disorders are very rare: their study has revealed that many rare disorders depart from classical genetic expectations, showing instead complex modes of inheritance.
- The growing genome complexity is changing the fundamentals based on the paradigm of Mendelian genetics "one gene = one disorder" moving far beyond this concept.
- In the end most Mendelian disorders will be ascertained via "genotype-driven" delineation and not via "phenotype-driven".

- Genetic variation at genes linked to Mendelian disorders plays an important role in driving susceptibility to complex disorder. The more we do rare-disorder research, the more we then identify cellular pathways and processes that can apply to more-common disorder.
- The big challenge in Mendelian disorders diagnosis remains the clinical and functional interpretation of the huge amount of genetic data generated from these studies.

References

- 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. A map of human genome variation from population-scale sequencing. Nature. 2010 467(7319):1061-73. doi: 10.1038/nature09534. Erratum in: Nature. 2011 473(7348):544.
- Al-Sayed MD1, Al-Zaidan H, Albakheet A, Hakami H, Kenana R, Al-Yafee Y, Al-Dosary M, Qari A, Al-Sheddi T, Al-Muheiza M, Al-Qubbaj W, Lakmache Y, Al-Hindi H, Ghaziuddin M, Colak D, Kaya N. Mutations in NALCN cause an autosomal-recessive syndrome with severe hypotonia, speech impairment, and cognitive delay. Am J Hum Genet. 2013 93(4):721-6. doi: 10.1016/j.ajhg.2013.08.001. PMID:24075186 doi:10.1016/j.ajhg.2013.08.001
- Alves RM, Uva P, Veiga MF, Oppo M, Zschaber FCR, Porcu G, Porto HP, Persico I, Onano S, Cuccuru G, Atzeni R, Vieira LCN, Pires MVA, Cucca F, Toralles MBP, Angius A, Crisponi L. Novel ANKRD11 gene mutation in an individual with a mild phenotype of KBG syndrome associated to a GEFS+ phenotypic spectrum: a case report. BMC Med Genet. 2019 20(1):16. doi: 10.1186/s12881-019-0745-7. PMID:3064227
- Angius A, Cossu S, Uva P, Oppo M, Onano S, Persico I, Fotia G, Atzeni R, Cuccuru G, Asunis M, Cucca F, Pruna D, Crisponi L. Novel NALCN biallelic truncating mutations in siblings with IHPRF1 syndrome. Clin Genet. 2018 93(6):1245-1247. doi: 10.1111/cge.13162. PMID:29399786
- Angius A, Uva P, Oppo M, Buers I, Persico I, Onano S, Cuccuru G, Van Allen MI, Hulait G, Aubertin G, Muntoni F, Fry AE, Annerén G, Stattin EL, Palomares-Bralo M, Santos-Simarro F, Cucca F, Crisponi G, Rutsch F, Crisponi L. Exome sequencing in Crisponi/cold-induced sweating syndrome-like individuals reveals unpredicted alternative diagnoses. Clin Genet. 2019 95(5):607-614. doi: 10.1111/cge.13532. PMID:30859550

- Angius A, Uva P, Oppo M, Persico I, Onano S, Olla S, Pes V, Perria C, Cuccuru G, Atzeni R, Serra G, Cucca F, Sotgiu S, Hennekam RC, Crisponi L. Confirmation of a new phenotype in an individual with a variant in the last part of exon 30 of CREBBP. Am J Med Genet A. 2019 179(4):634-638. doi: 10.1002/ajmg.a.61052. PMID:30737887
- Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet. 2011 12(11):745-55. doi: 10.1038/nrg3031. Review. PMID:21946919
- Bamshad MJ, Nickerson DA, Chong JX. Mendelian Gene Discovery: Fast and Furious with No End in Sight. Am J Hum Genet. 2019 105(3):448-455. doi: 10.1016/j.ajhg.2019.07.011. PMID: 31491408
- Banka S, Sayer R, Breen C, Barton S, Pavaine J, Sheppard SE, Bedoukian E, Skraban C, Cuddapah VA, Clayton-Smith J. Genotype-phenotype specificity in Menke-Hennekam syndrome caused by missense variants in exon 30 or 31 of CREBBP. Am J Med Genet A. 2019 179(6):1058-1062. doi:10.1002/ajmg.a.61131. PMID:30892814 doi:10.1002/ajmg.a.61131
- Boycott KM, Parboosingh JS, Chodirker BN, Lowry RB, McLeod DR, Morris J, Greenberg CR, Chudley AE, Bernier FP, Midgley J, Møller LB, Innes AM. Clinical genetics and the Hutterite population: a review of Mendelian disorders. Am J Med Genet A. 2008 146A(8):1088-98. doi: 10.1002/ajmg.a.32245. Review. PMID:18348266
- Buers I, Persico I, Schöning L, Nitschke Y, Di Rocco M, Loi A, Sahi PK, Utine GE, Bayraktar-Tanyeri B, Zampino G, Crisponi G, Rutsch F, Crisponi L. Crisponi/cold-induced sweating syndrome: Differential diagnosis, pathogenesis and treatment concepts. Clin Genet. 2019 doi: 10.1111/cge.13639. Review.PMID:31497877
- Buske OJ, Manickaraj A, Mital S, Ray PN, Brudno M. Identification of deleterious synonymous variants in human genomes. Bioinformatics. 2013 29(15):1843-50. doi: 10.1093/bioinformatics/btt308. Erratum in: Bioinformatics. 2015 31(5):799. PMID:23736532
- Chambers C, Jansen LA, Dhamija R. Review of Commercially Available Epilepsy Genetic Panels. J Genet Couns. 2016 25(2):213-7. doi: 10.1007/s10897-015-9906-9. PMID:26536886 doi:10.1007/s10897-015-9906-9
- Chelly J, Khelfaoui M, Francis F, Chérif B, Bienvenu T. Genetics and pathophysiology of mental retardation. Eur J Hum Genet. 2006 14(6):701-13. Review.PMID:16721406

- Chial, H. Rare Genetic Disorders: Learning About Genetic Disease Through Gene Mapping, SNPs, and Microarray Data. Nature Education. 2008) 1(1):192
- Chiurazzi P, Pirozzi F. Advances in understanding - genetic basis of intellectual disability. F1000Res. 2016 7;5. pii: F1000 Faculty Rev-599. doi: 10.12688/f1000research.7134.1. eCollection 2016. Review. PMID:27127621
- Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, Nayir A, Bakkaloğlu A, Ozen S, Sanjad S, Nelson-Williams C, Farhi A, Mane S, Lifton RP. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A. 2009 106(45):19096-101. doi: 10.1073/pnas.0910672106. PMID:19861545
- Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, Harrell TM, McMillin MJ, Wiszniewski W, Gambin T, Coban Akdemir ZH, Doheny K, Scott AF, Avramopoulos D, Chakravarti A, Hoover-Fong J, Mathews D, Witmer PD, Ling H, Hetrick K, Watkins L, Patterson KE, Reinier F, Blue E, Muzny D, Kircher M, Bilguvar K, López-Giráldez F, Sutton VR, Tabor HK, Leal SM, Gunel M, Mane S, Gibbs RA, Boerwinkle E, Hamosh A, Shendure J, Lupski JR, Lifton RP, Valle D, Nickerson DA; Centers for Mendelian Genomics, Bamshad MJ. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. Am J Hum Genet. 2015 97(2):199-215. doi: 10.1016/j.ajhg.2015.06.009. Review. PMID:26166479
- Chung KF, Adcock IM. How variability in clinical phenotypes should guide research into disease mechanisms in asthma. Ann Am Thorac Soc. 2013 Suppl:S109-17. doi: 10.1513/AnnalsATS.201304-087AW. Review. PMID:24313760
- Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012 6(2):80-92. doi: 10.4161/fly.19695. PMID:227
- Claes L, Del-Favero J, Ceulemans B, Lagae L, Van Broeckhoven C, De Jonghe P. De novo mutations in the sodium-channel gene SCN1A cause severe myoclonic epilepsy of infancy. Am J Hum Genet. 2001;68:1327-32. PMID:11359211 doi:10.1086/320609
- Costain G, Jobling R, Walker S, Reuter MS, Snell M, Bowdin S, Cohn RD, Dupuis L, Hewson S, Mercimek-Andrews S, Shuman C, Sondheimer N, Weksberg R, Yoon G, Meyn MS, Stavropoulos DJ, Scherer SW, Mendoza-Londono R, Marshall CR. Periodic reanalysis of whole-genome sequencing data enhances the diagnostic advantage over standard clinical

genetic testing. Eur J Hum Genet. 2018 26(5):740-744. doi: 10.1038/s41431-018-0114-6. PMID:29453418

- Cuccuru G, Orsini M, Pinna A, Sbardellati A, Soranzo N, Travaglione A, Uva P, Zanetti G, Fotia G. Orione, a web-based framework for NGS analysis in microbiology. Bioinformatics. 2014 30(13):1928-9. doi: 10.1093/bioinformatics/btu135. PMID:24618473
- Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, Bolduc V, Waddell LB, Sandaradura SA, O'Grady GL, Estrella E, Reddy HM, Zhao F, Weisburd B, Karczewski KJ, O'Donnell-Luria AH, Birnbaum D, Sarkozy A, Hu Y, Gonorazky H, Claeys K, Joshi H, Bournazos A, Oates EC, Ghaoui R, Davis MR, Laing NG, Topf A; Genotype-Tissue Expression Consortium, Kang PB, Beggs AH, North KN, Straub V, Dowling JJ, Muntoni F, Clarke NF, Cooper ST, Bönnemann CG, MacArthur DG. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. Sci Transl Med. 2017 9(386). pii: eaal5209. doi: 10.1126/scitranslmed.aal5209. PMID:28424332
- de Ligt J, Boone PM, Pfundt R, Vissers LE, Richmond T, Geoghegan J, O'Moore K, de Leeuw N, Shaw C, Brunner HG, Lupski JR, Veltman JA, Hehir-Kwa JY. Detection of clinically relevant copy number variants with whole-exome sequencing. Hum Mutat. 2013 34(10):1439-48. doi: 10.1002/humu.22387. PMID:23893877
- Deltas C. Digenic inheritance and genetic modifiers. Clin Genet. 2018 93(3):429-438. doi: 10.1111/cge.13150. Review. PMID:28977688
- Dragojlovic N, Elliott AM, Adam S, van Karnebeek C, Lehman A, Mwenifumbo JC, Nelson TN, du Souich C, Friedman JM, Lynd LD. The cost and diagnostic yield of exome sequencing for children with suspected genetic disorders: a benchmarking study. Genet Med. 2018 20(9):1013-1021. doi: 10.1038/gim.2017.226.
- Engel P, Bagal S, Broback M, Boice N Physician and patient perceptions regarding physician training in rare diseases: the need for stronger educational initiatives for physicians. J Rare Disord 2013 1(2):1–15.
- Epi4K Consortium; Epilepsy Phenome/Genome Project, Allen AS, Berkovic SF, Cossette P, Delanty N, Dlugos D, Eichler EE, Epstein MP, Glauser T, Goldstein DB, Han Y, Heinzen EL, Hitomi Y, Howell KB, Johnson MR, Kuzniecky R, Lowenstein DH, Lu YF, Madou MR, Marson AG, Mefford HC, Esmaeeli Nieh S, O'Brien TJ, Ottman R, Petrovski S, Poduri A, Ruzzo EK, Scheffer IE, Sherr EH, Yuskaitis CJ, Abou-Khalil B, Alldredge BK, Bautista JF, Berkovic SF, Boro

A, Cascino GD, Consalvo D, Crumrine P, Devinsky O, Dlugos D, Epstein MP, Fiol M, Fountain NB, French J, Friedman D, Geller EB, Glauser T, Glynn S, Haut SR, Hayward J, Helmers SL, Joshi S, Kanner A, Kirsch HE, Knowlton RC, Kossoff EH, Kuperman R, Kuzniecky R, Lowenstein DH, McGuire SM, Motika PV, Novotny EJ, Ottman R, Paolicchi JM, Parent JM, Park K, Poduri A, Scheffer IE, Shellhaas RA, Sherr EH, Shih JJ, Singh R, Sirven J, Smith MC, Sullivan J, Lin Thio L, Venkat A, Vining EP, Von Allmen GK, Weisenberg JL, Widdess-Walsh P, Winawer MR. De novo mutations in epileptic encephalopathies. Nature. 2013 501(7466):217-21. doi: 10.1038/nature12439. PMID:23934111

- Esplin ED, Oei L, Snyder MP. Personalized sequencing and the future of medicine: discovery, diagnosis and defeat of disease. Pharmacogenomics. 2014 15(14):1771-1790. PMID:25493570 doi:10.2217/pgs.14.117
- EuroEPINOMICS-RES Consortium. Epilepsy Phenome/Genome Project; Epi4K Consortium; EuroEPINOMICS-RES Consortium. De Novo Mutations in Synaptic Transmission Genes Including DNM1 Cause Epileptic Encephalopathies. Am J Hum Genet. 2017 100(1):179. doi: 10.1016/j.ajhg.2016.12.012. PMID:28061363
- Fernandez-Marmiesse A, Gouveia S, Couce ML. NGS Technologies as a Turning Point in Rare Disease Research , Diagnosis and Treatment. Curr Med Chem. 2018 25(3):404-432. doi: 10.2174/0929867324666170718101946. Review. PMID:28721829
- Fournier T, Schacherer J. Genetic backgrounds and hidden trait complexity in natural populations. Curr Opin Genet Dev. 2017 47:48-53. doi: 10.1016/j.gde.2017.08.009. Review. PMID:28915487
- Fuentes Fajardo KV, Adams D; NISC Comparative Sequencing Program, Mason CE, Sincan M, Tifft C, Toro C, Boerkoel CF, Gahl W, Markello T. Detecting false-positive signals in exome sequencing. Hum Mutat. 2012 33(4):609-13. doi: 10.1002/humu.22033. PMID:22294350
- Gardiner M. Molecular genetics of infantile nervous system channelopathies. Early Hum Dev. 2006 82(12):775-9. Review. PMID:17049761
- Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. Eur J Hum Genet. 2012 20(5):490-7. doi: 10.1038/ejhg.2011.258. Review. PMID:22258526
- Gobbi G, Sarajlija J, Leonardi S, Di Pietro E, Zara F, Striano P. La genetica delle epilessie e le epilessie genetiche. Neurologia pediatrica 2014; 44:226-239.

- Gonorazky HD, Naumenko S, Ramani AK, Nelakuditi V, Mashouri P, Wang P, Kao D, Ohri K, Viththiyapaskaran S, Tarnopolsky MA, Mathews KD, Moore SA, Osorio AN, Villanova D, Kemaladewi DU, Cohn RD, Brudno M, Dowling JJ. Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease. Am J Hum Genet. 2019 104(5):1007. doi: 10.1016/j.ajhg.2019.04.004. PMID:31051109
- Helbig I, Tayoun AA. Understanding Genotypes and Phenotypes in Epileptic Encephalopathies. Mol Syndromol. 2016 7(4):172-181. PMID:27781027 doi:10.1159/000448530
- Helbig KL, Farwell Hagman KD, Shinde DN, Mroske C, Powis Z, Li S, Tang S, Helbig I. Diagnostic exome sequencing provides a molecular diagnosis for a significant proportion of patients with epilepsy. Genet Med. 2016 18(9):898-905. doi: 10.1038/gim.2015.186. PMID:26795593 doi:10.1038/gim.2015.186
- Ji J, Shen L, Bootwalla M, Quindipan C, Tatarinova T, Maglinte DT, Buckley J, Raca G, Saitta SC, Biegel JA, Gai X. A semiautomated whole-exome sequencing workflow leads to increased diagnostic yield and identification of novel candidate variants. Cold Spring Harb Mol Case Stud. 2019 1;5(2). pii: a003756. doi: 10.1101/mcs.a003756. PMID:30755392
- Joshi RS, Garg P, Zaitlen N, Lappalainen T, Watson CT, Azam N, Ho D, Li X, Antonarakis SE, Brunner HG, Buiting K, Cheung SW, Coffee B, Eggermann T, Francis D, Geraedts JP, Gimelli G, Jacobson SG, Le Caignec C, de Leeuw N, Liehr T, Mackay DJ, Montgomery SB, Pagnamenta AT, Papenhausen P, Robinson DO, Ruivenkamp C, Schwartz C, Steiner B, Stevenson DA, Surti U, Wassink T, Sharp AJ. DNA Methylation Profiling of Uniparental Disomy Subjects Provides a Map of Parental Epigenetic Bias in the Human Genome. Am J Hum Genet. 2016 99(3):555-566. doi: 10.1016/j.ajhg.2016.06.032. PMID:27569549
- Katsanis N. The continuum of causality in human genetic disorders. Genome Biol. 2016 17(1):233. PMID:27855690
- Kochinke K, Zweier C, Nijhof B, Fenckova M, Cizek P, Honti F, Keerthikumar S, Oortveld MA, Kleefstra T, Kramer JM, Webber C, Huynen MA, Schenck A. Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. Am J Hum Genet. 2016 98(1):149-64. doi: 10.1016/j.ajhg.2015.11.024.PMID:26748517
- Köhler S. Improved ontology-based similarity calculations using a study-wise annotation model. Database (Oxford). 2018 doi:10.1093/database/bay026.PMID:29688377

- Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP; NHLBI Exome Sequencing Project, Quinlan AR, Nickerson DA, Eichler EE. Copy number variation detection and genotyping from exome sequence data. Genome Res. 2012 22(8):1525-32. doi: 10.1101/gr.138115.112. PMID:22585873
- Kwong AK, Ho AC, Fung CW, Wong VC. Analysis of mutations in 7 genes associated with neuronal excitability and synaptic transmission in a cohort of children with non-syndromic infantile epileptic encephalopathy. PLoS One. 2015 10(5):e0126446. doi: 10.1371/journal.pone.0126446. eCollection 2015. PMID:25951140
- Li MX, Gui HS, Kwan JS, Bao SY, Sham PC. A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. Nucleic Acids Res. 2012 40(7):e53. doi: 10.1093/nar/gkr1257. PMID:22241780 28672
- Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. Hum Mutat. 2013 34(9):E2393-402. doi: 10.1002/humu.22376. PMID:23843252
- Liu Z, Zhang N, Zhang Y, Du Y, Zhang T, Li Z, Wu J, Wang X. Prioritized High-Confidence Risk Genes for Intellectual Disability Reveal Molecular Convergence During Brain Development. Front Genet. 2018 9:349. doi: 10.3389/fgene.2018.00349. eCollection 2018. PMID:30279698
- Lupski JR1, Belmont JW, Boerwinkle E, Gibbs RA. Clan genomics and the complex architecture of human disease. Cell. 2011 147(1):32-43. doi: 10.1016/j.cell.2011.09.008. PMID:21962505 doi:10.1016/j.cell.2011.09.008
- Magi A, Tattini L, Palombo F, Benelli M, Gialluisi A, Giusti B, Abbate R, Seri M, Gensini GF, Romeo G, Pippucci T. H3M2: detection of runs of homozygosity from whole-exome sequencing data. Bioinformatics. 2014 30(20):2852-9. doi: 10.1093/bioinformatics/btu401. Epub 2014 Jun 24. PMID:24966365
- Maulik PK, Mascarenhas MN, Mathers CD, Dua T, Saxena S. Prevalence of intellectual disability: a meta-analysis of population-based studies. Res Dev Disabil. 2011 32(2):419-36. doi: 10.1016/j.ridd.2010.12.018. Epub 2011 Jan 13. Review. Erratum in: Res Dev Disabil. 2013 34(2):729.PMID:21236634
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework

for analyzing next-generation DNA sequencing data. Genome Res. 2010 20(9):1297-303. doi: 10.1101/gr.107524.110. PMID:20644199

- Menke LA, van Belzen MJ, Alders M, Cristofoli F; DDD Study, Ehmke N, Fergelot P, Foster A, Gerkes EH, Hoffer MJ, Horn D, Kant SG, Lacombe D, Leon E, Maas SM, Melis D, Muto V, Park SM, Peeters H, Peters DJ, Pfundt R, van Ravenswaaij-Arts CM, Tartaglia M, Hennekam RC. CREBBP mutations in individuals without Rubinstein-Taybi syndrome phenotype. Am J Med Genet A. 2016 170(10):2681-93. doi: 10.1002/ajmg.a.37800. PMID:27311832
- Menke LA; DDD study, Gardeitchik T, Hammond P, Heimdal KR, Houge G, Hufnagel SB, Ji J, Johansson S, Kant SG, Kinning E, Leon EL, Newbury-Ecob R, Paolacci S, Pfundt R, Ragge NK, Rinne T, Ruivenkamp C, Saitta SC, Sun Y, Tartaglia M, Terhal PA, van Essen AJ, Vigeland MD, Xiao B, Hennekam RC. Further delineation of an entity caused by CREBBP and EP300 mutations but not resembling Rubinstein-Taybi syndrome. Am J Med Genet A. 2018 176(4):862-876. doi: 10.1002/ajmg.a.38626. PMID:29460469
- Myers CT, Mefford HC. Genetic investigations of the epileptic encephalopathies: Recent advances. Prog Brain Res. 2016;226:35-60. doi: 10.1016/bs.pbr.2016.04.006. Review. PMID:27323938
- Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet. 2010 42(1):30-5. doi: 10.1038/ng.499. PMID:19915526
- Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009 461(7261):272-6. doi: 10.1038/nature08250. PMID:19684571
- Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, Murphy D, Le Cam Y, Rath A. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur J Hum Genet. 2019 doi: 10.1038/s41431-019-0508-0. PMID:31527858
- Nieh SE, Sherr EH. Epileptic encephalopathies: new genes and new pathways. Neurotherapeutics. 2014;11:796-806. Review. PMID:25266964
- Posey JE. Genome sequencing and implications for rare disorders. Orphanet J Rare Dis. 2019 14(1):153. doi: 10.1186/s13023-019-1127-0. Review. PMID:31234920

- Schacherer J. Beyond the simplicity of Mendelian inheritance. C R Biol. 2016 339(7-8):284-8. doi: 10.1016/j.crvi.2016.04.006. Review. PMID:27344551
- Schulze TG, McMahon FJ. Defining the phenotype in human genetic studies: forward genetics and reverse phenotyping. Hum Hered. 2004;58(3-4):131-8. PMID:15812169 doi:10.1159/000083539
- Sidore C, Busonero F, Maschio A, Porcu E, Naitza S, Zoledziewska M, Mulas A, Pistis G, Steri M, Danjou F, Kwong A, Ortega Del Vecchyo VD, Chiang CWK, Bragg-Gresham J, Pitzalis M, Nagaraja R, Tarrier B, Brennan C, Uzzau S, Fuchsberger C, Atzeni R, Reinier F, Berutti R, Huang J, Timpson NJ, Toniolo D, Gasparini P, Malerba G, Dedoussis G, Zeggini E, Soranzo N, Jones C, Lyons R, Angius A, Kang HM, Novembre J, Sanna S, Schlessinger D, Cucca F, Abecasis GR. Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. Nat Genet. 2015 47(11):1272-1281. doi: 10.1038/ng.3368. PMID:26366554
- Smedley D, Schubach M, Jacobsen JOB, Köhler S, Zemojtel T, Spielmann M, Jäger M, Hochheiser H, Washington NL, McMurry JA, Haendel MA, Mungall CJ, Lewis SE, Groza T, Valentini G, Robinson PN. A Whole-Genome Analysis Framework for Effective Identification of Pathogenic Regulatory Variants in Mendelian Disease. Am J Hum Genet. 2016 99(3):595-606. doi: 10.1016/j.ajhg.2016.07.005. PMID:27569544
- Thiffault I, Farrow E, Zellmer L, Berrios C, Miller N, Gibson M, Caylor R, Jenkins J, Faller D, Soden S, Saunders C. Clinical genome sequencing in an unbiased pediatric cohort. Genet Med. 2019 21(2):303-310. doi: 10.1038/s41436-018-0075-8. PMID:30008475
- Topper S, Ober C, Das S. Exome sequencing and the genetics of intellectual disability. Clin Genet. 2011 80(2):117-26. doi: 10.1111/j.1399-0004.2011.01720.x. Review. PMID:21627642
- Trump N, McTague A, Brittain H, Papandreou A, Meyer E, Ngoh A, Palmer R, Morrogh D, Boustred C, Hurst JA, Jenkins L, Kurian MA, Scott RH. Improving diagnosis and broadening the phenotypes in early-onset seizure and severe developmental delay disorders through gene panel analysis. J Med Genet. 2016 53(5):310-7. doi: 10.1136/jmedgenet-2015-103263. PMID:26993267 doi:10.1136/jmedgenet-2015-1
- van Bokhoven H. Genetic and epigenetic networks in intellectual disabilities. Annu Rev Genet. 2011;45:81-104. doi: 10.1146/annurev-genet-110410-132512. Review. PMID:21910631

- Vissers LE, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. Nat Rev Genet. 2016 17(1):9-18. doi: 10.1038/nrg3999. Review. PMID:26503795
- Von Deimling M, Häsler R, Steinbach V, Holterhus PM, von Spiczak S, Stephani U, Helbig I, Muhle H. Gene expression analysis in untreated absence epilepsy demonstrates an inconsistent pattern. Epilepsy Res. 2017 132:84-90. doi: 10.1016/j.eplepsyres.2017.02.008. PMID:28324682 03263
- Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in children. Nat Rev Genet. 2018 19(5):325. doi: 10.1038/nrg.2018.12. doi: 10.1038/nrg.2017.116.

WEB resources

EURORDIS, https://www.eurordis.org

https://www.rarediseaseday.org

https://www.globalrarediseasecommission.com

ORPHANET, https://www.orpha.net/

OMIM, https://www.omim.org/

HPO, http://www.human-phenotype-ontology.org/

GATK, https://software.broadinstitute.org/gatk/

UCSC, https://genome.ucsc.edu/

RefGene, http://varianttools.sourceforge.net/Annotation/RefGene

Ensembl, https://www.ensembl.org/index.html

dbSNP, https://www.ncbi.nlm.nih.gov/projects/SNP/

1000 Genomes Project, http://www.internationalgenome.org/

ESP6500, http://evs.gs.washington.edu/EVS/

ExAC, http://exac.broadinstitute.org/

CLINGEN, https://www.clinicalgenome.org/

DGV, http://dgv.tcag.ca/dgv/app/home

ClinVar, https://www.ncbi.nlm.nih.gov/clinvar/

LOVD, https://www.lovd.nl/

DECIPHER, https://decipher.sanger.ac.uk/

Acknowledgments

I would like to thank all the patients, families and clinicians involved in this study.
I've been lucky to have worked with an extraordinary group, and I want to thank you each member, Andrea Angius, Paolo Uva, Manuela Oppo, Ivana Persico, Stefano Onano, Gianmauro Cuccuru, Rossano Atzeni for a great time spent working together and for the important results we were able to achieve.
I would like to thank Prof. Francesco Cucca for the precious opportunity he gave me to carry out this project.
Last but not least, I thank all the persons that directly or indirectly helped me to get these results.

 Angius A, Cossu S, Uva P, Oppo M, Onano S, Persico I, Fotia G, Atzeni R, Cuccuru G, Asunis M, Cucca F, Pruna D, Crisponi L. Novel NALCN biallelic truncating mutations in siblings with IHPRF1 syndrome. Clin Genet. 2018 93(6):1245-1247. doi: 10.1111/cge.13162.

 Angius A, Uva P, Oppo M, Persico I, Onano S, Olla S, Pes V, Perria C, Cuccuru G, Atzeni R, Serra G, Cucca F, Sotgiu S, Hennekam RC, Crisponi L. Confirmation of a new phenotype in an individual with a variant in the last part of exon 30 of CREBBP. Am J Med Genet A. 2019 179(4):634-638. doi: 10.1002/ajmg.a.61052.