DIPARTIMENTO DI NEUROSCIENZE UMANE

Dottorato di Ricerca in Neuroscienze Clinico- Sperimentali e Psichiatria

Curriculum Disturbi Neurologici, Neuroevolutivi e Psichiatrici in età evolutiva

Ciclo XXXII

"Identification and functional characterization of new disease genes and variants implicated in pediatric encephalopathies: from exome sequencing to *in vivo* investigation"

Tutor:

Prof. Vincenzo Leuzzi

Candidata:

Martina Venditti

Co-Tutor:

Dott. Marco Tartaglia

“Don't let anyone rob you of your imagination,

your creativity or your curiosity.”

Mae Jemison

INDEX

Abstract

Rare diseases are a heterogeneous group of clinical conditions that affect pediatric patients in about two-thirds of the cases and, most of them, have a genetic cause. These are chronic conditions, often complex to treat. The diagnosis is very challenging for many of these patients and, in about half of the pediatric cases presenting syndromes associated with cognitive impairment, identifying the cause of the disease remains complex. The *Ospedale Pediatrico Bambino Gesù* (OPBG) and above all the Unit of "Molecular genetics and functional genomics", directed by Dr. Marco Tartaglia, carries out projects dedicated to find the genes implicated in rare diseases in orphan patients and explore the underlying pathogenetic mechanisms. In my PhD thesis, I have contributed to discover the molecular cause of forms of complex encephalopathies participating to whole-exome sequencing (WES) efforts that allowed to identify new disease genes of previously unknown pathologies. For one case, I investigated the pathogenesis using zebrafish as an *in vivo* model under expert supervision on animal experimentation. Specifically, the goal of my thesis was: 1. Initial identification, through the use of WES, of new genes responsible, when mutated, for rare neurodevelopment disorders (different forms of pediatric, isolated and syndromic encephalopathies); 2. The functional validation of a panel of mutations affecting one of these genes, through the use of small freshwater fish, *Danio rerio* (zebrafish), as an ideal model system for studying pathophysiology in embryonic development programs.

During the first part of my PhD activity, I became confident with the WES technology starting from the DNA preparation library (from patients' peripheral blood DNA samples) to sequencing reactions and row data analysis (derived from bioinformatics core of our research group). In this context, I focused my PhD work on the study of news forms of encephalopathies caused by mutations in genes encoding channels or subunits of ion channel (***KCNK4***; Bauer., *et al*., 2018) and encephalopathies caused by mutations in genes encoding proteins involved in the control of cytoskeletal dynamics and intracellular trafficking (***CLTC***; Manti., *et al*., 2018; and ***ARF3***; manuscript in preparation). The identification of new disease-causing genes or genetic variants allow me to begin with the *in vivo* functional validation.

Indeed, I focused the last period of my PhD work on the functional study of the variants identified in the *ARF3* gene modeled in zebrafish. First, I performed studies of arf3 sequence to define the conservation among orthologs genes in human and in zebrafish. Second, I actively participated in the analysis of the phenotype of *ARF3* transiently expressed mutants generated by my colleagues. Given the clinical features of patients (short stature with skeletal abnormalities and microcephaly), I investigated morphological characteristics of embryonic development leading to body axis formation and the possible impact on cephalic development. I deepened my analyses by looking at markers of early axis establishment (convergence and extension movements during early embryogenesis) via whole-mount *in situ* hybridization of specific mRNAs (*krox20* and *myoD)* used as markers to reveal abnormalities in those processes. An enlargement of the medio-lateral (ML) axis as compared to a shortening of the anterior-posterior (AP) axis, which was evident in early mutant embryos after gastrulation as early as 15h of development. Fish had a shortened body axis and microcephaly.

In summary, in my PhD I have contributed to find genetic diagnosis in some previously unsolved cases via next-generation sequencing and have actively participated in the *in vivo* validation of a new disease-gene causing a severe form of encephalopathy.

1. Introduction

1.1 Undiagnosed Diseases Program

Rare diseases are heterogeneous clinical conditions, which, in the European definition, affect 1 person per 2.000 (Rommel K, *et al*., 2010). They are pediatric conditions in about two thirds of the cases and have a genetic origin approximately in 90% of those. Currently, over 8.000 nosological entities are known. Their overall number is a social problem of the national healthcare (approximately at least 750.000 people are affected in Italy excluding rare cancer forms). About 5-10% of rare patients are orphan of diagnosis and cure, but the percentage of patients with disabilities and without diagnosis even reaches 40%. Furthermore, a conspicuous number of them, about 5.000 are ultra-rare (<1/100.000). In Italy, there are ~1.000.000 individuals affected by a form of rare disease, 60% of which in pediatric age (Eurordis, 2007; Boycott *et al*., 2017).

Generally, more than 50% of children who approach a pediatric hospital is affected by a medical condition of genetic cause or by a multifactorial disorder with a strong incidence of genetic component (Chong *et al*., 2015). These forms include diseases caused by chromosomal (*i.e.* number and structural anomalies) and genetic alterations (*i.e.* specific mutations in a single/multiple gene/s). Despite a significant investment of time and money, many of these rare pathologies are orphans of diagnosis for the lack of disease-causing gene and knowledge on the pathogenetic mechanisms involved.

The Bambino Gesù Children's Hospital (OPBG) is the largest pediatric institution in Italy that welcomes~ 13.000 patients suffering from these chronic diseases, mostly of high clinical complexity and difficult to cure (Italian register of rare diseases in 2017). Diagnosis remains challenging for most of these patients (~ 50% of the cases). Understanding the molecular causes and the pathogenetic mechanisms of a disease is an essential requirement for diagnosis, correct stratification of patients, definition of natural history and its clinical variability for the optimization of the successful treatment protocols. Indeed, the identification of new disease-causing genes and their pathogenetic mechanism can: 1) promote the development of new diagnostic tests, 2) characterize

new nosological entities, 3) differentiate similar diseases, outlining new correlations between genetic alterations and clinical features, 4) contribute to implement early diagnosis offering targeted genetic counseling to the family, 5) offer regular prevention programs as well as 6) monitor any future high-risk pregnancy.

Since 2015, the "*Molecular genetics and functional genomics*" research unit in particular has focused part of the research activities on an ambitious program dedicated to undiagnosed patients. This program recruited 560 patients, selected in the context of 48 multidisciplinary multi-center teleconsultation sessions. The causes of disease in 32% of the recruited patients were investigated at least at two levels. The first one comprised the **analysis of the coding regions of genes known to be associated with human disease** (from now on called *clinical exome*), based on a specific diagnosis hypothesis. This investigation step allowed to reach a diagnosis in 66% of cases. On the other hand, for the rest of patients (~ 42%), with unclear clinical features and whose clinical exome investigation did not highlight any known genetic aberration in disease genes, a deeper level of investigation towards previously undescribed variants likely causative of the unknown disease, object of the study. This approach is referred to as **"*research exome* analysis"**. This second level of analysis aimed to understand the molecular mechanisms of disease caused by known genes already associated to a disease in that 42% of patients and identify potential new disease-causing genes in the remaining 13% of patients. Overall, combining different approaches, this program has led to the identification of several new disease-causing genes and new clinical-genetic classification of previously uncharacterized pediatric conditions (Kortüm *et al*., 2015; Chong *et al*., 2016; Flex *et al*., 2016; Muto *et al*., 2018; Niceta *et al*.,2015). Alongside the development and application of these genomic technologies in the diagnostic and research fields, there is a growing need to develop complementary experimental *in vitro* and *in vivo* models aimed at the functional validation of new disease-causing genes /gene variants identified by genomic sequencing approach.

1.2 Infantile encephalopathies with neurodevelopmental defects

The developmental and epileptic encephalopathies (DEEs) are a heterogeneous group of rare and ultra-rare syndromes characterized by seizures, behavioral disturbances, or

abnormalities of spontaneous electrical neuro-activity. These syndromes involve the impaired brain development (*i.e.* developmental encephalopathies) or developmental regression and progressive cerebral dysfunction (*i.e.* epileptic encephalopathies) (Nashabat, *et al*., 2019); the cause of DEEs remains unknown in the majority of cases. Epilepsy is often associated with major comorbidities, most frequently intellectual disability (ID), which affects 25% of cases (Berg *et al*., 2008; Tuchman and Cuccaro, 2011). Conversely, the frequency of lifetime history of epilepsy ranges from 7%–15% for individuals with mild to moderate ID to 45%–82% for those with severe ID. The co-occurrence of epilepsy and ID can involve at least two non-exclusive mechanisms. In some cases, uncontrolled seizures can be detrimental to developing cortical networks and can lead to regression and poor cognitive outcomes in children (Ben-Ariand Holmes, 2006). The term epileptic encephalopathy (EE) has been used to designate disorders where the epileptic activity itself contributes to cognitive slowing or regression, and EE can occur in a child with or without preexisting developmental delay (Scheffer*et al*., 2017). In other cases, a single genetic or environmental process is sufficient to induce both seizures and cognitive impairment (Brooks-Kayal, 2011). For instance, mutations that induce specific synaptic defects might result in aberrant connectivity and seizures, as well as alter synaptic plasticity and cause learning disabilities. The term developmental encephalopathy (DE) has been proposed to designate disorders where developmental delay emerges before the presence of epileptic activity or in the presence of infrequent seizures. Because it is not always easy to dissect the contribution of each of these mechanisms and because some genetic disorders can involve both mechanisms in the same or in different individuals, the term developmental and epileptic encephalopathy (DEE) has been coined to refer to conditions characterized by ID and epilepsy where both mechanisms might play a role (Hamdan *et al*., 2017). Epileptic encephalopathies are complex and heterogeneous disorders that make difficult not only the diagnosis but also the treatment decisions. Recent exome sequencing data suggest that mutations causing epileptic encephalopathies are often sporadic, typically resulting from *de novo* dominant mutations in a single autosomal gene, although inherited autosomal recessive and X-linked forms also exist (Nieh and Sherr, 2014).

If, on the one hand, the use of modern genomic sequencing technologies generate valid informations and allow the identification of new disease-causing genes/genetic variants underlying different forms of pediatric encephalopathies, on the other it is important to

approach their functional validation. For most of these diseases, *in vitro* analyses have already provided the first level of information on pathogenetic mechanisms. However, studies based on the *in vivo* model are necessary for 1) the **understanding of the impact that these genetic variants** have on morphogenetic and developmental programs and physiology cellular, focusing on the cellular processes that coordinate the development of the central nervous system (CNS) and the neuronal functions, and 2) **identifying possible pharmacological approaches** able to contrast, slow or stop the onset or progression of the pathological phenotype. Although *in vitro* approaches are used successfully for this purpose, these are not highly informative in understanding patho-physiology and the pleiotropic effects on development programs, especially in the case of genes whose coded protein has not been fully characterized and the relationship between it and the biological processes behind the disease are not clear. For all these reasons, alongside the use of **exome sequencing**, the use of an *in vivo* model system becomes necessary. For this purpose, the small teleost fish of freshwater, **zebrafish**, represents an ideal animal model to study different forms of pediatric encephalopathies.

1.3 Next generation sequencing

Rare diseases, or orphan diseases, caused by altered functions of single genes can be chronically debilitating and life limiting. Some rare diseases are compatible with a good quality of life if they are diagnosed early and optimally managed. Although the individual diseases are rare, they collectively affect millions of individuals worldwide. Currently, >9000 diseases are estimated to exist, 75% of which affect children. Unfortunately, effective therapies for these diseases are themselves comparatively rare. Thus, in addition to the effects on patients and their families, these diseases have a tremendous cost for health care systems and societies (Boycott *et al*, 2017; Shen., 2014). It is very difficult to define the precisely number of rare genetic diseases. An interrogation of Online Mendelian Inheritance in Man (OMIM), a catalogue of human genes and associated genetic diseases, and Orphanet, a comprehensive reference portal for rare diseases, results in a best estimate of between 6,000 and 7,000 rare genetic diseases. The advent of next-generation sequencing (NGS) has changed the landscape of rare-genetic-disease research, with causative genes being identified at an accelerating rate.Next generation sequencing (NGS) represents an entirely new principle of

sequencing technology following Sanger (first generation) sequencing, which was first described in 1977 (Sanger *et al*., 1977). NGS became available to the community in 2009 when the first NGS machines entered the market (Lohmann and Klein, 2014). Technical improvements of this sequencing technology enabled automation of this approach, thereby increasing the sequencing capacity from a few hundred base pairs to several thousands of them within a single analysis (Metzker *et al*., 2005). NGS, or massively parallel sequencing, represents a method of simultaneously sequencing millions of fragments of DNA. It has become crucial in several fields, including human genetics, virology, system biology, and forensic biology, among others and it has been rapidly adopted in the clinical laboratory because of its ability to simultaneously analyze several genes or gene regions with a single test and in a short time compared to traditional methods. The Sanger sequence reaction produces DNA chains arbitrarily terminated at each of the different positions by introducing a dideoxy-nucleotide and, subsequently, separating the pool of these chains according to size by electrophoresis; conversely, NGS is based on the principle of "sequencing-by-synthesis". This concept means that the complementary integration of a single nucleotide during chain prolongation (*i.e.* the sequencing reaction) is directly monitored by the sequencing machine (Lin *et al*., 2008). In medical genetics, DNA sequencing has been proven as a powerful tool in finding the human disease-causing genes (Rabbani *et al.*, 2014). Understanding the pathogenetic mechanism underlying genetic diseases mostly depends on finding the causative genes/genetic variants associated with a specific phenotype. The identification of new genes causing human rare monogenic disorders is crucial to understand the biological pathways underlying diseases and, consequently, improve the genetic counseling in respect to diagnosis, prognosis, and risk assessment as well as treatment and therapeutic management (Rabbani *et al.*, 2014). Using NGS approach, it is possible to study human DNA through: 1) **Whole Genome Sequencing** (WGS), which is able to sequence the whole genome of an individual in a single experiment; by this sequencing strategy about 4 million variants per individual can be detected (Lam *et al*., 2012), and 2) **Whole Exome Sequencing** (WES), which captures and sequence the exome of the human genome (~ 1% of the genome). By this sequencing strategy it is possible to detect about 20.000 variants per individual. Both sequencing processes represent powerful and unbiased methods for detecting genetic variations (Boycott *et al.*, 2013). WGS/WES of a patient eliminates the sequencing of a targeted and limited genes selection, reducing time required to identify the disease-causing genes and

leading from a two-step to a one-step approach (Gilissen*et al.*, 2012).Currently, WES is the most used approach for the discovery of rare disease-causing genes because it has been estimated that 85% of the disease-causing genes/genetic variants are located in the coding portion of the human genome because of its feasibility in terms of computational infrastructures, costs and interpretation effort (Boycott *et al.*, 2013; Rabbani *et al.*, 2014). WES presents fundamental limitations compared to WGS such as the inability to assess the impact of non-coding alleles. Despite this, WES is a well-justified strategy for discovering rare alleles underlying Mendelian phenotypes and complex traits (Bamshad *et al.*, 2011) as well as predisposing genetic variants in common diseases and cancers (Boycott *et al.*, 2013). The WES analysis in a research or diagnostic setting provides for a) a wet laboratory workflow, including library preparation and the actual sequencing of the library (refers to par. 1.3.1), and b) a dry laboratory part involving informatics and bioinformatics analyses (sequence alignments and single nucleotide variant calling), and variant filtering and interpretation (annotation, gene variant mapping against genetic variation databases) (refers to par. 1.3.2). The key steps of exome sequencing are shown in figure 1:

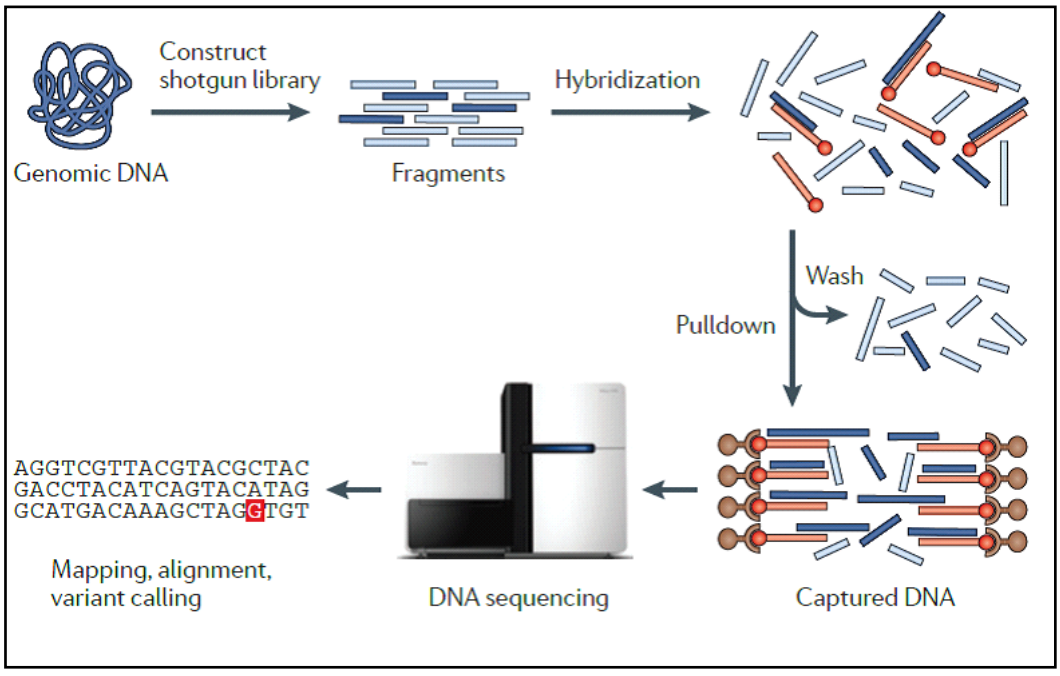


Figure 1: Workflow for the exome sequencing. First step is DNA fragmentation and the construction of a library. Fragments corresponding to exons (in dark blue) are hybridized to DNA or RNA probes (in orange) and then they are recovered through a streptavidin-based pulldown assay. Following steps are represented by amplification and massive parallel sequencing of the enriched and amplified library and the mapping and calling of candidate causal variants (Bamshad *et al.*, 2011).

1.3.1 "Wet lab" workflow: library preparation, exome capture and sequencing technology

The **library preparation** is the first step and refers to the process of DNA fragment preparation (by randomized fragmentation) and collection followed by a ligation of specialized adaptors (called also *linkers*) to both fragment ends. Few micrograms of genomic DNA are sufficient to construct an *in vitro* shotgun library. Next, as the second step, the DNA fragment library is enriched for sequences corresponding to exons through aqueous-phase hybridization capture and the exon fragments are hybridized to biotinylated DNA or RNA single-stranded oligonucleotides (also called *probes* or *baits*). Recovery of targeted fragment-probe hetero duplexes (through a streptavidin beads pulldown assay) is followed by the amplification reaction.

The **exome capture** sequesters protein coding regions by hybridization of genomic DNA library to biotinylated oligonucleotide baits that are DNA or RNA complementary to targeted exons. Magnetic streptavidin beads are used to selectively pull-down and enrich baits to which the target sequences are linked. The sample preparation methods are highly similar across the different technologies but there are considerable differences among those technologies in terms of target regions, and in the total number of bases targeted (Chilamakuri *et al.*, 2014). Currently, there are four major solution-based human exome capture systems available: *Nimble Gen's SeqCap EZ Exome Library, Agilent's SureSelect Human All Exon*, *Illumina's TruSeq Exome Enrichment*, and *Illumina's Nextera Exome Enrichment*. The major differences between the technologies correspond to the choice of their respective target regions, bait lengths, bait density, molecules used for capture, and genome fragmentation method (Table 1).

	NimbleGen	Agilent	Illumina TruSeq	Illumina Nextera
Bait type	DNA	RNA	DNA	DNA
Bait length range (bp)	NP	114-126	95	95
Median bait length (bp)	NP	119	95	95
Number of baits	NP	554,079	347,517	347,517
Total bait length (Mb)	NP	66.48	33.01	33.01
Target length range (bp)	59–742	114–21,747	2–37,917	2–37,917
Median target length (bp)	171	200	135	135
Number of targets	368,146	185,636	201071	201,071
Total target length (Mb)	64.19	51.18	62.08	62.08
Fragmentation method	Ultrasonication	Ultrasonication	Ultrasonication	Transposomes
Automation	++	++	++	+++
Throughput	+++	+++	+++	+++
Flexibility	Custom available	Custom available		Custom available
Species	Human, mouse, 3 plant species	Human, mouse, 14 other species custom	Human	Human
Costs	$$	$$	$	$

Table 1: Performance and design comparison of four exome capture systems for deep sequencing (Chilamakuri *et al.*, 2014).

The sequencing of the enriched and amplified library (range from 25-500 base pair, bp) can be performed through different technologies based on fluorescence or chemiluminescence methods (Bamshad *et al.*, 2011). Among the most used **sequencing platforms** two NGS systems are normally used, which differ in the underlying chemistry technology and offer different advantages and disadvantages.

- Illumina platform (including the *HiSeq, MiSeq, and NexSeq* benchtop sequencers): it uses a flow cell to immobilize each DNA fragment and clonally amplify it for generating a large enough signal for the detection. In detail, the flow cell contains sequences that hybridize to a part of the adaptor on the DNA fragments. The clonal amplification step creates a cluster with approximately 1000 identical copies of a unique parent DNA molecule that are physically isolated from other molecules. Illumina platform, taking advantage of the principle of the "sequencing by synthesis" with fluorescent detection, it produces sequence reads of several hundreds of bp length (up to 300 bp) from tens of millions of simultaneously amplified DNA fragments.

- Ion Torrent technology (including the *IonPGM, IonProton*, and *IonS5* systems): it uses a bead emulsion for the immobilization and the clonal amplification. Conversely, to Illumina platform, Ion Torrent sequencing adds a single base during the chain prolongation in each round. When a base is incorporated, a hydrogen ion is released,

accompanied by a pH change that is detected for each bead within a well; if a base is not incorporated there is no voltage generated (Yohe and Thyagarajan, 2017).

1.3.2 "Dry lab" workflow: sequence alignment, variant calling and gene annotation

The exome sequencing experiments provide a large amount of reads, 140 million on average, depending on capture and sequencing technologies used (Chilamakuri *et al.*, 2014). The several parameters to evaluate NGS experiments include sequencing depth and coverage. Depth represents the average number of reads that align to a specific position in the reference genome, while coverage refers to the degree of enrichment of target sequences (exons in case of WES). In the first step of WES analysis, the raw data of reads (fastq files) are analyzed to remove sequence adaptors and low quality reads (DePristo *et al.*, 2011). The raw data of reads deriving from the sequencers undergo a series of bioinformatics processes (also referred as a *bioinformatic pipeline*) to ultimately deliver a variant call file (VCF), which contains the genomic variants and allows to describe in tabular format the most common genomic variants of a genome, together with the possibility of inserting annotations and metadata (Danecek *et al*, 2011).The bioinformatic pipeline include three key phases (summarized in Fig.2; Meena, *et al*., 2017).

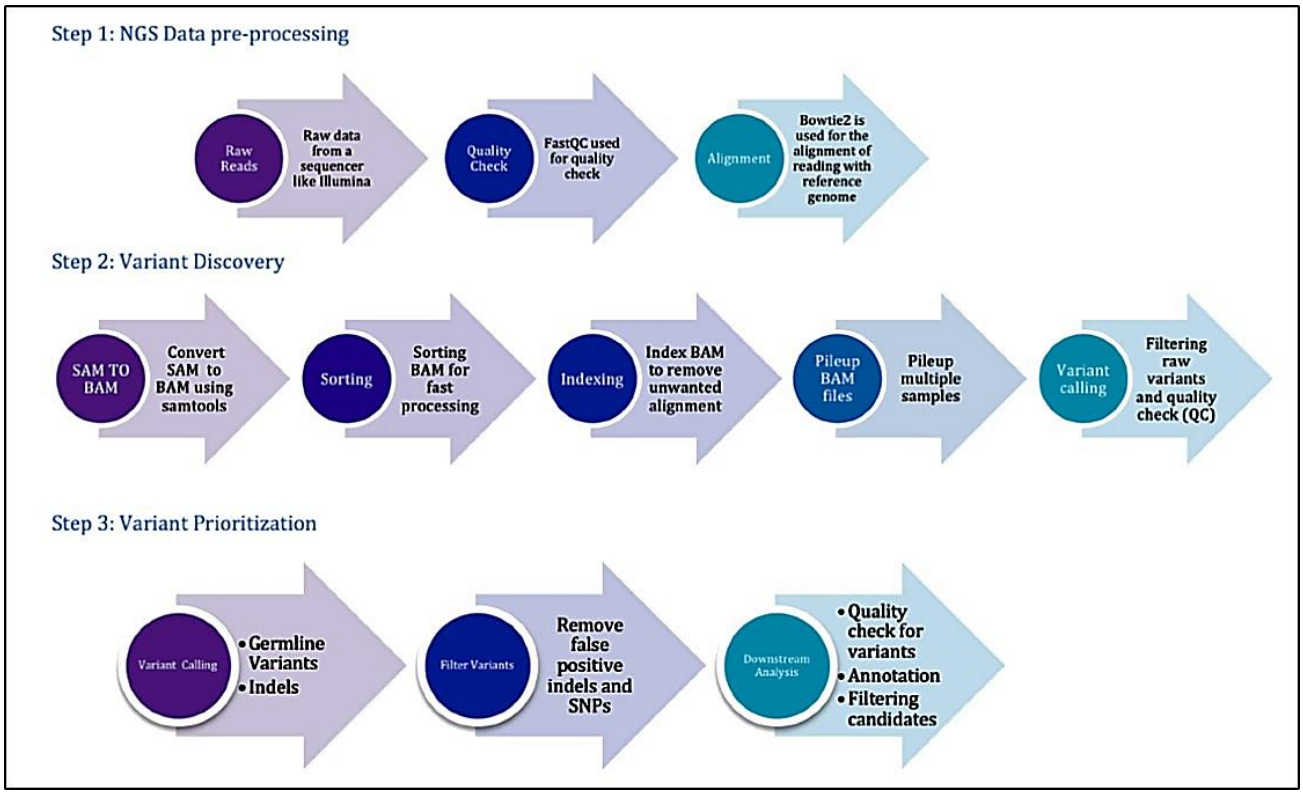


Figure 2: The pipeline involving three important phases, viz. preprocessing, variant discovery and prioritization of variants. (Meena, *et al*., 2017).

1) **NGS data pre-processing,** *demultiplexing*, defined as a process for the separation of an individual sample's reads from the pooled reads of multiple samples by unique identifier codes that were attached before pooling), quality analysis, mapping of the reads to a reference genome (also called *re-sequencing*), in the first step of WES analysis, raw reads data (FASTQ files, files with consensus assessment of sequence and variation) are analyzed to remove sequence adaptors and low quality reads (DePristo *et al.*, 2011). The next step is represented by the alignment to the reference genome, a process that determines the exact position of each read on the human genome (Ruffalo *et al.*, 2011). There are different tools that can achieve this task, and that take into account several factors, e.g. genetic variation in the population, sequencing error, short read length and the huge volume of short reads to be mapped (Ruffalo *et al.*, 2011).

2) **Variant discovery,** the output is a BAM file (Binary version of sequence alignment/map), of about 6 GB, containing all retained reads (Sophia Yohe; Bharat Thyagarajan., 2017).

3) **Variant prioritization,** the next step is the variant calling that consists in the identification of DNA sequence variation relative to the reference genome. Variations that can be recognized by WES are single-nucleotide variants and small insertion-deletions (indels); the output file of this analysis is a Variant Calling File (VCF), a text format file containing variants calling (Van der Auwera *et al.*, 2013).

The final step is characterized by the annotation of variants and of genes in which the variants localize, in order to characterize them and to investigate their possible involvement in a specific pathology or biological process. Variant interpretation is complex when applied to whole genes (as opposed to well-defined hotspots) and a large number of genes. There are several kinds of functional annotations that can be retrieved for DNA variants such as position in the genome (e.g. genes, intron/exon, etc.), effect on amino acid sequence (silent, non-synonymous, missense, etc.) and conservation (Torkamani *et al.*, 2011). Variants can also be annotated using information present in population databases as dbSNP (Database of Single Nucleotide Polymorphisms - http://www.ncbi.nlm.nih.gov/SNP/), 1000 Genomes Project (http://www.1000genomes.org/) and ExAC (Exome Aggregation Consortium - http://exac.broadinstitute.org/). At this point, the interpretation of the variants, which concerns with evaluating whether they could be potentially pathogenic and therefore causative of the disease studied) is a critical step. To this aim, the genetic information is collected and evaluated for categorization of the effect on an encoded protein, for in silico prediction of the consequences on protein function, and for previously reported knowledge on the gene and the specific mutation in question from databases. For this step can be used a functional annotation to predict the potentially damaging effect of identified variants on protein function (CADD score - Combined Annotation Dependent Depletion - http://cadd.gs.washington.edu/), the involvement in pathogenic mechanisms (OMIM - Online Mendelian Inheritance in Man - http://www.ncbi.nlm.nih.gov/omim), animal disease models (MGD - Mouse Genome Database - http://www.informatics.jax.org/ and ZFIN database-https://zfin.org/), and gene expression (GXD Gene Expression Database - http://www.informatics.jax.org/mgihome/GXD/aboutGXD.shtml) (Torkamani *et al.*, 2011). These approaches explore the predicted proteins changes caused by specific amino acid substitutions and enrich for functional sites at which observed variants are more likely to affect phenotype (Bamshad *et al.*, 2011). This allows to quantitatively

prioritize functional, deleterious, and disease causal variants across a wide range of variants.

Finally, when one or a few candidate nucleotides changes are selected as potential pathogenic variants, a comprehensive functional validation is needed to confirm the pathogenicity and the causative link with the disease object of the study. This functional investigations include: 1. re-sequencing the human samples to exclude false positive; 2. Setup an experimental workflow to investigate the impact of the new variants found in the disease causing mechanism using ideally both *in vitro* and *in vivo* models.

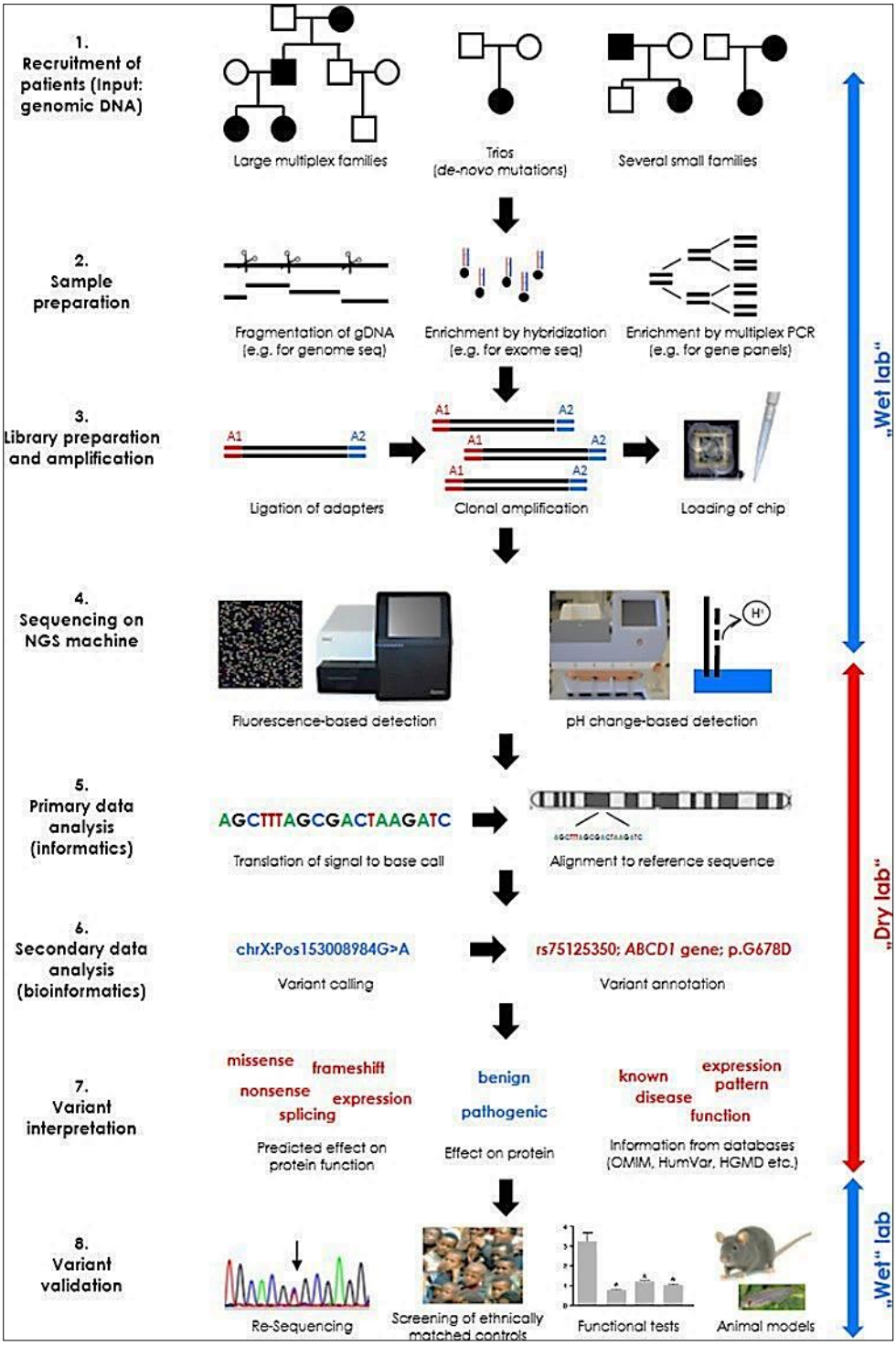


Figure 3: Workflow of NGS analysis. The figure provides a simplified overview of the main phases in NGS and demonstrates some of the different options at each step (*Figure modified by Katy Lohmand Christine Klein., 2014).*

1.4 The model organism Zebrafish (*Danio Rerio*)

The zebrafish, *Danio Rerio*, is a small tropical freshwater fish belonging to *Cyprinidae* family. The zebrafish is native to the streams of the south-eastern Himalayan region and it is found in parts of India, Pakistan, Bangladesh, Nepal, and Burma. This specie arose in the Ganges region in eastern India and commonly inhabits streams, canals, ditches, ponds, and slow-moving or stagnant water bodies, including rice fields.

The zebrafish is so called because of five uniform, pigmented, horizontal, blue or purple stripes on the side of the body, which are reminiscent of a zebra's stripes and which extend to the end of the caudal fin. Its shape is fusiform and laterally compressed, with its mouth directed upwards. The male is torpedo-shaped, with gold stripes between the blue stripes; the female has a larger, whitish belly and silver stripes instead of gold (Figure 4). The zebrafish can grow to 2.5 cm in length, although it seldom grows larger than 4 cm in captivity. Its lifespan in captivity is around two to three years, although in ideal conditions, this may be extended to five years (Spence *et al.*, 2008). Controlled cross-breeding has given rise to different varieties: wild-type (with the classic stripes, from which it derives its name), leopard, longfin (characterized by long fins), and the cross-breeding of longfin type with spotted skin leopard variety.

Females can spawn every 2-3 days and a single clutch may contain several hundred eggs. The **generation time is short**, typically 3-4 months, making it suitable for selection experiments. Zebrafish eggs are large relative to other fish, (0.7 mm in diameter at fertilization) and **optically transparent**, a characteristic that makes the zebrafish a convenient research model in genetics. Furthermore, taking advantage of the **external fertilization**, live embryos are accessible to manipulation and can be monitored through all developmental stages under a dissecting microscope (Kimmel *et al.*, 1995). The **organogenesis** and **development is rapid** (within 72 hours post fertilization, hpf) and larvae display food seeking and active avoidance behaviors within five days post fertilization (dpf), 2-3 days after hatching (Kimmel *et al.*, 1995). The sex of juveniles cannot be distinguished and sex determinants are not clearly understood. The zebrafish are omnivorous, primarily eating zooplankton, insects, larvae and small crustaceans (*i.e.* Artemia salina).

Figure 4: Male and female zebrafish

Zebrafish development is divided into seven broad periods of embryogenesis: the zygote, cleavage, blastula, gastrula, segmentation, pharyngula, and hatching. These divisions highlight the changing spectrum of major developmental processes that occur during the first 3dpf (Kimmel *et al*., 1995):

1) ZYGOTE PERIOD (0-3/4 h): the newly fertilized egg is in the zygote period until the first cleavage occurs about 40 minutes after fertilization. The zygote is about 0.7 mm in diameter at the time of fertilization. In this period many changes are occurring. Fertilization also activates cytoplasmic movements. Non-yolky cytoplasm begins to stream toward the animal pole, segregating the blastodisc from the clearer yolk granule rich vegetal cytoplasm;

2) CLEAVAGE PERIOD *(3/4-2$^{1/4}$* h): after the first cleavage the cells, or blastomeres, divide at about 15-minute intervals. The cytoplasmic divisions are meroblastic; they only incompletely undercut the blastodisc (Kimmel and Law, 1985). The six cleavages that comprise this period frequently occur at regular orientations and are characterized by different numbers of cells (2-cells stage,4-cells stage, 8-cells stage, 16-cells stage, 32-cells stage and 64-cells stage);

3) BLASTULA PERIOD *(2 1/4-5 1/4*h): a period in which the blastodisc begins to look ball-like, at the 128-cell stage, and until the time of onset of gastrulation. Important processes occur during this blastula period; the embryo enters mid-blastula transition (MBT), the yolk syncytial layer (YSL) forms, and epiboly begins. This stage ends at 30 % epiboly;

4) GASTRULA PERIOD *(5 1/4-10* h): the epiboly goes on with the following period of gastrula stage, ranging from 5 to 10 hpf. During this developmental stage, the morphogenetic cell movements of involution, convergence and extension take place, leading to the formation of mesoderm, ectoderm and endoderm. The beginning of involution defines the onset of gastrulation. It is also possible to recognize the anterior (that will generate the head) from the posterior (that will generate the tail) part of the main body axis, the ventral from the dorsal and medial tissue from the lateral ones. (Figure5)

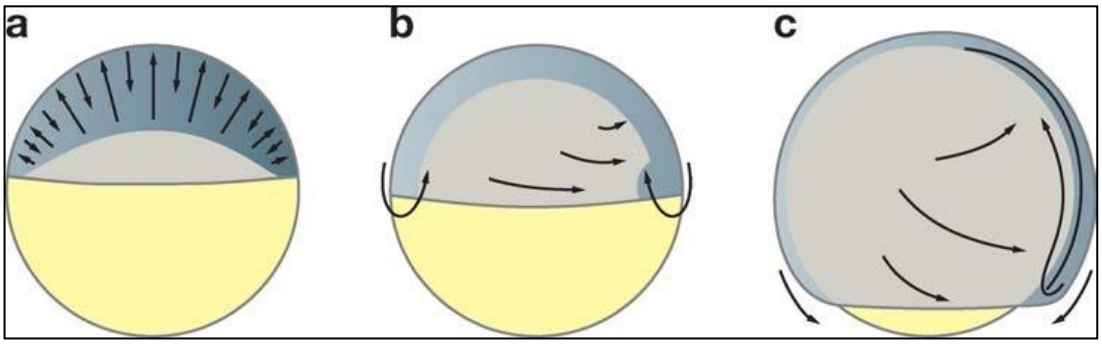


Figure 5: Gastrulation movements. (*a*) Dome stage. Cells intercalate radially, contributing to epiboly. (*b*) Shield stage. Cells at the margin internalize and migrate toward the animal pole. Cells converge dorsally, with lateral mesodermal cells starting convergence at later stages than cells closer to the shield (282). (*c*) 90% epiboly stage. Epiboly, internalization, convergence and extension continue. (Schier and Talbolt, 2005).

5) SEGMENTATION PERIOD (10-24 h): the segmentation period is so called because of segmentation of the nervous system (Figure6). It's characterized by a variety of morphogenetic movements, the formation of somites, which are the structures giving rise to the trunk muscles, the rudiments of the primary organs become visible, the tail bud becomes more prominent and the embryo elongates. The anterior-posterior and dorso-ventral axes are unambiguous. The first cells

differentiate morphologically, and the first body movements appear;

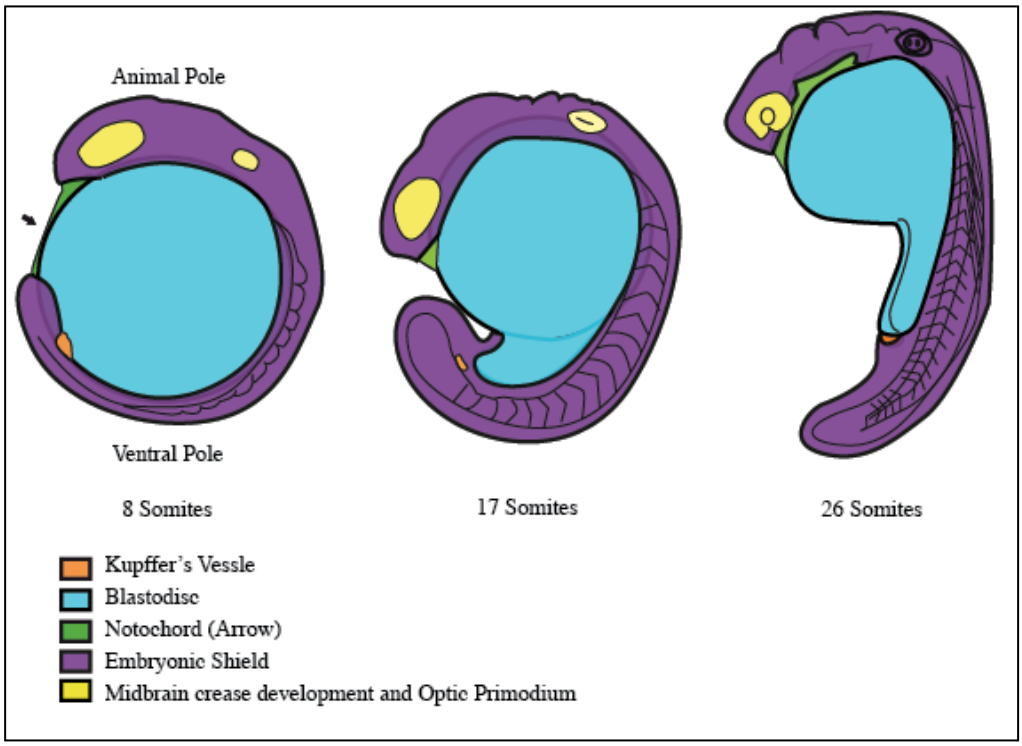


Figure 6: https://embryology.med.unsw.edu.au/embryology/index.php/Zebrafish_**Segmentation_Period.** The image shows the development of the Kupffer Vessel and the theformation of the subdivisions of the brain located at the animal pole. The straightening out of the posterior trunk also occurs. The lumps along the dorsal neural tube show the formation of the hindbrain rhombomeres; divided segments of the neural tube within the hindbrain

6) PHARYNGULA PERIOD (24-48 h): the term "pharyngula" refers to the embryo that has developed to the phyolotypic stage, when it possesses the classic vertebrate *bauplan*. The embryo is most evidently now a bilaterally organized creature, entering the pharyngula period with a well-developed notochord, and a newly completed set of somites that extend to the end of a long post-anal tail. The nervous system is hollow and expanded anteriorly. The embryo is fully transparent up to 22 hpf, by which stage the melanocytes start to differentiate and the embryo begins to be pigmented. The movements are irregular and non-controlled. Towards the end of this period all organs

were formed but the intestine;

7) HATCHING PERIOD (48-72 h): this is the last embryonic phase, in which the embryo comes out of the chorion (membrane surrounding the zygote) and the embryonic development is completed. During the hatching period the embryo continues to grow at about the same rate as earlier. The morphogenesis of many organ rudiments is now rather complete. After hatching, the embryo becomes a larva able to swim freely and all anatomical structures are fully established (Kimmel *et al.*, 1995).

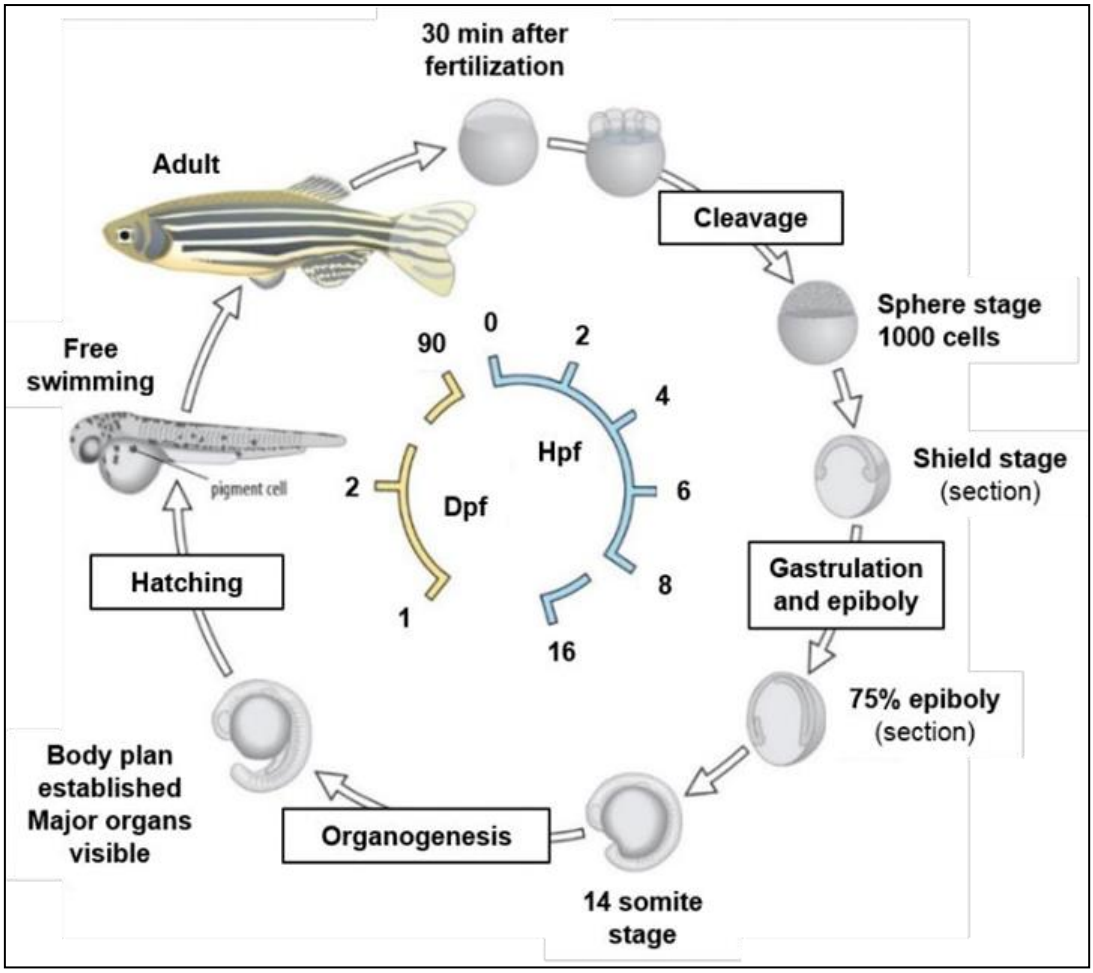


Figure 7: Zebrafish developmental stages. (Ramcharran, 2016).

1.4.1 Why does zebrafish make such good animal model?

In recent years, zebrafish has become widely used for human diseases. This organism has convincingly demonstrated the usefulness of fish for improving the understanding of the molecular and cellular mechanisms leading to pathological conditions, and for the development of new diagnostic and therapeutic tools (Schartl, 2014). Zebrafish are widely used in research due to their numerous beneficial properties:

- Fully sequenced genome: the zebrafish share~ 70% of their genes with humans, and 84% of disease-causing genes have a zebrafish counterpart. The zebrafish genome has been fully sequenced and over 140,000 genes have been mutated to study their function in development and disease (Hannah Simmons, M.Sc.). In teleosts, as in zebrafish, the specific genome duplications led to the generation of paralogous gene copies, whose mechanism of divergence resulted in small ("subfunctionalization") or large ("neofunctionalization") changes in paralogues; this feature represent a good advantage for studying the contribute of each paralogue. (MátéVarga, *et al* 2018):

- Easy manipulation of its genome: transparent zebrafish embryos allows study of the different developmental stages starting from the early embryogenesis. Until now, it has been possible to generate more than 10.000 mutants in protein-coding genes and several transgenic lines of zebrafish to study human diseases (Howe et al., 2013). The easy availability of multiple strains (wild type, transgenic and mutant) of zebrafish is another important advantage of this species, which allows studying (considering the transparency of embryos) several biological process *in vivo* (Teame, *et al*., 2019).

- High fecundity and external fertilization: breeding and getting eggs from the zebrafish is relatively easy. A single reproductive event can ensure ~ 150 - 200 eggs, externally fertilized and produced regularly in large numbers (multiple hundreds of eggs/week). The external fertilization also helps the study of developmental processes avoiding the use of invasive techniques (Clark and Ekker. 2015) and for pharmacological high-throughput screenings. (MacRae and Peterson 2015)

- Short generation time and rapid and transparent embryonic development: zebrafish embryos form complete organ systems, including heart, intestine and blood vessels

within 48-72 hpf. Their organs share the same main features as humans and, for this reason, it is powerful to study human developmental processes (Clark and Ekker, 2015). Importantly, all the main phases of embryogenesis (from blastula formation, gastrulation movements to segmentation) are readily visible in the externally developing zebrafish embryos and occur in relatively short period of time. This makes the model ideal for addressing specific alterations of embryogenesis processes (including gastrulation defects and body axis formation, cell migration and early differentiation signaling for organogenesis) (Clark and Ekker, 2015). For instance, zebrafish is even used for studying hematopoiesis because these fish have the same sequential multi lineage hematopoiesis process as humans, investigating cardiovascular diseases because the gene determinants involved in cardiac development and differentiation are evolutionary conserved (Sakai and Hoffman., 2018). In addition, zebrafish is an excellent model also for modelling neurological disorders given the high conservation of the main neuronal cell populations and circuit architecture with other vertebrates and the underlying genetics (Veldmanand Lin, 2008). Zebrafish individuals grow rapidly and reach the sexual maturity about at third month of the development. This shortens the overall experimental processes and is particularly useful in mutant generation (Sakai and Hoffman., 2018).

- Easy maintenance: it is very affordable to maintain zebrafish colonies in a relatively small amount of laboratory space. Although zebrafish lines require relatively easy management, it is extremely important to optimize a healthy diet and adequate water quality to ensure the fish health and growth (Spence, *et al*., 2008).

1.4.2 Use of zebrafish to investigate pediatric rare disease affecting the neurodevelopment

The advent of the NGS technology allows to sequence large regions of DNA rapidly and economically, which has been particularly advantageous for the diagnosis of rare genetic diseases. This possibility triggered rapid progress in the identification of their causative mutations in recent years (Koboldt *et al*., 2013) and led to the establishment of several large-scale collaborative initiatives, including national and international networks of clinicians and researchers, with the aim of accelerating the discovery of the

rare disease-causing genes. In many cases, these new sequencing technologies allowed a fast and accurate identification of new disease-causing genes and/or new variants in genes already associated with diseases (Máté Varga *et al.*, 2018). Together with the integration of these technologies in the clinical practice as diagnostic tool, development of complementary experimental approaches is representing an important step to shed light on the pathological mechanisms underlying the newly discovered genes and variants. This is crucial for future test and development of pharmacological molecules; given that for the majority of these pediatric rare disease a cure is missing. For new genes and variants found, the functional approaches, based on *in vitro* and *in vivo* model systems, are therefore necessary to 1. Confirm the causative role of genetic variants in the rare diseases orphan of diagnosis as well as 2. For the understanding of the pathogenetic mechanisms involved and find effective cures. Over the past four decades zebrafish has become one of the most in-demand genetic organisms (Kinth *et al*., 2013; Grunwald *et al*., 2002; Varga *et al.*, 2018; Lieschke *et al.*, 2007). Zebrafish became so important in the genetics research because it has several characteristic of an ideal vertebrate genetic model organism (for details see paragraph 1.4.1). Especially, the high genetic conservation make it an ideal genetic model organism to study pediatric disease. The sequencing of the zebrafish genome revealed that 71% of all human proteins and 82% of those that cause disease have a zebrafish ortholog (Howe, K. *et al*., 2013). About 75% (13.217 / 17.727) of the genes included in the PedAM database, published on the record of pediatric diseases, have a zebrafish ortholog (Jia*et al*., 2018) (Figure 8). Taking advantage of its unique features zebrafish is an ideal *in vivo* model for the study of genetic disorders (for details see par. 1.4.1), especially with respect to genetic neurodevelopmental disorders (Sakai and Hoffman., 2018).

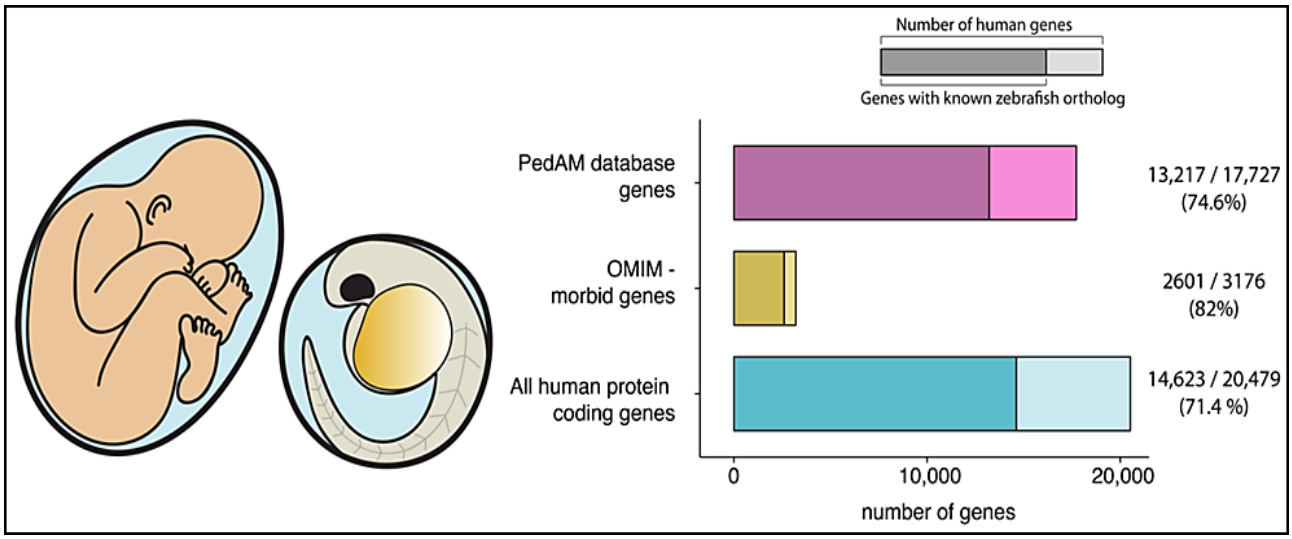


Figure 8: A high level of genetic conservation makes zebrafish an ideal genetic model organism to study pediatric disease. (OMIM–Online Mendelian Inheritance in Man database) (MátéVarga, *et al* 2018).

Indeed, several successful zebrafish models exist for neurodevelopmental disorders depending on genetic causes: 1) Autism Spectrum Disorders (ASD), a devastating group of disorders affecting the neurodevelopment characterized by marked impairments in social behavior and communication, and by the presence of restricted, repetitive behaviors; 2) epilepsy accompanied by intellectual disability, a common neurological condition characterized by recurrent seizures and 3) Attention-Deficit/Hyperactivity Disorder (ADHD), marked by an ongoing pattern of inattention and/or hyperactivity-impulsivity that interferes with functioning or development, and 4) schizophrenia, a psychotic disorder characterized by hallucinations, disorganized thought processes or behavior which severely impacts overall functioning. Most studies of the zebrafish models affecting the neurodevelopment have focused on early morphological and behavioral phenotypes to identify the molecular and cellular mechanisms underlying these disorders. Recent advances in neuro-functional imaging allow also to follow the altered biological processes 'in real time' leading to the neuronal circuit-level characterization of phenotype resulting from risk gene disruption (Simvoulidis *et al*., 2017, Sakai and Hoffman., 2018).

1.5 Early neurogenesis in zebrafish

To understand how to model a pediatric neurological disorder from the molecular to the morphological level in zebrafish, it is essential to understand how the fish brain develops as compared to other vertebrates and humans. “Neurogenesis” describes the process by which undifferentiated neural progenitor cells generate mature and functional neurons, starting at the beginning of gastrulation, around 6 hpf. The induction of neural progenitors and the cell division, that enlarges the pool of progenitors, are the early phases of neurogenesis. Subsequently, a sequence of progenitor specification and commitment to differentiation of post-mitotic neurons occurs. Each of these phases is spatially and temporally orchestrated to generate the multiple neuronal and glial cell types that will populate the mature CNS.

1) The first step in the development of the vertebrate nervous system is the **neural induction by extrinsic factors**, or specification of the neuroectoderm**,** during the early embryonic development (Schmidt *et al*., 2013). At the onset of gastrulation, the forming mesodermal layer involutes and meets the overlying ectoderm (Doniach T and Musci TJ, 1995; Lumsden A and Krumlauf R, 1996; Spemann H and Mangold H, 2001). This presumptive mesodermal layer secretes important factors locally to induce or inhibit neural induction in the ectodermal layer. The extrinsic signaling factors, which are involved in the neural induction, are members of the bone morphogenetic protein (BMP), wingless-integrated (Wnt) and fibroblast growth factor (Fgf) families (Streit A, *et al*., 2000; Wilson SI, *et al.*, 2001). Ventral secretion of BMPs, especially BMP2, 4 and 7, blocks neural induction by inducing an epidermal fate (Sasai and De Robertis, 1997; Wilson, *et al*., 1997). BMP antagonists such as Noggin and Chordin are produced early in the dorsal pre-organizer region, which later forms the Spemann organizer, corresponding to the shield organizer in fish. These secreted proteins act permissively for the establishment of the neural fate in the dorsal ectoderm and allow the formation of the neural plate (Wessely O, *et al.*, 2001).
2) The **neural fate commitment** is determined by intrinsic factors involving especially members of the SRY-box containing genes B1 (SoxB1) family (Avilion AA, *et al*., 2003; Streit A, *et al*., 1997). Together with the extrinsic factor participation, intrinsic networks, programmed by both transcriptional and epigenetic factors, participate in the regulation of the neural fate commitment as well. Recent findings focus on the initiation of the nervous system, elaborately regulated by the intrinsic programs, which are

mediated by transcriptional factors such as Sox2, Zfp521, Sip1 and Pou3f1, as well as epigenetic modifications, including histone methylation/demethylation, histone acetylation/deacetylation, and DNA methylation/demethylation (Tang, *et al*., 2015).

3) Once specified, the neural ectoderm forms the **neural plate**, a pseudostratified epithelial structure in zebrafish. During the process of neurulation, the zebrafish differs from other vertebrates: instead of folding the neuronal plate immediately in a tube with a lumen, it is formed first a solid neural keel. However, the topological arrangement of cells in zebrafish during neuronal formation of the keel from the neural plate is similar to that of other vertebrates. In the so-called "secondary neurulation", the neural rod of the fish swells and forms a typical vertebrate tube. Thus, although there are differences, neurulation in fish and mammals it leads to the formation of a highly similar structure in each case, the neural tube (Papan C and Campos-Ortega JA, 1999). After the formation of the lumen, there is an increasing number of cells dividing asymmetrically into the neural tube; indeed the most apically derived daughter cell becomes the neuron, while a more basal daughter fills the apex pool of progenitors. Important for this asymmetric division is the Notch signaling: basal self-renewal daughter cells show high Notch activity while apical daughter cell differentiation shows low Notch activity (Alexandre, *et al.*, 2010; Dong, *et al.*, 2012).
4) Following neural plate formation, the **early anterior-posterior patterning of the neural plate** is established. The early segmentation in the brain emerge morphologically during early gastrulation as a series of bulges along the A-P axis of the neural tube, including 7-8 hindbrain rhombomeres. Each rhombomere contains the same basic cell types (e.g., commissural interneurons, branchiomotor neurons), but also have distinct identities and contribute to different cranial nerves (Lumsden 2004). In the embryo, each rhombomere also forms a compartment, within which cell lineages are confined and separated by distinct boundary regions (Kimelman, and Martin, 2012).

   As development proceeds, the major zebrafish brain structures develop and, from 10 hpf, CNS structures can already be identified. By 24 hpf, three primary vesicles develop in the rostral portion of the neural tube: the prosencephalon (forebrain), mesencephalon (midbrain) and rhombencephalon (hindbrain). The **forebrain**, the anterior-most part of the embryonic brain, is perhaps one of the most complex domains of the embryonic brain, as it develops into the telencephalon, the diencephalon, the hypothalamus, and the retina. These structures represent an important functional part of the brain, responsible for receiving and processing sensory information and directing behavior. The zebrafish

telencephalon composed by the pallium, the subpallium, and the olfactory bulb. The thalamus, pineal body, and habenula are part of the zebrafish diencephalon. The olfactory bulb is responsible for perceiving odor information, and it contains neurons that transmit that information to other regions of the brain, such as the telencephalon, thalamus, and habenula (Friedrich and Korsching, 1997; Miyasaka, *et al*., 2009). The telencephalon is associated with regulating the social behavior, memory, and emotion, while the hypothalamus regulates fundamental aspects of physiological homeostasis and behavior. The **midbrain** is a relatively small structure in the adult zebrafish brain, but important for vision and hearing. The major structures derived from the midbrain are the tectum and the tegmentum. The tectum, or optic tectum, is the visual processing and response center, with retinal ganglion cells connecting this structure to the retina. The tectum and its connecting neurons are especially important for survival, since they make up the startle and reflex response center (Karlstrom, 1997; Portugues and Engert, 2009; Kita, 2015). The **hindbrain** is one of the most studied parts of the zebrafish brain and can easily be identified during embryonic development since it is physically separated anteriorly from the midbrain by the midbrain-hindbrain boundary (MHB). The hindbrain is composed of the posteriorly located medulla oblongata, the ventro anterior pons, and the dorso anterior cerebellum. The medulla and pons are often referred to as the "brainstem", containing a conglomerate of cell groups that form a complex network termed the reticular formation, which is involved in higher order behaviors such as respiration, circulation, and wakefulness. Indeed, the primary role of the cerebellum is to coordinate movement, and it thus receives information from many sources while transmitting information primarily to motor cortical areas (Moens and Prince 2002). The cerebellum is easily identified by the physical specification of eight compartments along the neural tube called rhombomeres (r1–r7), as early as 10 hpf (Kimmel, C.B. *et al*., 1995; Moens and Prince 2002). Each rhombomere gives rise to cells that differentiate into specific neurons, which innervate the head but also project through the spinal cord to other parts of the body (Oxtoby and Jowett, 1993) (Figure9).

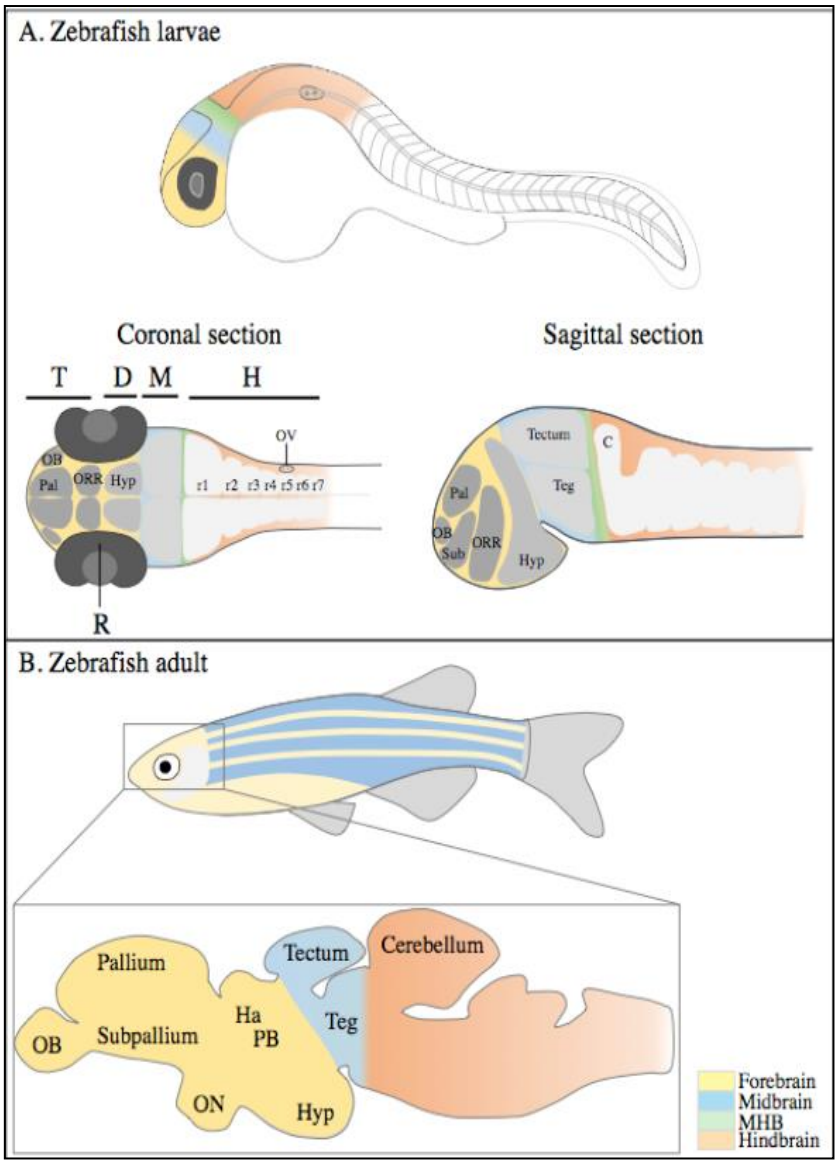


Figure 9. **Development of the zebrafish brain.** (A) Schematic representation of the embryonic brain (30 hpf), showing the forebrain (in yellow), midbrain (in blue), MHB (in green), and hindbrain (in orange). Coronal and sagittal section schemes show brain structures primordia. Forebrain is composed by the telencephalon (in darker gray) and the diencephalon (containing the hypothalamus, lighter grey). (B) Simplified representation of the adult brain and main domains.C: cerebellum; D: diencephalon; M: midbrain; MHB: midbrain-hindbrain boundary; H: hindbrain; Ha: habenula; Hyp: hypothalamus; OB: olfactory bulb; ON: optic nerve; ORR: optic recess region; OV: otic vesicle; Pal: pallium; PB: pineal body; R: retina; r1–r7: rhombomeres 1 to 7; Sub: sub-pallium; T: telencephalon; and Teg: tegmentum. (Raquel Vaz *et al.*, 2019).

1.6 Morphogenesis of axes formation in zebrafish

The basic vertebrate body plan of the zebrafish embryo is established in the first 10 hours of development. Especially during the gastrulation periods, the movements of epiboly, internalization, convergence, and extension transform the radially symmetric blastula into the gastrula embryo with clear dorso-ventral and anterior-posterior axes (Adams RJ and Kimmel CB. 2004). The morphogenetic movements of gastrulation are a series of co-ordinated movements of cell groups that give rise to the three-germ layersectoderm, mesoderm and endodermand overtlyshape the embryonic axis, mostly orchestrated by the non-canonical Wnt (Schier and Tabolt, 2005). During the past 10 years, a combination of genetic, embryological, and molecular analyses has provided detailed insights into the mechanisms underlying this process. Maternal determinants control the expression of transcription factors and the location of signaling centers, responsible for patterning of the blastula and gastrula. Instead, Bmp, Nodal, FGF, canonical Wnt, and retinoic acid signals generate positional information that leads to the restricted expression of transcription factors that control cell type specification. One such movement is convergent extension, when cells of the mesoderm and ectoderm accumulate on the dorsal side of the gastrula by means of highly directed and integrated movements. This results in both **medio–lateral** (ML) narrowing (convergence) and **anterior–posterior** (AP) elongation (extension) of tissues to create the embryonic axis.
This mechanism is regulated by a process that is called "medio-lateral intercalation behavior (MIB)", which underlies the rearrangements (Figure10). In according to MIB hypothesis, a single force-generating cellular machine, distributed across a field of cells, produces both convergence and extension, both narrowing and elongation of the field. By the MIB hypothesis, as applied particularly to the domain of notochord-forming cells within dorsal mesoderm, motile and adhesive cells become polarized along one particular axis, the medio-lateral (ML) axis. The polarity may depend on, and be coordinated within the field, by a non-canonical Wnt signaling planar polarity pathway (Choi and Han, 2002; Heisenberg *et al.*, 2000). The cells take on a bipolar shape, elongating along the ML axis. This process requires that the individual cells all correctly orient actin-based cytoskeletal machinery that mediates motility, and perhaps also orient associated adhesion complexes on their plasma membranes (Zalik *et al.*, 1999).

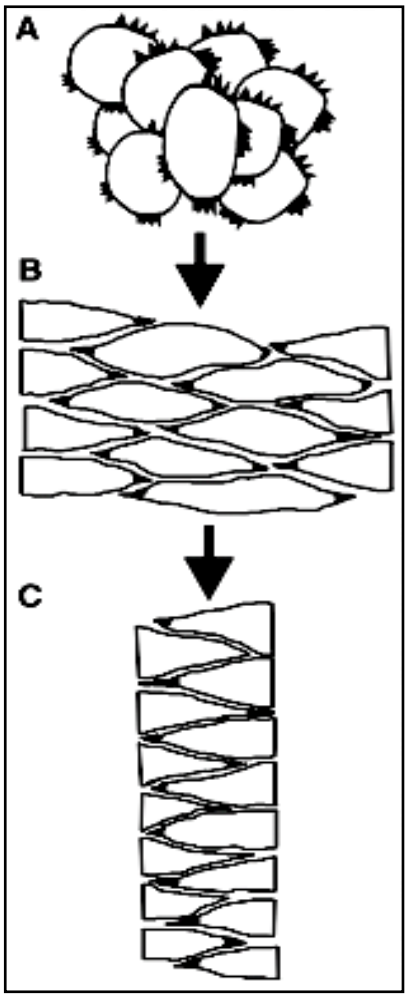


Figure 10: The mediolateral intercalation behavior (MIB) hypothesis. (A) Early protrusive activity is random. (B) The cells take on a bipolar shape as their protrusions are restricted to the ML axis. (C) The field narrows and elongates (converges and extends) as the cells exert traction upon one another and pulltogether (Nathalia S. Glickman *et al.*, 2002).

With this phenomenon, cells intercalate medio laterally and, to accomplish this, they all protrude filopodial processes both medially and laterally that extend between immediate cellular neighbors, and new cell-cell contacts are made, perhaps involving adhesion molecules of the cadherin/protocadherin superfamily localized to their filopodial tips. The newly contacting cells then contract their processes. They exert traction upon one another and pull together. Such events, which occur across the entire field of cells, narrow the field (convergence). The resulting intercalations push previously neighboring cells apart along the AP axis, lengthening the field along this axis (extension). The cells disperse along the AP axis by intercalating with neighbors to form a discontinuous AP string, as expected if MIB mediates extension (Kimmel and Warga, 1986; Kimmel *et al.*, 1994). It is clear that convergent extension of the mesoderm is the major driving force of vertebrate gastrulation. During this process, mesodermal cells move toward the future dorsal side of the embryo, then radically change behavior as they initiate extension of the body axis.

AP patterning in the vertebrate embryo can be roughly divided into two major phases:

1) **An initiation phase**, in which the embryo is generally divided into the head and the body. In zebrafish, the initiation phase occurs prior to the start of gastrulation, such that by the start of gastrulation, the different territories of the final body plan can be roughly mapped onto the embryo. The mesoderm of the head, which comprises part of a very important signaling center called "the Organizer". It is first specified near the equator on what is defined as the dorsal side of the embryo (Saude *et al*., 2000). These cells migrate toward the animal pole during gastrulation, where the brain forms. In contrast, the major mesodermal derivative of the body (the fast muscle) and the spinal cord, are at the gastrula stage oriented with their AP axes along what is termed the dorsal-ventral axis (Kimelman and Martin, 2012);
2) **An elaboration phase**, in which the body progressively forms toward the posterior end, forming the trunk and tail (this process involves the formation of blocks of muscle tissue called somites and it is defined the somitogenesis stages).

1.7 AP patterning and segmentation of the vertebrate head and trunk: role of *krox20* and *myoD*.

The neuroepithelium of the developing hindbrain undergoes a transient AP segmentation into seven segmental progenitor cell compartments, derived largely, or entirely, from ectoderm, named rhombomeres (r1–r7). (Lumsden and Krumlauf, 1996). This conserved organization underlies the segmental organization of cranial nerves in the developing head of all vertebrate embryos. The vertebrate hindbrain segmentation is an evolutionarily conserved process that involves a complex interplay of transcription factors and signaling pathways that show spatially restricted patterns of expression along the AP axis, with limits corresponding to prospective or established boundaries between adjacent rhombomeres (Lumsden and Krumlauf, 1996; Schilling, 2008). They constitute compartments and developmental units for neuronal differentiation, branchiomotor nerve organization and neural crest specification (Lumsden and Keynes, 1989). The rhombomeres are units of cell lineage restriction, they are serially reiterated and each of them containing the same basic cell types (*e.g.* commissural inter-neurons, branchiomotor neurons), but also have distinct identities and contribute to different cranial nerves (Lumsden 2004). In the embryo, each rhombomere also forms a

compartment, within which cell lineages are confined and separated by distinct boundary regions (Fraser et al. 1990). Before boundary formation, cells can intermingle into adjacent territories, but once sharp borders with distinct cellular properties are formed between adjacent rhombomeres, neuroepithelial cells and their progeny can no longer cross into adjacent segments. This cellular segregation process is paralleled by progressive establishment and sharpening of rhombomere-specific gene expression domains, providing distinct segmental identities to each compartment (Lumsden and Krumlauf, 1996). The rhombomere identities are determined, at least in part, by a combinatorial code of transcription factors such as the Hox genes (Lumsden 2004). Stripes of rhombomere-specific Hox gene expression domains appear in the neural ectoderm by the end of gastrulation and these are positioned by early signals along the AP axis. In detail, three signals form an AP patterning network: 1) Wnt signals initially repress expression of anterior genes, 2) fibroblast growth factor (FGF) and 3) retinoic acid (RA) activate posterior genes in a concentration-dependent manner (Cho and De Robertis 1990; Sive, *et al*. 1990; Holowacz and Sokol 1999; Domingos, *et al.* 2001; Kudoh, *et al.* 2002). FGF and RA, in particular, are good candidates for the molecular link between AP patterning and segmentation, potentially acting as graded ‘‘morphogens’’ because they establish the boundaries of segments with distinct concentration thresholds. **FGF signalling** plays a major role because its signaling is necessary to promote the expression of the transcription factor *krox20* (also known as *Egr2*) which is required for the formation and specification of two segmental units: rhombomeres 3(r3) and 5(r5) (Aragon and Pujades, 2009; Marin and Charnay, 2000; Charlotte Labalette, *et al*., 2011). *Krox20*encodesa protein with three C2H2-type zinc finger. This protein was shown to bind to a specific DNA sequence and to act as a transcriptionfactor. It belongs to a small subfamily of proteins, with similar zinc fingers, which recognize identical or very closely related GC-rich sequences (Oxtoby and Jowett, 1993). This process is regulated by a specific signaling pathway: during segmentation, *krox20* expression in r3 and r5 represses the expression of the RA-degrading enzyme genes, *cyp26b1* and *cyp26c1*, thus maintaining high RA levels. *Krox20* auto regulates itself and represses the r4 marker hoxb1. In r4, high *cyp26b1* and *cyp25c1* expression levels contribute to maintain a low RA environment, with is permissive for hoxb1 expression. Hoxb1 also auto regulates itself and represses*krok20* expression. When an isolated r3-derived cell intermingles into r4, it becomes exposed to environmental low RA signaling from surrounding cells, and the RA level is reduced

within the ectopic cell. This allows *Hoxb1* expression in the r3-derived cell, which in turn represses *krox20*, switching to an r4 fate. (Addison, *et al*., 2018; Kitazawa and Rijli, 2018). (Figure 11).

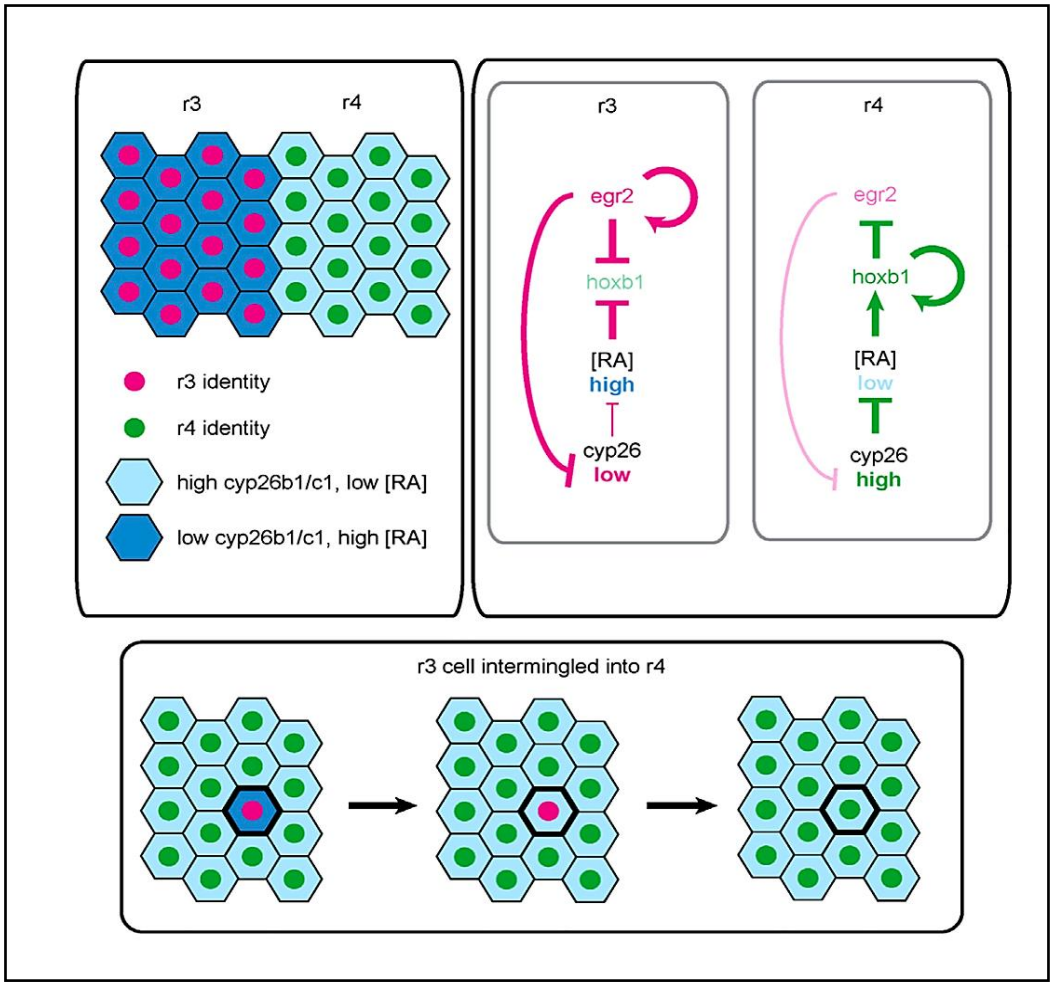


Figure 11: Model of *cyp26b*1, *cyp26c1*, *erg2* (*krox20*) and hoxb1 in cell identity switching (Addison. *et al*., 2018).

While the hindbrain subdivides into rhombomeres, the spinal cord remains unsegmented. However, both are patterned along the AP axis by Hox genes. **FGF and RA** also influence Hox gene expression and neurogenesis in the prospective spinal cord, as well as segmentation of the paraxial mesoderm into somites, suggesting that common signals control AP patterning in head and trunk. Unlike the hindbrain, however, the gradients of RA and FGF oppose one another in the trunk after gastrulation, with RA levels highest in the most anterior somites and FGF highest posteriorly. This difference

in the spatial deployment of at least one morphogen, RA, suggests a fundamental difference in the mechanisms by which these signals establish segmentation in the head and trunk. RA gradient between the head and trunk is also a crucial feature of recent models of somite formation and segmentation of the paraxial mesoderm in the trunk and tail. The somites, a segmental structure formed from the paraxial mesoderm in an anterior-to-posterior sequence, and their derivatives (*e.g.* vertebrae, axial muscles) are the most clearly segmented portions of the vertebrate body. Somite formation depends on a ‘‘clock and wave front’’ mechanism involving oscillating patterns of gene expression that move progressively toward the posterior end. An oscillating signal combines with a moving wave front from anterior to posterior that sets the boundaries of somites within the presomitic mesoderm. A working model at the molecular level suggests that the clock involves Notch signaling, the wave front involves RA and FGF signaling, and the two are interconnected through cycles of Wnt signaling (Palmeirim, *et al.* 1997; Dequeante and Pourquie 2008; Schilling, 2008).

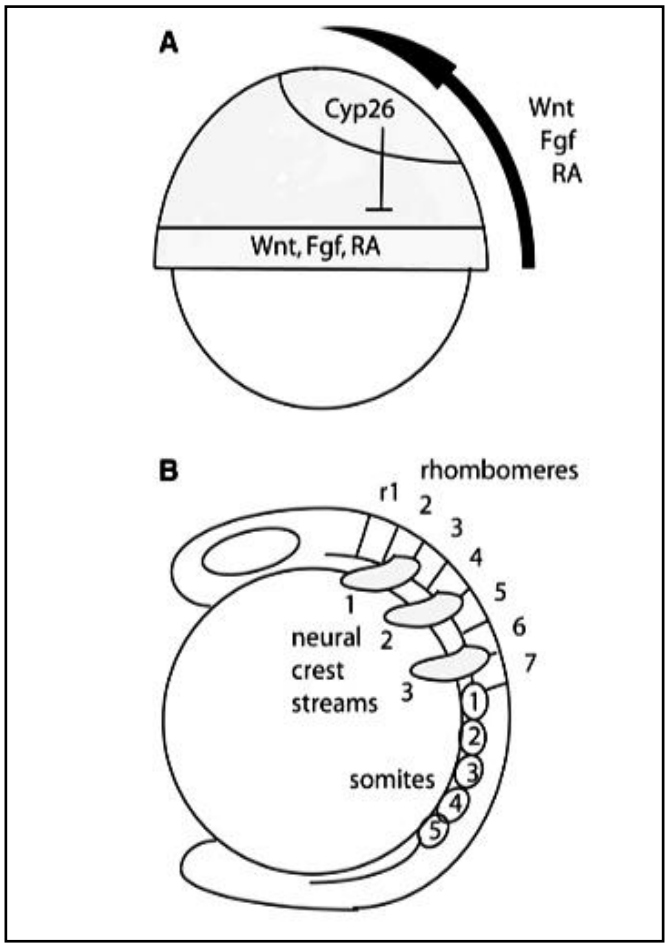


Figure 12: AP patterning signals and hindbrain segmentation. (A) Diagrams of gastrula-stage zebrafish embryos in lateral view, dorsal to the right, illustrating the proposed distribution of RA, Fgf, and Wnt signaling in the neural ectoderm (Kudoh et al. 2002). All three signals are made at the margin of the gastrula, the future posterior, and are thought to diffuse anteriorly to promote posterior development. (B) Segmentally organized structures in vertebrates, including rhombomeres of the hindbrain, streams of migrating cranial neural-crest cells, and mesodermal somites in the trunk, are schematized in a zebrafish at the five-somite stage (Thomas F. Schilling, 2008).

In zebrafish, the overall process of somitogenesis is similar to mammals, but the timing and the specification of myogenic progenitors show particular differences. In zebrafish, the first somite forms shortly after the end of gastrulation (Kimmel *et al*., 1995). The paraxial mesoderm develops from the cells around the edge of the early gastrula, which converge towards the dorsal side to form the paraxial mesoderm, adjacent to the axial mesoderm. As somitogenesis proceeds, the trunk begins to lift off the yolk and the tail extends. At the end of the first day of development, somitogenesis is complete, and the somitesare subdivided into sclerotome and myotome, where the myotome is already innervated and functional. A group of structurally related proteins, the myogenic basic helix-loop-helix (bHLH) family of transcription factors, has a major regulatory role in myogenesis (Emerson, 1993; Weintraub, 1993; Olson and Klein, 1994). Genes encoding four myogenic bHLH proteins have been found in mammals: *myoD*, *myogenin*, *myf-5* and *MRF4*. These genes encode proteins with a basic domain that mediates sequence-specific DNA binding and a helix-loop-helix domain that regulates dimerization. Several studies have demonstrated that *myoD* is sufficient to promote slow muscle formation from adaxial cells, and that it is required for fast muscle differentiation (Hammond CL, *et al*., 2006; Groves JA, *et al*., 2005). The expression of this gene occurs in several phases. In the first phase, which lasts from mid gastrula to just before somite formation, *myoD* is activated in rows of cells (termed ad axial cells) adjacent to the notochord or notochord precursor mesoderm. *MyoD* transcript is first detected in small triangular patches on each side of the embryonic shield at about 7-7.5 hpf. By 8.5 hpf, the patches of cells containing *myoD* transcript elongate and form a pair of narrow longitudinal rows, which lie on either side of the prospective notochord. At 10.5 hpf, prior to the formation of somites, the rows of *myoD*-expressing cells are two cells wide along the mediolateral axis, and increase in thickness along the dorso ventral axis from two cells (anteriorly) to six cells (posteriorly). The second phase of *myoD*

expression occurs between 10.5 and 12 hpf, the period in which the first 6-7 somites rapidly form. During this interval, *myoD* transcripts are simultaneously detected in a series of 5-7 faint bands of cells that project laterally from the longitudinal rows. Between 12 and 30 hours, two somites are formed per hour in a regular fashion (Hanneman and Westerfield, 1989; Westerfield, 1993). In a third phase of *myoD* expression, starting at 12.5 hours (7- to 8-somite stage), additional bands of cells express the gene. The shapes of the somites, and the locations of *myoD* expressing cells within these somites, differ along the AP axis and thus with respect to time since formation of each somite. The changes in *myoD* expression within the somite during these stages are coincident with changes in somite shape. *MyoD* expression persists at these sites during later stages of development, at least until 60 hours (Weinberg, *et al.*, 1996).

2. Aims of my PhD thesis

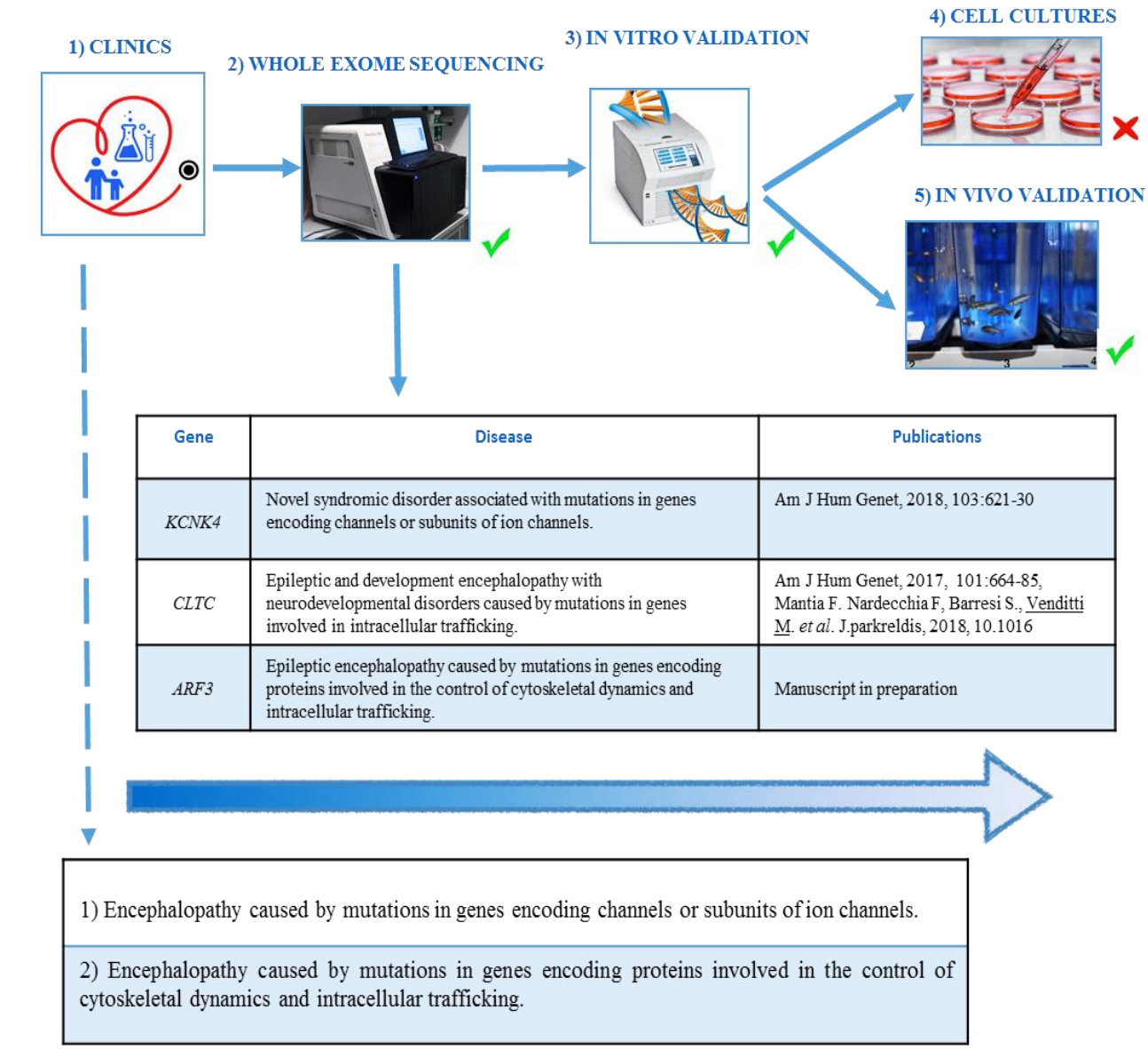


Figure 13: **Schematic representation of my PhD work**. The image represents the entire workflow used during my work. The images are numbered to indicate the different steps of the project. The symbol ✓ refers to the experimental activities that I have directly conduct during my PhD, while the symbol ✗ indicates the activities that are carried out by other colleagues. The part 1 (Clinics) was carried out by expert clinicians; in this part I only contributed to participate in the meetings planned for the clinical data discussion.

Since 2015, OPBG is dedicated to a multi-center and multi-disciplinary health-care and research project aimed at providing answers to more than 200 undiagnosed rare patients orphan of diagnosis and cure, for a better health care path and for novel treatment development. Since then OPBG committed to collect patients' clinical histories, along with a full pedigree including, whether available, affected second- or third-degree family members. For all the patients, phenotypic features are carefully summarized by molecular geneticists or a genetic counselor and recorded by means of Human Phenotype Ontology (HPO) terms for a standardized comparison of the phenotypes within internal cohorts and external databases. When needed, patients undergo a whole-exome sequencing (WES) analysis; different type of WES analysis (*e.g.* singleton. sib-pair, trio or most distantly related family members) are decided on a case-by-case level in agreement with the clinical and research staff. Careful examination of the WES results via internal bioinformatic research core and functional validations and disease mechanism discovery using multiple models represent crucial steps of the program. In this context, during my PhD work I primarily carried out research activity in the Unit of "*Molecular genetics and functional genomics*" (Dr. Marco Tartaglia). Here, the effectiveness of advanced sequencing techniques and sophisticated analysis in place have allowed identification of many new disease variants, pathology classification, disease genes and molecular bases of diseases in previously undiagnosed children world-wide. The global scope of my thesis was to highlight and investigate the variants and the underlying molecular/cellular alterations, principally involved in brain formation alterations causing a group of rare neurodevelopmental disorders in pediatric patients.

Therefore, as a part of this program, my PhD focused on: 1) **genomics analysis wet lab pipeline** through the WES approach. In this context, it was crucial to identify new disease-causing genes or genetic variants associated to a group of rare brain disorders. During the first part of my PhD activity, I became confident with the whole WES wet lab technology starting from the DNA preparation library (from patients' peripheral blood DNA samples) to sequencing reactions and row-data analysis (derived from bioinformatics core of our research group). The identification of new disease-causing genes or genetic variants allow me to begin with 2) ***in vivo* functional validation**, under

the supervision of the Dr. Antonella Lauri and Giulia Fasano (experts in zebrafish brain development), of a newly discovered candidate disease gene, *ARF3*, encoding for a GTP-binding protein involved in intracellular trafficking and likely causing a form of pediatric encephalopathy with microcephaly and overall body axis malformation. In this phase, I acquired a basic knowledge of the zebrafish model, including biology and physiology, genetics, husbandry and breeding techniques. Furthermore, as recommended by the D.lgs 26/2014, I obtained documented competence on using the zebrafish model for scientific purposes, released by the Animal Welfare Office here of OPBG. During my research activity on zebrafish, I focused on genotype-phenotype correlation in ARF3 zebrafish mutants generated by my collegues under Dr. Antonella Lauri guidance and carrying patients' specific mutations previously identified by WES. In detail, I investigated the impact of specific patients' mutations on the brain morphogenesis and body axes formation employing molecular biology, immunohistochemistry, in situ mRNA hybridization, light microscopy and statistical analysis.

3. Materials and Methods

3.1 Patients selection

Since 2015, meetings have been held twice a month, in total about 560 cases have been presented. Of these 321 undiagnosed patients, suspected for monogenic rare disorders or showing complex phenotypes, have been selected to make WES. Phenotypic features were carefully summarized by a genetic counselor and recorded by means of Human Phenotype Ontology (HPO) terms, for a standardized comparison of the phenotypes within our cohort and external databases, as well as for WES data interpretation. Which approach to use for whole exome sequencing (e.g. singleton, sib-pair, trio or most distantly related family members) was a case-by-case decision in agreement between clinical and research staff.

3.2 Whole exome sequencing (WES)

Exome sequencing was performed from the DNA of patients and, when available, of parents and affected and healthy relatives. Genomic DNA (gDNA) was extracted from peripheral blood circulating leukocytes using "QIAamp DNA blood mini kit" (QIAGEN cat. no. 51104) according to the manufacturer's instructions. After extraction, gDNA samples were analyzed and diluted using two serial fluorometric assays with Qubit™ dsDNA BR Assay Kit (Thermo Fisher cat.no. Q32853) to determinate the initial concentration of each sample. This step was critical for the successful preparation of input DNA at the required concentration to ensure optimal fragmentation. All gDNA samples were diluted with nuclease-free water to a final concentration of 100ng/µl. Then, each DNA was adjusted with nuclease-free water to a final concentration of 25 ng/µl.

Overview of the workflow

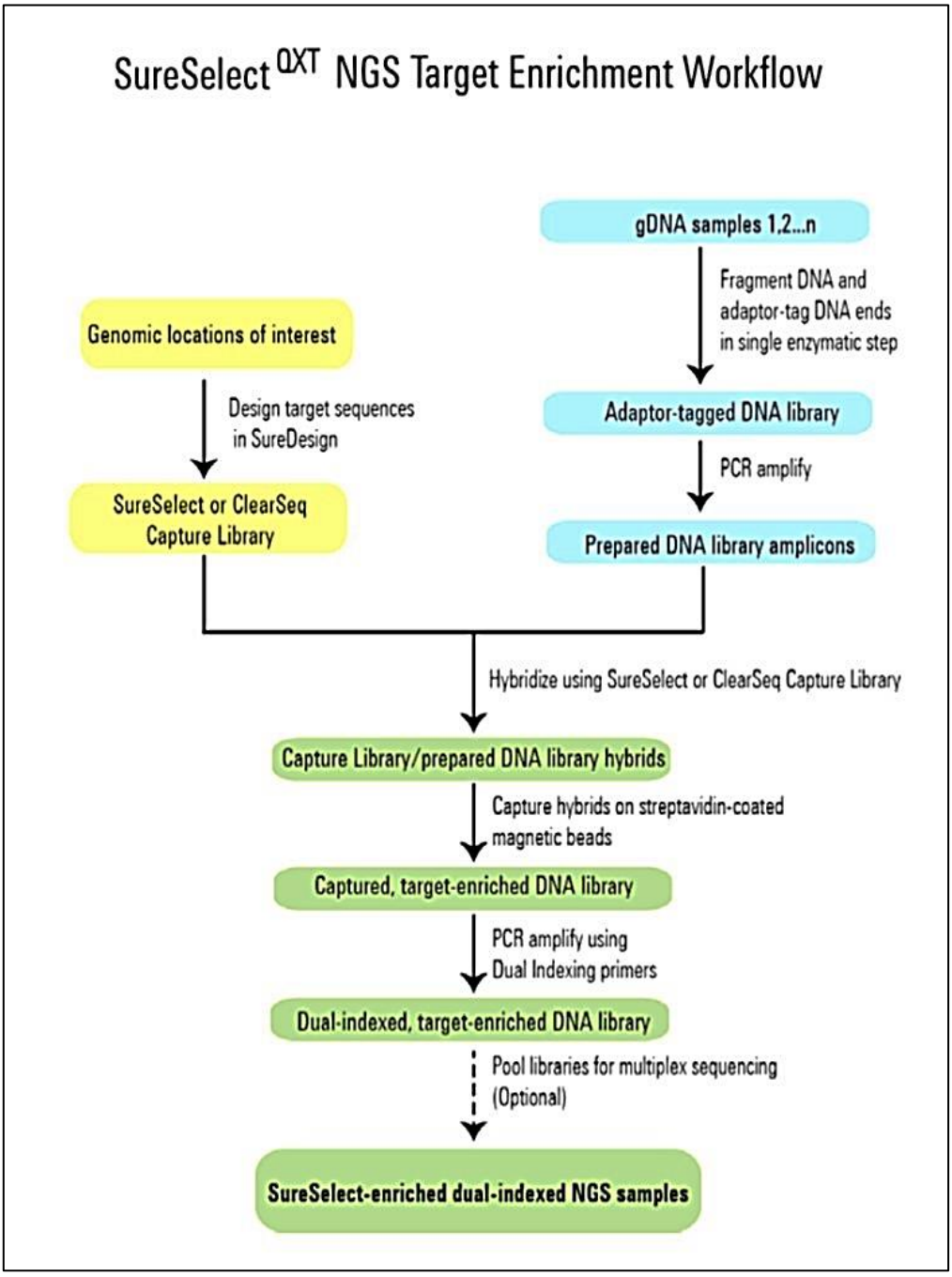


Figure 14: Overall target-enriched sequencing sample preparation workflow (© Agilent Technologies, Inc. 2014-2016, 2018).

3.2.1 Sample preparation

Exome enrichment and massive parallel sequencing were performed through enrichment kit SureSelect QXTTarget Enrichment for Illumina Multiplexed Sequencing 96rxn (Agilent Technologies cat.no.G9683B), using aqueous-phase hybridization capture system, and sequencing through platform NextSeq 500 (Illumina).

Step1. Fragmentation and adaptor-tag the genomic DNA samples

In this step, the gDNA was enzymatically fragmented and adaptors were added to ends of the fragments in a single reaction. This mechanism was mediated by the activity of transposon that was able to cut the genome into fragments of about 100 bp and simultaneously to insert adapter sequences.In this step were used specific SureSelect^{QXT} Reagent Kit components: SureSelect QXT Stop Solution, SureSelect QXT Buffer, SureSelect QXT Enzyme Mix ILM. The fragmentation reaction was made on ice using strip tubes:

-To each sample well were added 17µl of SureSelect QXT Buffer

-50 ng (2µl) of each DNA sample were added in sample well

- 2µl of SureSelect QXT Enzyme Mix ILM were added to each sample well and the mix was pipetted up and down 8 to 10 times to ensure complete transfer of the viscous solution to the well.

- Strip tubes were mixed thoroughly by vortexing at high speed for 20''

-The samples were placed in a thermal cycler with a specific program for DNA fragmentation that provided:

45°C -10 minutes, 4°C- 1 minute, 4°C- Hold.

-After 1' at 4°C, the samples were placed on ice and 32ul of 1X SureSelect QXT Stop Solution were added to each fragmentation reaction and all samples were incubated at room temperature for 1'.

Step 2. Purification the adaptor-tagged library using AMPure XP beads

The AMPure XP beads (Beckman Coulter) were incubated at room temperature for at least 30 minutes and fresh 70% ethanol was prepared.

-52 µl of the homogeneous bead suspension were added to each well containing the DNA samples and they were incubated for 5' at room temperature.

- Strip tubes were put on the magnetic stand at room temperature and two washes with 70% ethanol were performed to purify fragmented samples.
- Finally, each cleared supernatant was removed (approximately 10 µl), transferred to wells of a fresh strip tube and kept on ice; the beads were discarded.

Step 3. Amplification the adaptor-tagged DNA library

In this step, the adaptor-tagged DNA library was repaired and PCR-amplified.

- The appropriate volume of PCR reaction mix was prepared, as described below, on ice.

Reagent	Volume for 1 reaction
Nuclease-free water	25 µl
Herculase II 5× Reaction Buffer	10 µl
100 mM dNTP Mix (25 mM each dNTP)	0.5 µl
DMSO	2.5 µl
SureSelect QXT Primer Mix	1 µl
Herculase II Fusion DNA Polymerase	1 µl
Total	**40 µl**

Table 2: Preparation of pre-capture PCR Reaction mix.

-40 µl of the pre-capture PCR reaction mix were added to each 10 µl purified DNA library sample.

- The mix was incubated in the thermal cycler with the following program:

Segment Number	Number of Cycles	Temperature	Time
1	1	68°C	2 minutes
2	1	98°C	2 minutes
3	8	98°C	30 seconds
		57°C	30 seconds
		72°C	1 minute
4	1	72°C	5 minutes
5	1	4°C	Hold

Table 3: Thermal cycler program for pre-capture PCR.

Step 4. Purification of amplified library with AMPure XP beads

The AMPure XP beads were incubated at room temperature for at least 30 minutes and fresh 70% ethanol was prepared.

-50 μl of the homogeneous bead suspension were added to each well containing the DNA samples and they were incubated for 5' at room temperature.

- Strip tubes were put on the magnetic stand at room temperature and two washes with 70% of ethanol were performed to purify amplified samples.

- Finally, each cleared supernatant was removed (approximately 13 μl), transferred to wells of a fresh strip tube and kept on ice; the beads were discarded.

Step5. Assessing of library DNA quantity and quality

The quality check of samples was performed with D1000 Screen Tape (Agilent, cat.no.067-5582) and associated reagent kit (cat. no 067- 5583) to analyze the amplified libraries using the Agilent 2200 TapeStation system, an automated platform that carries out electrophoretic separation of nucleic acids. Samples were prepared as instructed in the user manual. 1μl of each amplified library DNA sample, diluted with 3 μl of D1000 sample buffer,was used for the analysis.

Tube strips were loaded into TapeStation with D1000 Screen Tape and loading tips, as instructed in the instrument user manual and the run was started. Eventually, the electropherogram was verified, the DNA fragments should have a profile between 245 and 325 bp, but it is possible to have variability.

Illustrative sample electropherograms are shown in figure 15.

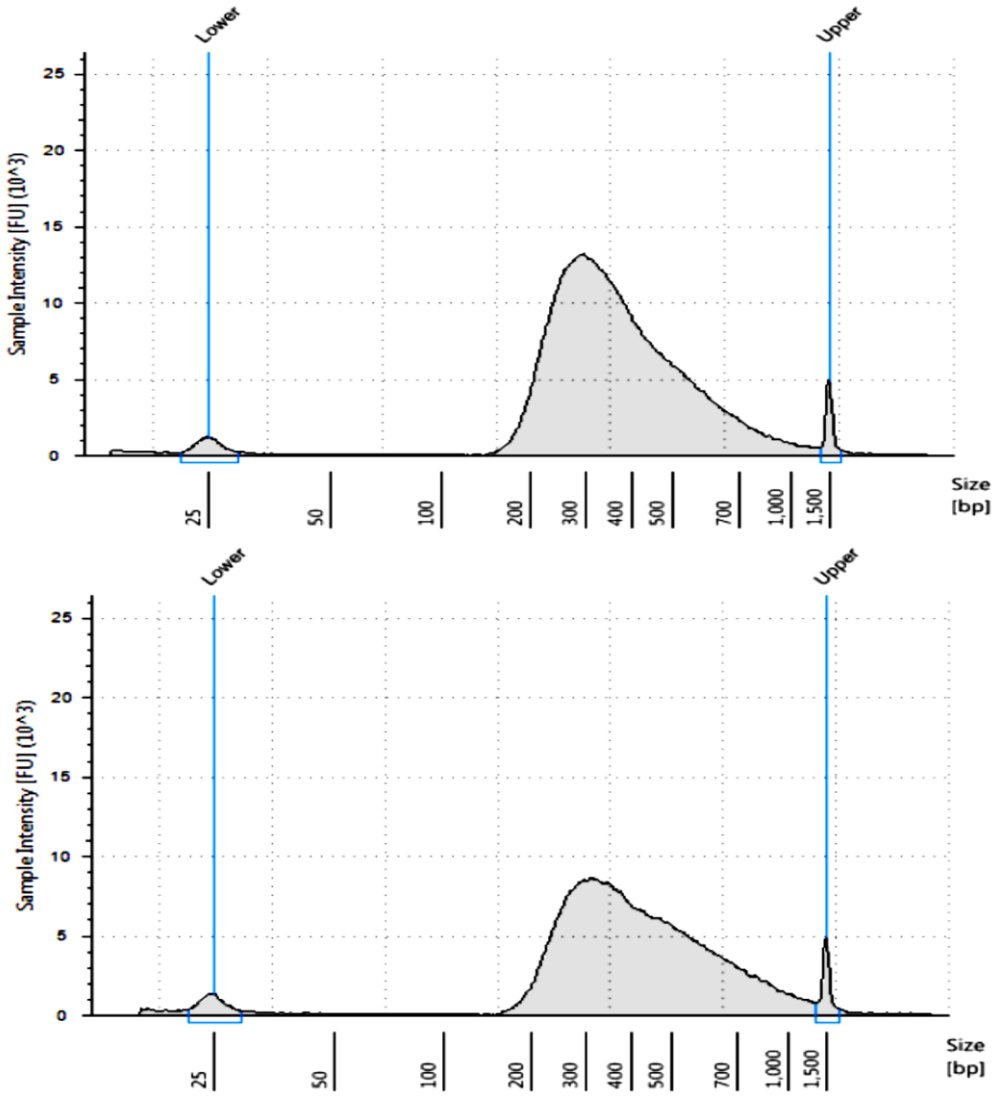


Figure 15: Representative sample electropherograms showing pre-capture analysis of amplified library DNA using the 2200 TapeStation with a D1000 ScreenTape. A peak DNA fragment size significantly less than 245 bp may indicate too little gDNA in the fragmentation reaction and may be associated with increased duplicates in the sequencing data. In contrast, a peak DNA fragment size significantly greater than 325 bp may indicate too much gDNA in the fragmentation reaction and may be associated with decreased percent-on-target performance in sequencing results.

3.2.2 Hybridization and Capture

Step 1. Aliquot prepared DNA samples for hybridization

The gDNA library was quantified to value the exact concentration of gDNA that need to be used for the hybridization step. The quantification was made with Qubit™ dsDNA BR Assay Kit. The amount of prepared gDNA library used in the hybridization reaction varies according to the size of the Capture Library. For libraries with a size greater than 3 Mb, as in our case, up to 1500 ng were required. We had placed up to 1500 ng (amount 750 to1500 ng) of prepared DNA library into the designated well and then brought the final volume to 12 μl, using nuclease-free water according to protocol.

Step 2. Hybridization of DNA samples to the Capture Library

In this step was used a specific SureSelect QXT Reagent Kit components: SureSelect Fast Hybridization Buffer, SureSelect QXT Fast Blocker Mix, SureSelect RNase Block and Capture Library (Clinical Research Exome SSELXTV2, #16, Agilent cat.no 5190-9491).

-To each adaptor-tagged DNA sample well, 5 μl of SureSelect QXT Fast Blocker Mix were added.

-The mix was pipetted up and down 8 to 10 times and was vortexed at high speed for 5’’.

-Sample wells containing 17 μl of prepared DNA Fast Blocker mixture, were transferred to the thermal cycler and was started the following program, using a heated lid.

Segment Number	Number of Cycles	Temperature	Time
1	1	95°C	5 minutes
2	1	65°C	10 minutes
3	1	65°C	1 minute (**PAUSE cycler here**)
4	60	65°C	1 minute
		37°C	3 seconds

Table. 4: Thermal cycler program for Hybridization.

-During PCR, was prepared a 25% solution of SureSelect RNase Block, containing 0.5 µl of SureSelect RNase Block and 1.5µl of nuclease-free water per sample (1 part RNase Block:3 parts water).

-The Capture Library Hybridization Mix appropriate for >3Mb Capture Libraries was prepared according to the table 5.

-All reagents were combined at room temperature by pipetting and vortexing at high speed.

-13µl of Capture Library Hybridization mix was added to each sample.

Reagent	Volume for 1 reaction
25% RNase Block solution (from step 3)	2 µl
Capture Library ≥3 Mb	5 µl
SureSelect Fast Hybridization Buffer	6 µl
Total	**13 µl**

Table 5: Preparation of Capture Library Hybridization Mix for Capture Libraries >3 Mb.

Step 3. Prepare streptavidin-coated magnetic beads for DNA hybrid capture

For hybrid capture protocol were used SureSelect Binding Buffer, SureSelect Wash Buffer 2 and SureSelect Wash Buffer 1 in addition to the streptavidin-coated magnetic beads.

-The DynabeadsMyOne Streptavidin T1 magnetic beads (Thermo Fisher Scientific) were vigorously resuspended on a vortex mixer.

-For each hybridization sample, were added 50 µl of the resuspendedbeads.

-The beads were washed with 200µl of SureSelect Binding Buffer for three times.

-Finally, the beads were resuspended in200 µl of SureSelect Binding Buffer.

Step 4. Capture of the hybridized DNA using streptavidin-coated beads

-After the hybridization step was completed and the thermal cycler had reached the 65°C hold step, the samples were transferred to room temperature.

- The hybridizedsamples were maintained at room temperature, on a vortex mixer, for 30’ while to transfer the entire volume (approximately 30 μl) of each hybridization mixture to wells containing 200 μl of washed streptavidin beads.

- During the 30’ capture, Wash Buffer 2 was pre-warmed at 65°C.
- The samples were incubated for 10' at 65°C.
- Strip tubes containing the samples were put in the magnetic separator at room temperature for 1', when the solution was clear, the supernatant was removed and discarded.
- This step was repeated 3 times.
- At the end of the washes, after removing the supernatant from the final wash, tubes with samples were put on the magnetic stand to remove any remaining wash buffer droplets.
- 23 μl of nuclease-free water was added to each sample well. In this step captured DNA was retained on the streptavidin beads.

3.2.3 Indexing and Sample Processing for Multiplexed Sequencing

Step 1. Amplification of captured libraries to add index tags

In this step, the SureSelect-enriched DNA libraries were PCR amplified using the appropriate pair of dual indexing primers.

-The indexing amplification reaction was prepared for each DNA library.

-The appropriate index assignments were determined for each sample using a different indexing primer combination for each sample to be sequenced in the same lane. Sequences of the index portion of the P7 and P5 indexing primers used to amplify the DNA libraries in this step, for NextSeq platform, were annotated in the following tables:

The nucleotide sequence of each SureSelectQXT P7 / IndexesSureSelectQXT P5 Indexes 13 to 20

Index Number	Sequence
P7 Index 1 (P7 i1)	TAAGGCGA
P7 Index 2 (P7 i2)	CGTACTAG
P7 Index 3 (P7 i3)	AGGCAGAA
P7 Index 4 (P7 i4)	TCCTGAGC
P7 Index 5 (P7 i5)	GTAGAGGA
P7 Index 6 (P7 i6)	TAGGCATG
P7 Index 7 (P7 i7)	CTCTCTAC
P7 Index 8 (P7 i8)	CAGAGAGG
P7 Index 9 (P7 i9)	GCTACGCT
P7 Index 10 (P7 i10)	CGAGGCTG
P7 Index 11 (P7 i11)	AAGAGGCA
P7 Index 12 (P7 i12)	GGACTCCT

Index Number	Sequence
P5 Index 13 (P5 i13)	GCGATCTA
P5 Index 14 (P5 i14)	ATAGAGAG
P5 Index 15 (P5 i15)	AGAGGATA
P5 Index 16 (P5 i16)	TCTACTCT
P5 Index 17 (P5 i17)	CTCCTTAC
P5 Index 18 (P5 i18)	TATGCAGT
P5 Index 19 (P5 i19)	TACTCCTT
P5 Index 20 (P5 i20)	AGGCTTAG

-25 µl of the PCR reaction mix, prepared according to table 3, was added to the 23µl of bead suspension in each sample well.

-1 µl of the appropriate P7 dual indexing primer (P7 i1 to P7 i12) was added to each PCR reaction mixture well (only one of the twelve possible P7 primers to each reaction well).

-1 µl of the appropriate P5 dual indexing primer (P5 i13 to P5 i20) was added to each PCR reaction mixture well (only one of the eight possible P5 primers to each reaction well).

Reagent	Volume for 1 reaction
Nuclease-free water	13.5 µl
Herculase II 5× Reaction Buffer	10 µl
100 mM dNTP Mix (25 mM each dNTP)	0.5 µl
Herculase II Fusion DNA Polymerase	1 µl
Total	**25 µl**

Table 6: Preparation of post-capture PCR Reaction mix

-PCR reaction was mixed well by pipetting to ensure the beads were fully resuspended, then were transferred to a thermal cycler with PCR amplification program that provided:
-98°C 2', 1X
-98°C 30''
-58°C 30''/10X
-72°C 1'
-72°C 5', 1X
-4°C Hold

- When the PCR amplification program was completed, the streptavidin-coated beads were removed by placing the strip tube on the magnetic stand at room temperature. After 2 minutes, time to obtain a clear solution, each supernatant (approximately 50 µl) was removed and transferred to wells of a fresh strip tube. The beads were discarded.

Step 2. Purification of amplified captured libraries using AMPure XP beads

The AMPure XP beads were incubated at room temperature for at least 30 minutes and fresh 70% ethanol was prepared.
-60µl of the homogeneous bead suspension were added to each well containing the DNA samples and they were incubated for 5' at room temperature.
-Strip tubes were put on the magnetic stand at room temperature and two washes with 70% of ethanol were performed to purify fragmentation samples.
-25µl of nuclease-free water were added to each sample well.
-Finally, each cleared supernatant was removed (approximately 25 µl), transferred to wells of a fresh strip tube and kept on ice; the beads were discarded.

Step 3. Assess indexed library DNA quantity and quality

The quality check of samples was performed with High Sensitivity D1000 Screen Tape (cat.no. 5067-5584) and reagent kit (cat.no. 5067-5585) to analyze the amplified libraries using the Agilent 2200 Tape Station. Tape Station samples were prepared as instructed in the user manual. 2µl of each indexed DNA sample, diluted with 2 µl of High Sensitivity D1000 sample buffer, were used for the analysis.

Tube strips were loaded intoTape Station with the High Sensitivity D1000 Screen Tape, and loading tips, as instructed in the instrument user manual and the run was started. Eventually, the electropherogram was verified, the DNA fragments should have a profile between 325 and 450bp, but it is possible to have variability.
Sample electropherogram are shown in figure 16.

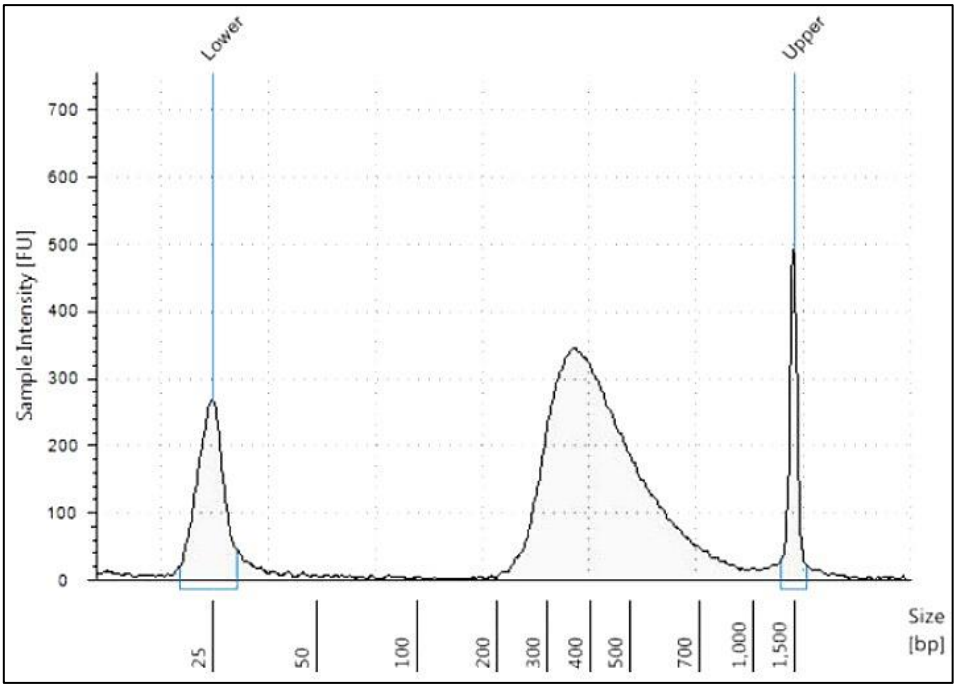


Figure 16: Post-capture analysis of amplified indexed library DNA using the 2200 TapeStation with a High Sensitivity D1000 Screen Tape.

Step 4. Pool samples for multiplexed sequencing

The number of indexed libraries that may be multiplexed in a single sequencing run was determined by the output specifications of the platform used, together with the amount of sequencing data required for the research design. In our case, according with sequencing kit used, output is of 80 Gb, target region is 67 Mb, target region Clinical Research Exome is 54 Mb. Therefore we considered these value to calculate the number of samples to perform a single sequencing, in particular: # samples = 80 Gb/ (54 Mb x 180X (which represents the desired depth in our experiment).The libraries were combined such that each index-tagged sample is present in equimolar amounts in the pool. For each library, the formula below was used to determine the amount of indexed sample to use:

$$\text{Volume of Index} = \frac{V(f) \times C(f)}{\# \times C(i)}$$

(Where *V(f)* is the final desired volume of the pool, *C(f)* is the desired final concentration of all the DNA in the pool,# is the number of indexes, and *C(i)* is the initial concentration of each indexed sample).

Step 5. Preparation of sequencing samples (denaturing and diluting of libraries for the NextSeq 500).

In this step, the library was denatured combining a precise volume of libraries, which depends on starting library concentrations, and freshly 0.2N NaOH in a microcentrifuge tube. In our case, we used 5µl of 4nM of starting library concentration that was denatured with 5 µl of 0.2 N of NaOH. Mix obtained was incubated for 5' at room temperature to denature library into single strands. Then the library was diluted at 20 pM adding 990 µl of pre-chilled HT1 (Hybridization Buffer provided of the NextSeq 500 kit (Illumina cat.no. 20024908). The optimal seeding concentration for SureSelectQXT target-enriched libraries varies according to the sequencing platform, run-type, and Illumina kit version. In our case, we have optimized using 1.6 pM of diluted library. According to this choice, we have diluted 20 pM library to 1.6 pM using pre-chilled HT1 in a final volume of 1.3 ml. Seeding concentration and cluster density may also need to be optimized based on the DNA fragment size range for the library and the desired output and data quality. After the denaturation and dilution of the library, the cluster was amplified using the appropriate Illumina Paired-End Cluster Generation Kit for kit configurations compatible with the recommended read length plus reads for the SureSelectQXT 8-bp dual indexes. The mechanism of Sequencing Paired End Libraries with Dual Index Read is showed in figure 17.

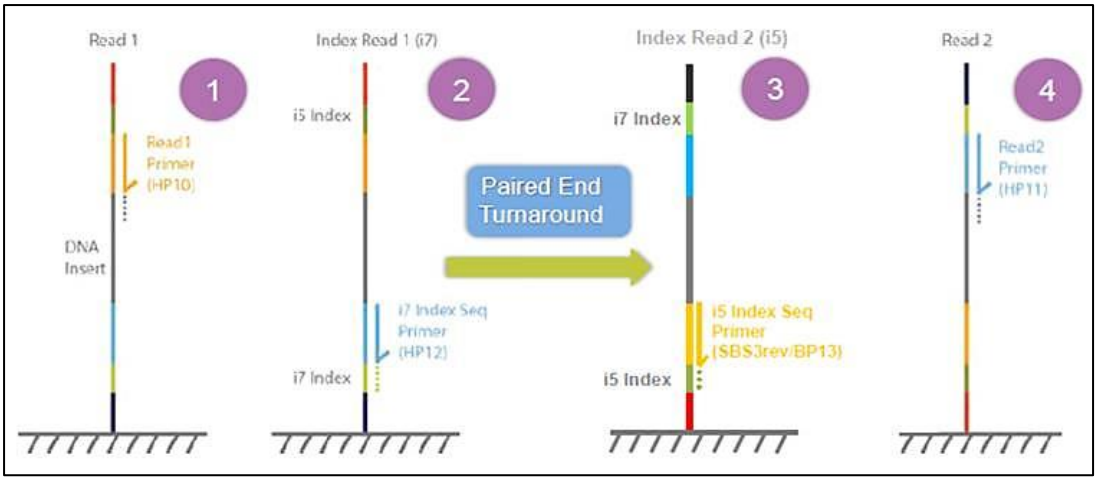


Figure 17: **Dual-Indexed Sequencing on a Paired-End Flow Cell. Read 1** follows the standard Read 1 sequencing protocol using SBS reagents. The Read 1 sequencing primer is annealed to the template strand during the cluster generation step. The Read 1 product is removed and the **Index 1 (i7)** sequencing primer is annealed to the same template strand. Following Index Read preparation, the Index 1 (i7) Read performs up to 20 cycles of sequencing. **The Index 1 (i7) Read product**is removed and the template anneals to the grafted **P5 primer** on the surface of the flow cell. The run proceeds through an additional 7 chemistry-only cycles, followed by up to 20 cycles of sequencing. The **Index Read product**is removed and the original template strand is used to regenerate the complementary strand. Then, the original template strand is removed to allow hybridization of the **Read 2 sequencing** primer. Read 2 follows the standard paired-end sequencing protocol using SBS reagents.

To do this step, it was important to refer to the manufacturer's instructions, using the modifications for use of the SureSelectQXT Read Primers with the Illumina Paired-End Cluster Generation Kits. See Table 7, NextSeq 500.

Platform	Run Type	Read Length*	SBS Kit Configuration	Chemistry	Seeding Concentration
HiSeq 2500	Rapid Run	2 × 100 bp	200 Cycle Kit	v2	10–13 pM
HiSeq 2500	High Output	2 × 100 bp	4 x 50 Cycle Kit†	v3	10–13 pM
HiSeq 2500	High Output	2 × 100 bp	250 Cycle Kit	v4	13–16 pM
HiSeq 2000	All Runs	2 × 100 bp	4 x 50 Cycle Kit†	v3	7–11 pM
HiSeq 2000	All Runs	2 × 100 bp	250 Cycle Kit†	v4	10–14 pM
MiSeq	All Runs	2 × 100 bp	300 Cycle Kit	v2	10–13 pM
MiSeq	All Runs	2 × 76 bp	150 Cycle Kit	v3	14–19 pM
NextSeq 500/550	All Runs	2 × 100 bp	300 Cycle Kit	v2	1.7–2.0 pM
HiSeq 3000/4000	All Runs	2 × 100 bp	300 Cycle Kit	v1	200 pM

Table 7: Illumina Kit Configuration Selection Guidelines.

Step 6. Using the SureSelectQXT Read Primers with Illumina's Paired-End Cluster Generation Kits

To sequence the SureSelectQXT libraries on Illumina's sequencing platforms, weneeded to use the following custom sequencing primers, provided in Sure Select QXT Library Prep Kit:

- Sure Select QXT Read Primer 1
- Sure Select QXT Read Primer 2
- Sure Select QXT Index 1 Read Primer
- Sure Select QXT Index 2 Read Primer

These SureSelectQXT custom sequencing primers are provided at 100 uM and must be diluted in the corresponding Illumina primer solution, using the platform-specific instructions below:

Table 8: NextSeq 500/550 High-Output v2 Kit custom sequencing primer preparation

Sequencing Read	Volume of SureSelect^{QXT} Primer	Volume of Illumina Primer	Total Volume	Final Cartridge Position
Read 1	3.9 µl SureSelect QXT Read Primer 1 (brown cap)	1296.1 µl BP10 (from well 20)	1.3 ml	well 7
Read 2	4.2 µl SureSelect QXT Read Primer 2 (black cap)	1395.8 µl BP11 (from well 21)	1.4 ml	well 8
Index 1+ Index 2	6 µl SureSelect QXT Index 1 Read Primer (clear cap) + 6 µl SureSelect QXT Index 2 Read Primer (purple cap)	1988 µl BP14 (from well 22)	2 ml	well 9

Table 9: NextSeq 500/550 Mid-Output v2 Kit custom sequencing primer preparation

Sequencing Read	Volume of SureSelect^{QXT} Primer	Volume of Illumina Primer	Total Volume	Final Cartridge Position
Read 1	2.7 µl SureSelect QXT Read Primer 1 (brown cap)	897.3 µl BP10 (from well 20)	0.9 ml	well 7
Read 2	3.3 µl SureSelect QXT Read Primer 2 (black cap)	1096.7 µl BP11 (from well 21)	1.1 ml	well 8
Index 1+ Index 2	4.8 µl SureSelect QXT Index 1Read Primer (clear cap) + 4.8 µl SureSelect QXT Index 2 Read Primer (purple cap)	1590.4 µl BP14 (from well 22)	1.6 ml	well 9

After preparing custom sequencing primers, they are added to the preloaded single-use reagent cartridge that contains all reagents for cluster generation, sequencing, and paired-end chemistry. Into cartridge was also loaded pool of indexed libraries (see figure 18).

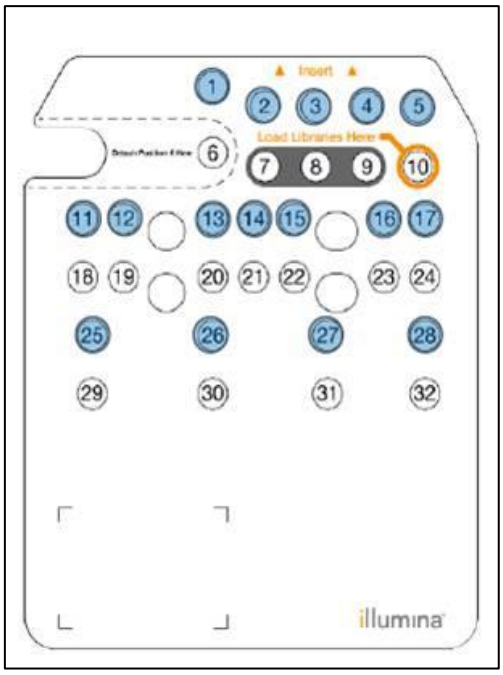


Figure 18: NextSeq 500 Reagent Cartridge.

Step 7. Set up the sequencing run and trim adaptors from the reads

It was necessary to refer to Illumina protocols to set up custom sequencing primer runs, using the additional guidelines provided by Illumina. For the NextSeq 500 platform the sequencing runs were set up using the specific provided settings:

Run Segment	Cycle Number
Read 1	100
Index 1 (i7)	8
Index 2 (i5)	8
Read 2	100

For the NextSeq platform, Cycle Number and custom sequencing primer settings can be specified on the Run Configuration screen of the instrument control software interface.

3.2.4 Clusters generation and sequencing

For cluster generation, Illumina technology uses solid-phase amplification producing clonally amplified clusters from fragment templates on a glass slide called *flow cell* to which were anchored two types of oligo complementary to the nucleic adapters (P5 and P7) at the end of the library fragments. The DNA strands were cloned amplified on the surface of the flow cell through a process called bridge amplification (Bridge-PCR). [The adapters are complementary to the oligos of flow cell anchorage, and the filament is polymerized starting from these oligos. The original filament dissociates and is washed away, while the neo-synthesized filament, covalently associated with the slide, folds back on itself by binding the second type of oligo in the flow cell, forming a bridge structure that will act as a mold for amplification]. Each strand was amplified on average 5000 times, generating millions of clusters. Illumina platform, taking advantage of the principle of the "sequencing by synthesis" with fluorescent detection, it produces sequence reads of several hundreds of bplenght from tens of millions of simultaneously amplified DNA fragments. For sequencing protocol, a cyclic reversible termination (CRT) approach was used followed by fluorescence imaging. After nucleotide incorporation, the unincorporated nucleotides were washed away followed by a cleavage step, which removes the terminating/inhibiting group and the fluorescent dye (Metzker, 2010). Raw image files produced were processed by software provided by Illumina, for base calling, with default parameters, generating reads [Reads are the short fragments of sequences obtained during sequencing, each of them representing one of the hundreds of millions of DNA molecules present in the library]. The Illumina sequencer output file is a FASTQ file that was processed by our bioinformatics, through a chain of bioinformatics tools, involved in the alignment, call of variants, annotation, filtering and prioritization of variants.

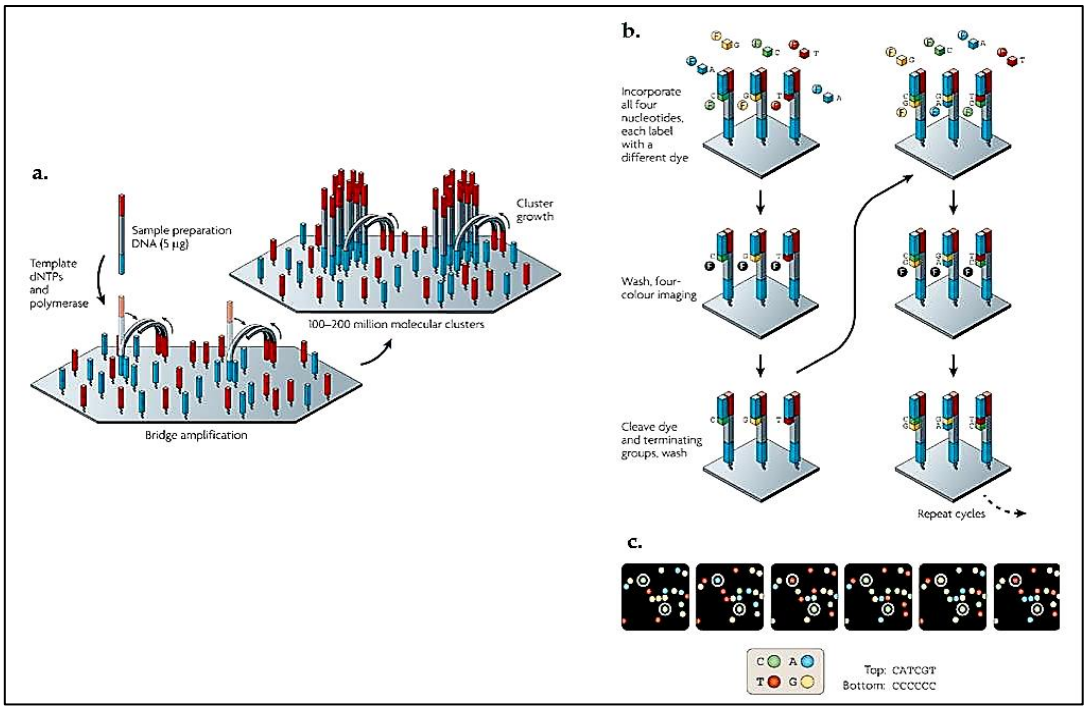


Figure 19: Cluster generation through solid-phase amplification and sequencing. a. Solid-phase amplification is composed of two basic steps: initial priming and extending of the single-stranded, single-molecule template, and bridge amplification of the immobilized template with immediately adjacent primers to form clusters. **b.** The four-colour cyclic reversible termination (CRT) method uses 3′-O-azidomethyl reversible terminator chemistry using solid-phase-amplified template clusters. Following imaging, a cleavage step removes the fluorescent dyes and regenerates the 3′-OH group. **c.** The four-colour images highlight the sequencing data from two clonally amplified templates (Metzker, 2010).

3.3 Exome sequencing data analysis

-Alignments and Variants calling

WES data analysis was performed using an in-house implemented pipeline (Niceta*et al.*, 2015; Kortüm*et al.*, 2015). Reads were aligned to the human reference genome (UCSC GRCh37/hg19) using the BWA-MEM *(Burrows-Wheeler Aligner*) (Li and Durbin, 2009). This is an algorithm that can convert FASTQ data intoBAM and SAM files. Through the SAM file you can see the alignment of the reads to the genome. Presumed PCR duplicates were removed using Picard's Mark Duplicates (http://picard.sourceforge.net.). The Genome Analysis Toolkit (GATK, McKenna *et al.*, 2010) was used for realignment of sequences encompassing indels and for base quality recalibration. Single nucleotide variants and small indels were identified by means of the GATK Haplotype Caller and/or Unified Genotyper algorithm (Kortüm *et al.*, 2015).

Variants were reported in a VCF (Variant Call Format) file. Two parameters were considered important indicators of good alignment: coverage and depth. Coverage means the percentage of target bases that have been sequenced for a given number of times, the greater the coverage, the greater the percentage of the exome that has been sequenced. Mainly, as the mean depth of target coverage, was considered the fraction of all target bases achieving 2x, 10x, 20x, 50x or greater coverage to evaluate the goodness of sequencing. The depth indicates the number of reads aligned to a specific region of the exome.

-Annotation, filtering and prioritization of variants

Annotation was the process by which each variant was associated with specific information obtained from public databases, such as the gene involved, the nature of the variant, its predicted effect and its frequency population. High-quality variants were filtered by comparing them to specific allelic frequency databases (*dbSNP 150* and *GnomAD V.2.0.1*), to deem the new and clinically associated variants un-known or very low frequency (*MAF*, Minor Alleic Frequency, <0.1%). Similarly, variants that have a lower frequency were considered 1% of the in-house database, which consists of the frequency data of approximately 1300 WES carried out in OPBG. Another selective phase on candidate variants was the consultation of genotype-phenotype binding databases, such as OMIM and MalaCards, in which were collected information about genetic disorders in humans. The use of these databases was very important to have a wider understanding of candidate variants and their genes, and to filter them, allowing to discard variants that do not have a correlation with the clinical phenotype under consideration.

At the same time, there was predictive analysis of the functional impact of the variant. The functional annotation of the variants was made using programs such SnpEff toolbox and dbNSFP (*SnpEff*v.4.3 e *dbNSFP*v.2.9) that were used to predict the functional impact of variants (synonymous, missense, nonsense, frame-shift, intronic and UTR) (Kortüm *et al.*, 2015) and to annotate variants and genes using information retrieved from selected public databases containing data on population frequency (*dbSNP* - Database of Single Nucleotide Polymorphisms, 1000 Genome Project and ExAC - Exome Aggregation Consortium), sequence conservation (phastCONS and PhyloP), cell pathways (*KEGG* - Kyoto Encyclopedia of Genes and Genomes), involvement in pathogenic mechanism (*HGMD* – Human Gene Mutation Database,

OMIM - Online Mendelian Inheritance in Man), gene function (*GO* - Gene Ontology), animal disease models (*MGD* - Mouse Genome Database), gene expression (*GXD* - Gene Expression Database, etc.), and from in-house databases. The functional impact has been predicted via Combined Annotation Dependent Depletion (CADD) v1.3 and M-CAP v.1.0 (Mendelian Clinically Pathogen Applicableicity). CADD is a framework that integrates multiple annotations into a single metric and is able to analyze any SNP within the genome and provide an accurate prediction impact of the variant by assigning a score. M-CAP is a tool that operates radially, associating with each information known in the literature and predictive data functional impact, providing a pathogenic score variant.

3.4 Zebrafish husbandry

Zebrafish were kept under standard conditions at the zebrafish facility at OPBG (Westerfield, 2000). Fish were allocated in a circulating system that continuously filters and aerates the system water to maintain the water quality required for a healthy environment. The tank temperature was maintained between 27-28.5 °C with lighting conditions of 14-hour light and 10-hour dark cycle (14/10). In the facility fine mechanical filter, carbon filtration and UV disinfection filter ensure water quality. Here the pH of the system water is checked daily and maintained between 6.8 and 7, the conductivity of the system is mantained between 250 to 500 with ocean salt (30gr/L). The fish were fed three times daily with a varied diet consisting of dry food and live shellfish (*Artemia salina*). Fish matured sexually and were employed for spawing at around three months. All experimental procedures described here have been conducted in accordance with valid regulations and guidelines and approved by the Italian Ministry of Health. We performed all the experiments on embryos at the stage of 15, 24, 48, 72- and 96-hours post fertilization (hpf).

3.5 Zebrafish breeding, embryos collection and staging

For spawning purposes, we put adult zebrafish males and females in the mating container with a separation between then and left them overnight. Females were told

apart from male fish by their bigger and whiter abdomen, while males were identified by their slender, more streamlined appearance, as well as their somewhat darker blue stripes with a pink tint in between. When the lights were on the following morning, males and females were left together in the mating container. To prevent the zebrafish from eating embryos, a spawning mesh placed in the middle of the container (Figure 20) allowed fertilized eggs to drop at the bottom of the mating container, away from the adult fish. The fish were left to mating undisturbed for about two hours. After breeding, the parents were returned to their original tanks in the rack of the facility.

Figure 20: mating tank

The embryos were collected in Petri dishes containing E3 medium (5M NaCl, 1M KCl, 1M $CaCl_2$, 1M $MgSO_4$, 0,5 ml Methylene-blue stock solution). Embryos were collected at the required developmental stages according to the need (see Kimmel *et al., 1995* for embryonic stage description).

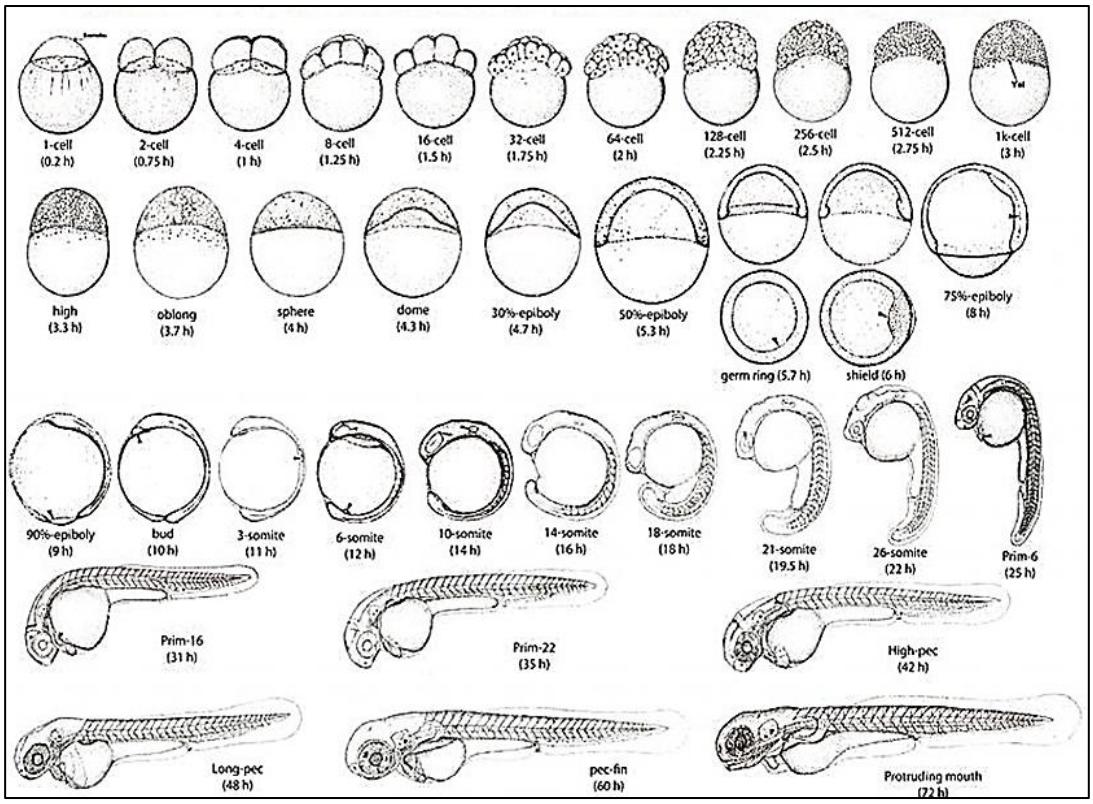


Figure 21: Zebrafish developmental stages. Modified from Kimmel et al., 1995.

3.6 Whole-mount in Situ Hybridization of Zebrafish Embryos

3.6.1 Sample preparation: early zebrafish embryos

Zebrafish Embryos were collected at the required developmental stages, 15hpf. (Kimmel et al., 1995). It was chosen this phase of development because this is the phase in which the gastrulation has ended, and the segmentation process has taken place. We can appreciate the complete formation of somites along the anterior-posterior axis and the development of major zebrafish brain structures. In this developmental stage, embryos are surrounded by an acellular envelope called chorion. For experiments, the chorion was removed manually using sharp forceps, making particular attention when opening the chorion because, at the stage of 15 hpf, embryos are easily damaged (Bonsignorio D., *et al*, 1996).

I followed this protocol:

- FIXATION: a 4% solution of paraformaldehyde (PFA) made in PBS (Phosphate Buffered Saline) overnight at 4°C or 3 hours at room temperature;
- WASHES: 2x for 5 minutes washes with PBSTween (PBT) 20 0.1% each at room temperature, 1x for 10 min washes with different dilutions of methanol (MetOH) in PBT (75%, 50% and 25%), 2x for 5 min in MetOH 100%.
Embryos were stored in 100% MetOH at-20°C until used for the future application.

3.6.2 RNA extraction and cDNA synthesis

Whole zebrafish embryos at different stages (6, 10, 15, 24, 48, 72 and 96 hpf) were frozen dry and were stored at-80°C at least for 24 h to preserve RNA and obtained a good yield. The RNA extraction was performed with RNeasy Plus Mini Kit extraction (Qiagen cat.n.74134/74136) according to the protocol. The embryos were defrosted in ice and were homogenized with a pestle in the appropriate volume of Buffer RLT Plus that was defined by the weight of the samples (<20mg add 350µl of Buffer RLT Plus, 20-30 mg add 600µl of Buffer RLT Plus). The lysate obtained was transferred to the QIAshredder column and centrifuged 3'at maximum speed, the supernatant in the column was eliminated while the flow-through was preserved. Then 1 volume (usually 350µl or 600µl) of 70% ethanol to the flow-through was added and mixed well. The samples were transferred to a Rneasy spin column and centrifuged for 15'' at 8000 x g, the flow-through was then discarded. Samples were then treated with 80ul of DNasi mix, composed by 10ul of Dnasi stock solution and 70ul of RDD buffer, and incubated at room temperature (15°-25°C) for 15'. Several washes of the column were performed afterwards: one wash with 700ul of Buffer RW1 to the RNeasy spin column and two washes with 500µl of Buffer RPE to the RNeasy spin column; the flow-through was discarded. At the end RNeasy spin columns was placed in a new 1.5 ml collection tube and were added 30-50µl of RNase-free water and the samples were centrifuged for 1' at 8000 x g to elute the extracted and concentrated RNA.
1µl of the the total RNA extracted was analyzed on Thermo Scientific™ NanoDrop 2000 spectrophotometer. The integrity of RNA was evaluated by 260/280 and 260/230 ratios, generally a ratio of ~2.0 is generally accepted as "pure" for RNA. The integrity of RNA was also evaluated using agarose gel electrophoresis.
Next, for cDNA synthesis, 1 µg of RNA was reverse transcribed using QuantiTect® Reverse Transcription Kit (Qiagen cat.no. 205311) according to manufacturer's

instructions using both oligo (dT) and random primers (RT Primer mix). In the first step a treatment of the RNA with a genomic DNA elimination reaction was performed, samples were incubated for 2' at 42°C and placed then immediately on ice. A reverse-transcription master mix was added next and the samples were incubated for 15' at 42°C and for 3' at 95°C to inactivate Quantiscript Reverse Transcriptase. The produced cDNA was aliquoted and stored at -20°C.

3.6.3 PCRs of zebrafish fragments of *krox 20* and *myoD*

Primers for the gene of interest, *krox20* and *myoD* were designed with a TM of around 60/64°C, and with a specific size around 600-900 pair of bases, that is an optimal size to generate RNA probes for In situ hybridization (ISH). The quality of the primers was assessed with: NEB Tm Calculator (https://tmcalculator.neb.com/), NCBI Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) and Primer3 input (https://primer3plus.com/primer3web/primer3web_input.htm).

- PCRs were performed using OneTaq® DNA Polymerase (NEB cat. no.M0480) following this protocol:

Component	50 ul reaction	Final concentration
5X One Taq Standard Reaction Buffer	10 ul	1 X
10mM dNTPs	1 ul	200uM
10uM Forward Primer	1 ul	0,2 uM
10 uM Reverse Primer	1 ul	0,2 uM
One Taq DNA Polymerase	0,25 ul	1,25 units/50 ul PCR
Template DNA	Variable	<1000ng
Nuclease-free water	to 50 ul	

Table 10: OneTaq DNA Polymerase reaction.

-Cycling program used (Table 11). Touchdown PCR was used to increase PCR specificity, a cyclic program was used in which the annealing temperature is gradually

reduced (2° C / every second cycle). The initial annealing temperature, 64°C, should be a few degrees higher than the estimated Tm of the primers (62°C). The annealing temperature is then gradually reduced to reach the calculated annealing temperature of the primers or some lower degrees (60°C). The amplification is then continued using this annealing temperature:

STEP	TEMPERATURE	TIME
Initial Denaturation	94°C	30 seconds
30 Cycles	94°C 64°-62°C-60°C 68°C	15-30 seconds 15-60 seconds 1 minute per Kb
Final Extension	68°C	5 minutes
Hold	4-10°C	

Table 11: Cycling program OneTaq Polymerase

The correct size of the PCR product was verified using agarose gel electrophoresis (refer to the Results section for details).

3.6.4 Cloning of zebrafish *krox20 and myoD* PCR products into PGEM T-easy vector

PGEM T-easy cloning Kit (Promega cat.no A1360, A1380, A3600, A3610, map below) was used to perform a cloning reaction and obtained a plasmid suitable for RNA probe preparation.

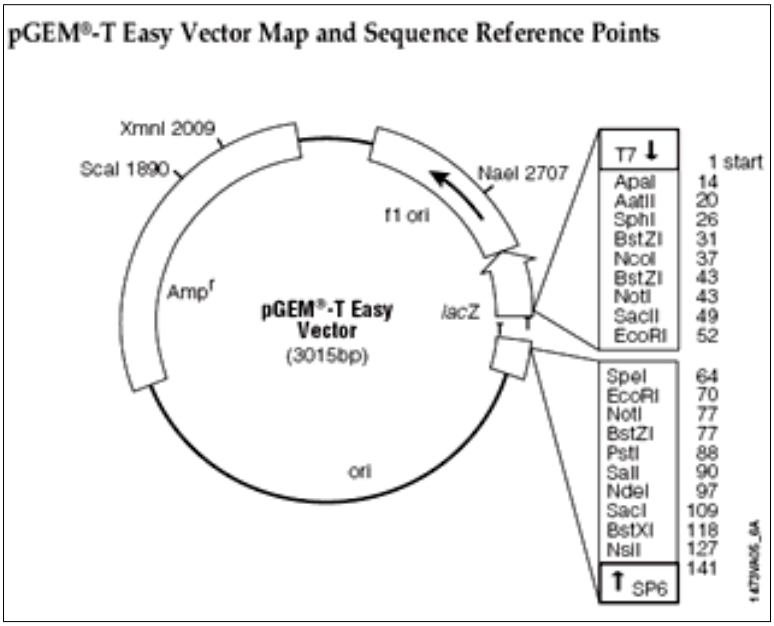


Figure 21: pGEM-T Easy Vector circle map and sequence reference points.

- Detailed ligation reaction:

The PCR products of *krox20* and *myoD* and the PGEM vector were combined as follows and incubated at 4°C overnight. To calculate the appropriate amount PCR product (insert) to include in the ligation reaction was used the following equation:
(ng of vector × kb size of insert ÷ Kb size of vector) × insert ÷ vector molar ratio = ng of the insert
PGEM-T easy vector is approximately 3Kb and is supplied at 50 ng/µl while *myoD* is 757bp and *krox20* is 797bp.*krox20* and *myoD* were added using a parameter for 3:1 insert: vector molar ratio, so for the ligation reaction were added 30 ng of *Krox-20* PCR product and 35 ng of *myoD* PCR product.

Reaction component	Standard reaction	Positive Control
2X Rapid Ligation Buffer, T4 DNA Ligase	5 ul	5 ul
pGEM-T Easy Vector (50ng)	1 ul	1 ul
PCR Product	Xul*	-
Control Insert DNA	-	2 ul
T4 DNA Ligase (3 Weiss units/ul)	1 ul	1 ul
Nuclease-free water to a final volume	10 ul	10 ul

Table 12: PGEM T-easy vector reaction. * For *krox20* were added 0,6ul (30ng); for *myoD* were added 0,4ul (35ng).

- Transformation reaction and clone selection

After ligation reaction, 2 µl of PGEM cloning product were introduced into 50 µl of JM109 High Competent Cells purchased from Promega. Transformation mix was placed on ice for 20', heat-shock was performed for 45'' in a water bath at exactly 42°C. Immediately tube was placed on ice for 2' and finally 950 µl room-temperature SOC medium were added to the tubes containing cells transformed with ligation reactions and were incubated for 1.5 hours at 37°C with shaking. After incubation 100 µl of transformation mix were plated on pre-warmed Luria Bertani (LB) agar plates containing 100ug/ml ampicillin and X-gal. [LB agar was prepared by adding 3,7gr of agar to 250 ml of LB medium. Finally, it was autoclaved, and plates were made using LB agar that was cooled to 42° C before adding ampicillin to a final concentration of 100 µg / ml. The medium was aliquoted in each plate, about 20 ml, then plates were incubated at room temperature to drying and then were put in 4°C. X-gal was added after, in according to PGEM's protocols: 20 µl of 50 mg / ml of X-Gal were spread on the surface of an ampicillin LB plate and allowed to absorb through 30 minutes at 37° C before use]. Plates were incubated at 37°C overnight. The empty PGEM vector contains the lack gene which encodes for B-galactoside. In the presence of X-gal (given during the transformation protocol), production of B-galactosidase forms blue color. DNA

ligated into plasmid (insert) disrupts the formation of functional B-galactosidase and white colonies are produced. Via white/blue colonies screening we could select colonies with our insert. To further confirm the positivity of the colonies (presence of insert cloned) we performed also a *PCR colony* using GoTaq® Green Master Mix (cat. No M7122) in 25 µl of final reaction in according to the manufacture's instruction. The correct size of the ligated fragment was determined by agarose gel electrophoresis. Agarose gel at 1% was made by adding 1 gr of agarose in 100ml of 10X TBE running buffer composed by**:** 890 mM Tris, 890 mM boric acid, 20 mM EDTA pH 8.0. Dilute to 1x for running and preparing agarose gels. After the screening, positive colonies were placed in the sterile tube containing LB media and were cultured at 37°C in the shaking incubator overnight at 200 rpm. Plasmid DNA was purified using Plasmid Plus Midi Kit (Qiagen cat.no12943 and 12945) in according to the manufacture's instruction. Following the plasmid purification, 3ug of this plasmidic DNA for each gene was linearized with the appropriate restriction enzyme that as a unique site located 3'(for sense probe) or 5'(for antisense probe) to the insert both genes were linearized with *SacII* enzyme (NEB cat. no. R0157L). The linearized DNA was purified using Monarch® PCR & DNA Cleanup Kit (5 µg) (NEB cat. no. T1030S) in according to the manufactures instructions and controlled by gel electrophoresis.

3.6.5. Synthesis of zebrafish *krox20* and *myoD* digoxygenin-labeled RNA probes

This linearized plasmid encoding zebrafish *krox20* and *myoD* were used as a template for the synthesis of the corresponding antisense RNA probes, labeled with digoxigenin-linked nucleotides. The synthesis of an antisense RNA probe was realized using DIG-RNA Labeling Kit (SP6/T7) (Roche cas.no.11175025910) as follows:

Component	**Amount per reaction**
Linear Template DNA	1 µg
5x Transcription buffer	2 µl
T3 or T7 RNA polymerase	2 µl
DIG-RNA labeling mix (UTP)	2 µl
RNAse inhibitor	1µl
DEPC water to a final volume *	20µl

Table 13: RNA Labeling reaction. *Diethyl pyrocarbonate (DEPC), was used to inactivate RNase enzymes in water by the covalent modification of histidine (most strongly), lysine, cysteine, and tyrosine residues (Chirgwin, John M,*et al.*1979; Wolf, Barry, *et al.* 1970). DEPC water was made by adding 1ml of DEPC in 1L of distilled water, it has been incubated at room temperature overnight, then it was autoclaved.

Reagents were mixed gently, and the RNA labeling reaction was incubate for 2h at 37°C. Then the remaining DNA template was removed by adding 10 µl of RDD buffer and 2,5 µl of DNase I stock solution, the reaction was incubated for 15' at room temperature. After the incubation, RNA labeling reaction was clean and concentrated using RNeasy MinElute Cleanup Kit (Qiagen cas.no.74204) in according to the manufacture's instruction.

Finally, the probe quality and integrity were verified by running 1ul on 1% agarose gel for about 15'. The concentration of the probe was quantified by measuring the amount of RNA using spectrophotometer (Thermo Scientific™ NanoDrop 2000 spectrophotometer).

3.6.6 Whole-mount in situ hybridization (protocol)

In situ hybridization (ISH) is a technique that allows to highlight the expression profile of a specific gene, by means of a chromogenic histochemical reaction in tissues fixed at a particular stage of development. In zebrafish it can be realized on whole embryos - whole mount - given the characteristic transparency, which allows the observation of marked cells even in the deepest tissues. The technique allows to detect the transcription product of a gene of interest.

PCR-amplified sequence of a gene of interest was used as a template for the synthesis of an antisense RNA probe, which was labeled with digoxigenin-linked nucleotides. Embryos were fixed and permeabilized before being soaked in the digoxigenin-labeled probe. The whole procedure takes normally 3 days and was carried out as follows according to established protocols Babykumari P, *et al*., 2013, Christine Thisse and Bernard Thisse, 2008, 2014 and the protocol of Dr. Antonella Lauri and Giulia Fasano (OPBG).

DAY 1 PRETREATMENT AND HYBRIDIZATION

1. Dechorionation and Rehydration

15hpf embryos *arf3* mutants, wt and not-injected controls, previously fixed in PFA 4% and stored in MetOH (as described in paragraph 3.7.1), were first **Dechorionation.** Embryos were then rehydrated stepwise with a different dilution of methanol / PBS (75%methanol/PBS, 50%methanol/PBS and 25%methanol/PBS) and finally put back in 100%PB-Tween (PBT).

2. Permeabilization reaction

Embryos were then incubated in proteinase K (stock 10µg/ml in PBT 1X). Time of the treatment with proteinase K depends on the developmental stage:

Stage	Time of treatment
15hpf	2’
24hpf	10’
48hpf	25’
72hpf	40’
96hpf	50’

3. Post-fixation and PBT washes

Proteinase K reactions was stopped by rinsing embryos in PBT followed by post fixation in 4% PFA in 1X PBS for 20 min at room temperature and by rinsing four times in PBT.

4. Pre-hybridization

Embryos were pre-hybridized for 3 h at 65°C (annealing probe’s temperature) in Hybridization Mixture (HM) pre-warmed at the same temperature.

Hybridization Mix for 100ml:

50% Formamide 50 ml

1.3x SSC 6.5 ml (20XSSC)

100 g/ml heparin (50 mg/ml) 200ul

50µg/ml Yeast RNA 5mg

0.2% Tween-20 (100%) 200ul
0.5% CHAPS (10%) 5ml
5mM EDTA pH8 (0,5M) 1ml
H20 DEPC to 100ml

5. Hybridization

Pre-hybridized embryos were then incubated overnight in hybridization mix (HM), containing 200 ng of the *krox20* and *myoD* probes diluted in HM, at 65°C (probe was denatured for 10 min at 95 °C).

DAY2 POST-HYBRIDIZATION WASHES

1. SSC (Saline-sodium Citrate) washes

Next day *krox20* and *myoD* probes were removed by 2 washes for 30' each in 100% HM pre-warmed at 65°C. Then these washes were performed in sequence:

- 1 for 5' in 75% HM and 25% SSC2X
- 1 for 5' in 50%HM and 50% SSC2X
- 1 for 5' in 25%HM and 75%SSC2X
- 1 for 10' in SSC2X
- 1 for 30' with SSC 0,2X pre-warmed at 65°C
- 1 for 30' with SSC0,2X at room temperature
- Then a final wash was performed in 10 mM PIPES and 0.5 M NaCl for 10' and, finally, in Maleic buffer tween-20 (MBT).

2. Pre-incubation

These washed zebrafish embryos were incubated in 2% Roche Blocking Reagent (Roche Applied Science, cat.no11096 /176001) for 2 hours at room temperature.

3. Incubation with anti-DIG Antibody

Embryos were incubated with a solution of anti-DIG antibody (Art.no. 11093274910) at a 5000-fold dilution in fresh MBT plus blocking reagents over night at 4°C with gentle agitation.

DAY3 STAINING

Next day, after washing away excess probe, hybrids was detected by immunohistochemistry using an alkaline phosphatase-conjugated antibody against digoxigenin and a chromogenic substrate.

1. Post-antibody washes and Pre-staining

Several washes in MBT were performed, embryos were incubated in a staining buffer for 30' with a gentle agitation.

2. Staining

Embryos were incubated in BM Purple chromogenic substrate (Roche Applied Science, co.no11442 074001) at room temperature in the dark until staining corresponding to the digoxigenin-labeled probe and therefore to the localization of zebrafish *krox20* and *myoD* was sufficiently developed (visible).

3. Stop reaction and Post-fixation

The embryos were observed under a stereomicroscope (Leica S8APO). After the desired staining intensity was reached, the reaction was stopped by several washes in PBT, one wash in PBS 1X/EDTA 10Mm for 10'. After stopping the reaction, embryos were post fixed in 4% PFA in 1X PBS for 20' and, finally, stored in 90% glycerol/PBT at 4° C.

4. Krox20 and myoD expression domain analysis in transient arf3 mutant embryos compared to wild type.

Whole mount zebrafish preparations mounted at the center of imaging slides with 6 layers of tape on the side in 80% glycerol (in PBS). Samples were dorsally mounted and subsequently imaged using microscope OLYMPUS TH4-200 microscope with 10x

objective. All the images and measures were conducted using Olympus CellSens Standard program. For each embryo (n = 10 for each experimental group) four parameters were taken into consideration, the length of the anteroposterior (AP) axis, the width of the mediolateral (ML) axis, the mid-lateral extension of the somites (measure calculated by performing a media of the extension of an anterior somite, a middle and a posterior one) and the number of the somites visible from the staining.

3.7 Statistical analysis

Statistical comparison of each *ARF3* mutant embryos group compared to wild type was conducted using GraphPad 8.0 version. Anderson-Darling test, D'agostino & Pearson test, Shapiro-Wilk test and Kolmogorov-Smirnov test were employed to assess the normality of the data. If data were normally distributed (based on the results from at least two out of three of the above-mentioned tests), they were analyzed using One-way ANOVA test with Sidak's *post hoc* test. Otherwise, the nonparametric Kruskall-Wallis test with Dunn's *post-hoc* test signed rank test was employed, after removing outliers' values. For *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$ (Ghasemi and Zahediasl., 2012). To compare the survival distributions of each experimental group Log-rank (Mantel-Cox) test was used (Bland and Altman., 2004). The graphs were generated with GraphPad 8.

4. RESULTS

4.1 Whole exome sequencing (WES) projects and results

Within the undiagnosed patient project so far, in 13 pediatric disorders pathogenic variants (11 of which were *de novo* mutations) were identified in novel genes not previously known to underlie a Mendelian phenotype, explaining a new syndrome, or found to be implicated in a known but unexplained phenotype.

In 27 cases, the causative variants were instead identified in genes known to be associated with reduced penetrance and variable expressivity, whose alteration cause disorders attributable to novel phenotypes, an expansion of the clinical spectrum of known phenotypes or an atypical presentation of the disorder. In this context, I focused my work on the investigation of new disease-causing genes or genetic variants associated with a group of the rare neurodevelopmental disorders. During my PhD work 3 new genes responsible for a new form of pediatric encephalopathies were identified via NGS. The results of this study to which I majorly contributed in the first part of my PhD project are shown in the table14.

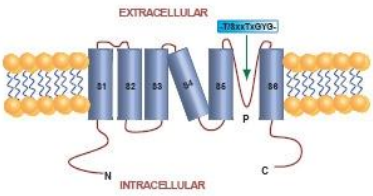
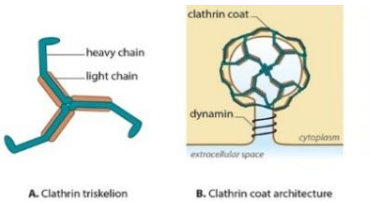
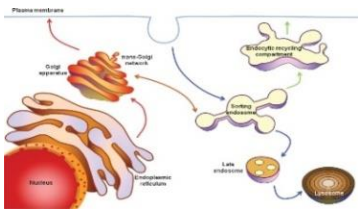
Gene	Protein	Clinical features	*De novo* variant
1) Encephalopathy caused by mutations in genes encoding channels or subunits of ion channels.			
KCNK4  	KCNK4:Potassium channel, subfamily K member 4	"FHEIG" syndrome: Facial dysmorphism, hypertrichosis, epilepsy, intellectual disability, developmental delay, gingival overgrowth.	Patient1: nucleotide substitution in heterozygosity **p.Ala172Glu** Patient2: nucleotide substitution in heterozygosity **p.Ala244Pro**
2) Encephalopathy caused by mutations in genes encoding proteins involved in the control of cytoskeletal dynamics and intracellular trafficking.			
CLTC  	CLTC: Clathrin heavy chain polypeptide gene	Epileptic spasms, myoclonus, intellectual disability, delay in global development, tonic-clonic crisis, refractory epilepsy.	Patient 1: nucleotide substitution in heterozygosity **p.Pro890Leu**
ARF3 	ARF3: ADP-ribosylationfactor 3	Microcephaly, cerebral malformation (cerebral cortical atrophy), developmental delay, severe intellectual disability, short stature, skeletal abnormalities (scoliosis) and epilepsy.	Patient 1: nucleotide substitution in heterozygosity **p.Lys127Glu** Patient 2: nucleotide substitution in compound heterozygosity in *cis* position **p.Leu12Val /p.Asp67Val**

Table 14: Summary of the WES projects and relative results carried out during the first phase of my PhD program.

4.2 Clathrin Heavy Chain *(CLTC)*

4.2.1 Case study

The proband is a 30-years-old woman. She was the first child of non-consanguineous Italian parents; she was born after a normal pregnancy and caesarean section delivery due to breech presentation. No dysmorphic features were detected throughout the follow-up. During early childhood, psychomotor delay and impairment of social skills were observed. At the age of 4 years, motor incoordination and proximal limb rigidity became evident. In the following years, she experienced cognitive development stagnation, and at the age of 11 years, mild intellectual disability, drooling, and slight gait ataxia were reported on clinical examination. An extensive metabolic work-up at the age of 13 detected a mild increase in blood phenylalanine, low urinary neopterin and low levels of homovanillic acid, 5-hydroxyindolacetic acid, and 5-methylthetrahydrofolic acid, with normal levels of biopterin and high phenylalaninein cerebrospinal fluid (CSF). Molecular examination of genes involved in tetrahydrobiopterin (BH4) and related biogenic amine pathways (*GCH1*, *FMRT*, *PTS*, *SPR*, *DHPR*, *TH*, *DDC*) detected no functionally relevant variants. Based on the hyperphenylalaninemia (HPA) in blood and CSF associated with low levels of monoamine neurotransmitter metabolites, a diagnosis of a mild form of phenylketonuria was suspected. Molecular data revealed compound heterozygosity in the phenylalanine hydroxylase (*PAH*) gene: a mild HPA-associated mutation (p.Thr380Met) and p.Asp151Glu, a novel variant affecting this gene. Lacking any alternative diagnostic clues, the condition was regarded as an atypical presentation of a very mild PAH deficit (L. Bonafè, *et al.*, 2001). In the following years, the girl experienced further cognitive decline and a relapsing-remitting disorder characterized by hypo- and bradykinesia, dysphagia, hyporexia with weight loss, and mood disorder. CSF examination at the age of 23 years confirmed the pattern of neurotransmitter alterations previously detected, showing low neopterin and biopterin with normal 5-methylthetrahydrofolic acid levels. L-dopa/carbidopa, 5-OH-tryptophan, pyridoxine, and folinic acid therapy failed to improve her neurological condition, which remained characterized by frequent fluctuations between periods of extreme bradypsychism, bradykinesia, anorexia with weight loss, and depression, and periods of relative wellbeing with hypomanic behaviour. At the age of 30, brain MRI did not detect any alteration. Neuropsychological assessment showed a moderate intellectual

disability, impaired sustained attention and semantic verbal fluency with relatively good adaptive skills.

The observation that her apparently unaffected 24 years-old younger brother carried the same *PAH* molecular defects and biochemical status prompted to a trio-based exome sequencing analysis that identified a *de novo* missense variant, c.2669C>T (NM_004859.3, **p.Pro890Leu**) in *CLTC* (F.F. Hamdan, *et al*., 2017; Filippo Manti, *et al*., 2018). The clinical diagnosis was defined at the Department of Child Neuro Psychiatry, Sapienza University of Rome, by Prof. Vincenzo Leuzzi, who provided us with all the clinical data, as well as the biological samples (peripheral blood and DNA) for carrying out the studies.

4.2.2 WES data

WES analysis was performed on the trio composed by the proband and her unaffected parents. DNA was obtained from peripheral leukocytes. Exomes were enriched using Nextera rapid capture v1.2 (Illumina), (library was performed using protocol described in Materials and Methods, chapter 3, par. 3.2, with same modifications according to manufacturer's instructions provided by Illumina protocols), and sequencing was performed on a NextSeq 500 platform (Illumina) in *Molecular Genetics and Functional Genomics* division (OPBG), resulting in 87x average depth on target (88,8% of which at least 20x depth). WES data were processed and analyzed by our bioinformatics, using an *in-house* pipeline set up (Kortüm F, *et al*., 2015; Niceta M, *et al*., 2015; Flex E, *et al*., 2016). Data obtained from WES analysis follow: the output VCF of our pipeline encompassed 116,719 high-quality variants that have been filtered using available public databases (*e.g.*, dbSNP147 and ExACV.0.3) to take into consideration only the new variants, those that are clinically associated and those with a predicted functional impact (*i.e.*, missense, nonsense, frameshift, indel variants, and putative splice site changes). The analysis retained variants with a low frequency (Minor Allelic Frequency MAF <0.1%) and those that present a frequency <1% in our *in-house* database, which includes the frequency data of about 1300 exomes. The variants were thus reduced to 14,618. Among these, 279 were private, rare and potentially interesting to the phenotype. These variants were prioritized using informatics tools such as CADD (Combined Annotation Dependent Depletion) and MetaSVM, and to each were

assigned scores indicative of their potential functional and structural effect on the gene product. Following variants prioritization, I took into consideration candidate variants with the highest score. I selected a certain number of variants based on the expression pattern of the mutated gene, the function of the encoded protein, the conservation of the amino acid residue affected by the mutation, the type of substitution predicted based on the nucleotide change, and the origin of the mutation (germinal *versus* somatic; *de novo* versus familial). A higher priority was thus assigned to mutations affecting genes whose expression was ubiquitous, with the encoded protein being implicated in a functionally relevant signal transduction pathway, and the variant affecting evolutionary conserved residue, whose replacement was predicted to be harmful. This strategy allowed us to identify a *de novo* variant, a heterozygous nucleotide substitution in the *CLTC* gene, encoding the clathrin heavy chain 1(c.2669C>T (NM_004859.3, **p.Pro890Leu**), also supported by the *in silico* predictors (CADD=34, metaSVM=-0.72). No other candidate variants were identified. The *de novo* origin of the variant was confirmed by Sanger sequencing.

4.2.3 Genotype-phenotype correlation

The Clathrin heavy chain polypeptide gene (*CLTC*) is highly expressed in the brain, as well as in various human tissues linked to growth and development. CLTC is involved in diverse cellular process including endocytosis, intracellular trafficking, and synaptic recycling (Kuenen S, *et al* 2008; Robinson MS. 2015). This gene has recently been associated with a variable clinical phenotype affecting neurodevelopment, with a variable neurological involvement ranging from learning disability to severe intellectual disability with or without epileptic encephalopathies. Genotype-phenotype correlations were reported. While mutations affecting the *N*-terminal portion of the light chain binding domain are generally associated with refractory epilepsy, truncating variants at the *C*-terminus of the protein are more prevalent in patients with hypotonia, developmental delay and/or isolated intellectual disability (Hamdan, *et al*., 2017). Functionally, a defective CLTC protein may result in the depletion of biogenic amines in the brain by altering their synaptic turnover. The maintenance of synaptic transmission, indeed, requires vesicles recycling after neurotransmitters release, and the majority of synaptic vesicles appears to be recycled through pathways requiring

clathrin, which is involved in the first step of endocytic vesicle formation at the level of the synaptic terminal. Consistently, *in vitro* studies have provided evidence that a deficiency in clathrin function causes a progressive decline in synaptic transmission and results in massive bulk membrane retrieval (Kasprowicz, *et al*., 2008).In line with that, *in vitro* and *in vivo* studies were performed in our laboratory to characterized the functional impact of the recurrent p.Pro890Leu change. Confocal microscopy analyses, performed by Dr. Valentina Muto, indicated a more scattered distribution of clathrin in patient's fibroblasts compared to control cells. Of note, a proliferative slowdown was observed in patient's cells, with flow cytometry analysis documenting a block in cell cycle transition from G0/G1 to S phase, suggestive of unreliable intake of growth factors. In line with that, cells from the affected subject displayed decreased transferrin uptake, indicating defective clathrin-mediated endocytosis. Finally, live-cell imaging spinning-disc microscopy analysis revealed that transferrin-containing vesicles are slower or completely still in patient's cells compared to controls. To explore the effect of the mutation *in vivo*, were generated knock-in *C. elegans* strains by CRISPR-Cas9, performed by Dr. Luca Pannone and Dr. Simone Martinelli (Istituto Superiore di Sanità). Based on the evidence that clathrin-mediated endocytosis plays a key role in synaptic vesicle recycling, animals were treated with aldicarb, an acetylcholinesterase inhibitor, or levamisole, a cholinergic receptor agonist, to assess synaptic transmission at the neuromuscular junction. Of note, mutant worms displayed resistance to the former drug and hypersensitivity to the latter, indicating a disruptive effect of the mutation on the release/recycling of both acetylcholine and GABA. Dopamine transmission was also affected, as revealed by the inability of mutant animals to slowdown in the presence of food, a dopamine-mediated behavior. Finally, nematodes carrying the mutation had defective synaptic plasticity, as revealed by learning assays performed following conditioning with high doses of benzaldehyde. These findings establish that the p.Pro890Leu *CLTC* mutation causes defective intracellular trafficking, which likely results in neurotransmitter depletion in the brain due to aberrant synaptic vesicles formation/turnover.

4.3 Potassium Channel, Subfamily K, Member 4 *(KCNK4)*

4.3.1 Case study

Two male children from non-consanguineous Italian parents with unremarkable family history shared a superimposable clinical phenotype including developmental delay/intellectual disability (ID), distinctive *facies*, gingival overgrowth, hypertrichosis, and EEG anomalies/epilepsy. WES analysis allowed to identify the underlying molecular cause. Clinical data and DNA samples were collected from the participating families after written informed consent was obtained. **Subject 1** is an 11-month-old boy with hypotonic face, bi temporal narrowing, thick and straight eyebrows, long eyelashes, short and deep filter, prominent upper and lower vermilion, and receding chin. **Subject 2** is a 5-year-old boy. He showed bi temporal narrowing, thick and straight eyebrows, long eyelashes, low-profile previews, short deep filtering and prominent upper and lower vermilion. The clinical diagnosis of subjects was checked at OPBG, in the *Molecular Genetics and Functional Genomics* division, by Dr. Francesca Clementina Radio.

4.3.2 WES data

The exome capture kits used were the SureSelect CREv2 and Illumina Nextera, achieving 85% and 89% 20x coverage for OPBG patient 1 and 2, respectively. WES data were processed and analyzed by our bioinformatics, using an in-house pipeline set up (Kortüm F, *et al*., 2015; Niceta M, *et al*., 2015; Flex E, *et al*., 2016). Reads were aligned to human genome build GRCh37/UCSC hg19, and variants were quality filtered according to GATK's 2016 best practices, annotated, and filtered against public (dbSNP150 and gnomAD v.2.0) and *in-house* (about 1,300 population-matched exomes) databases to retain private and rare (unknown frequency or MAF < 0.1%) variants located in exons with any effect on the coding sequence, and within splice site regions. The output VCF of our pipeline encompassed 13,484 variants (subject 1) and 14,574 variants (subject 2) with a predicted functional effect. Among these, 352 and 390 novel, clinically associated, and unknown/low frequency variants. The functional impact of variants was analyzed by Combined Annotation Dependent Depletion

(CADD) v.1.3, and M-CAP v.1.0, to obtain clinical interpretation. By applying our prioritization method, *KCNK4* (**subject 1**:c.515C>A, **p.Ala172Glu**; **subject 2**: c.730G>C, **p.Ala244Pro**) resulted as the only putative shared candidate gene compatible with autosomal-dominant transmission of the trait. The two variants were confirmed to be *de novo* by Sanger sequencing. They had not been reported in ExAC/gnomAD so far, affected residues conserved among vertebrate orthologs, with Ala_{244}invariantly occurring among members of the mechano-gated TRAAK/TREK subfamily, and were predicted to have a damaging impact on protein function (**subject 1**: CADD = 33, M-CAP = 0.028; **subject 2**: CADD = 29.5, M-CAP = 0.038) (Kircher *et al*. 2014; Jagadeesh *et al.* 2016).

4.3.3 Genotype-phenotype correlation

The human genome contains almost 80 genes encoding potassium (K^+) channels, which constitute the most diversified class of ion channels with regard to structure and gating characteristics (see OMIM database). Such diversity allows K^+ channels to serve various functions in both excitable and non-excitable cells (Hille,B. 2001; Tian, C.*et al*, 2014; Urrego, D.,*et al*, 2014). K^+ channels contribute to the maintenance and stabilization of the resting membrane potential, help to repolarize action potentials, mediate hyperpolarization, and determine cellular electrical activity or even cell proliferation. Moreover, K^+ channels also efficiently mediate modulations of membrane potential and cell function by controlling the K^+ flux through the cell membrane in response to multiple signals. Since the membrane potential has a key role in the control of a wide array of cellular processes, including cellular excitability, neurotransmitter release, hormone secretion, and electrolyte transport, it is not surprising that aberrant K^+ flux due to hyperactivating/inactivating mutations in genes coding for K^+ channels, may underlay a heterogeneous group of human disorders affecting neurotransmission and central nervous system function, cardiac electrophysiology, hormone secretion, and kidney function (Shieh, C.C. *et al*, 2000; Jentsch, T.J., 2000; Chiamvimonvat, N. *et al*, 2017). The aberrant function of some of these channels has recently been documented to affect development, underlying syndromic traits. Here, we report on the identification of dominantly acting mutations in *KCNK4* (MIM: 605720), which encodes a two-pore-domain (K2P) K^+ channel (Fink,M., *et al*, 1998; Feliciangeli, S. *et al*, 2015; Sepulveda,

F.V.,*et al*, 2015), as of the molecular cause underlying a new neurodevelopment disorederthat was called with the acronym "FHEIG" (facial dysmorphism, hypertrichosis, epilepsy, intellectual disability/developmental delay, gingival overgrowth) (Bauer et al. 2018). KCNK4 (also known as TRAAK, TWIK-related arachidonic acid-stimulated K^+ channel) belongs to the mechano-gated ion channels of the TRAAK/TREK subfamily of two-pore-domain (K2P) potassium channels (Fink, M. et al. 1998; Brohawn and MacKinnon, 2012). Based on available structures of KCNK4, both affected residues mapped in regions controlling KCNK4 activation (Lolicato *et al*., 2017), suggesting a disruptive effect of the mutations on channel gating (Niemeyer et al., 2016). Molecular dynamics simulations showed that the lateral fenestrations of the channel were predominantly closed for a significant fraction of the trajectory in the **Ala244Pro** and **Ala172Glu** mutants, while they were open for a large fraction in the wild-type protein; thus, mutations are predicted to cause sealing of the lateral intra-membrane fenestration proposed to negatively control K^+ flow by allowing lipid access to the central cavity of the channel. KCNK4 channel intervenes in the rapid depolarization response of neuronal cells in response to lipid (*e.g.*, arachidonic acid) stimuli and mechanical (*e.g.*, cell stretch) cues. *In vitro* electrophysiology assays have shown that *KCNK4*mutations lead to the constitutive opening of the ion channel through different molecular mechanisms that interfere with the normal adjustment of channel dynamics. The impressive gain-of-function caused by both mutations results in a dramatically increased basal K^+ conductance, which is expected to drain K^+ out of neurons increasing interstitial K^+ concentration, with a depolarizing effect also on neighbouring cells (Fink, M. *et al*. 1998).

4.4 ADP Ribosylation Factor 3 *(ARF3)*

4.4.1 Case study

During the work that concern with my PhD two independent individuals were analyzed. **Subject 1** was a 3 years-old girl born from non-consanguineous Italian parents. She presents a clinical picture characterized byhyposomia, microcephaly, cerebral malformation (pontocerebellar hypoplasia, expanded ventricular system, cyst of the pellucid septum, thin corpus callosum), congenital heart disease, severe psychomotor

delay, generalized hypertonus, epilepsy, cost-vertebral anomalies (11 pairs of left ribs; dorso-lumbar scoliosis) and inguinal hernia. The clinical diagnosis of subject 1 was checked at the Department of "*Molecular Genetics and Functional Genomics"*, Bambino Gesu Children's Hospital, by Dr. Francesca Clementina Radio. Given the absence of related cases in our cohort and the presence of a few information in the literature regarding this gene and its possible involvement in the observed phenotype, we shared genetic and clinical information on GeneMatcher (http://www.genematcher.org). This is a free website developed to allow the exchange of information between physicians and researchers from all over the world, to associate genes identified via NGS with Mendelian phenotypes, and allow the understanding of the molecular basis of rare and ultra-rare diseases. Thanks to the use of this informatics tool, we found another individual with a similar phenotype a mutation affecting the same gene. **Subject 2** was 18 months -old child born from non-consanguineous French parents, microcephaly, cerebral malformation (cerebral cortical atrophy), developmental delay, severe intellectual disability, short stature, skeletal abnormalities (scoliosis) and epilepsy.

4.4.2 WES data

DNA of the affected **subject 1** and his parents was extracted from circulating leukocytes for trio-based WES analysis. Exome capture was carried out using SureSelect Clinical Research Exome (Agilent) and raw data were processed and analyzed by our bioinformatics using an in-house implemented pipeline previously described up (Kortüm F, *et al*., 2015; Niceta M, *et al*., 2015; Flex E, *et al*., 2016), which is based on the GATK Best Practices (Van der Auwera GA et al, 2013). The UCSC GRCh37/hg19 version of genome assembly was used as a reference for reads alignment by means of BWA-MEM (Li H, 2013) tool and the subsequent variant calling. We used SnpEff v.4.3 (Cingolani P *et al*, 2012) and dbNSFP v.3.5 (Liu X, Jian X, and Boerwinkle E, 2013), tools for variants functional annotation, including Combined Annotation Dependent Depletion (CADD) v.1.4, Mendelian Clinically Applicable Pathogenicity (M-CAP) v.1.0 and Intervar v.2.0.1 for functional impact prediction (Kircher M, *et al* 2014; Jagadeesh K, *et al*, 2016; Li Q and Wang K, 2017). Thereby, the analysis was narrowed to variants which either affect coding sequences or splice site

regions. Moreover, high-quality variants were filtered against public databases (dbSNP150 and gnomAD V.2.0.1) so that only variants with unknown frequency or having MAF <0.1%, as well as variants occurring with frequency <1% in our population-matched database (~1300WES) were considered. The exome enrichment evaluation for this experiment showed 147x average coverage and at least 20x depth of coverage for 94% of the target. Among 55,639 high-quality variants, 13,469 affects either CDS or splice sites, whereas 282 have low or unknown frequency according to aforesaid threshold values. These variants were prioritized using informatics systems such as CADD and MetaSVM, and to each were assigned scores indicative of their potential functional and/or structural effect on the gene product. Following variants prioritization, I took into consideration candidate variants with the highest score. I selected a certain number of variants based on the expression pattern of the mutated gene, the function of the encoded protein, the conservation of the amino acid residue affected by the mutation, the type of substitution predicted based on the nucleotide change, and the origin of the mutation (germinal *versus* somatic; *de novo* versus familial). A higher priority was thus assigned to mutations affecting genes whose expression was ubiquitous, with the encoded protein being implicated in a functionally relevant signal transduction pathway, and the variant affecting evolutionary conserved residue, whose replacement was predicted to be harmful. This strategy allowed us to identify *de novo* variant in the *ARF3* gene (ADP Ribosylation Factor 3) (c.379A>G, **p.Lys127Glu**), also supported by the *in silico* predictors (CADD= 31; MetaSVM=0.66) .The variant was confirmed to be *de novo* by Sanger sequencing. The identification of a second subject carrying the compound heterozygous mutations in *cis* position, c.34C>G and c.200A>T (**p.Leu12Val** and **p.Asp67Val**), as *de novo* events in the same gene further supported the causal association between *ARF3* mutations and the trait.

4.4.3 Genotype-phenotype correlation

ADP-ribosylation factor (Arf) family proteins are master regulators of membrane trafficking, each of them controlling all the steps of vesicular transport. The intracellular trafficking pathways can be divided in two main routes: an outbound secretory route from the endoplasmic reticulum (ER) to the plasma membrane, through the Golgi apparatus; and the endocytic pathway, which transports cargo from the plasma

membrane to lysosomes, through early (or sorting) and late endosomes (Fig.19). Finally, cargo that does not follow the endocytic pathway to lysosomes can be recycled back to the plasma membrane through the perinuclear endocytic recycling compartment (ERC) or recycling endosome. Vesicular transport between donor and acceptor membrane-bound compartments can be divided into several steps that are regulated by different families of proteins: (1) budding of a vesicle containing sorted cargo from the donor compartment; (2) transport of this vesicle via the cytoskeleton; (3) tethering and docking to the acceptor compartment and (4) fusion, which releases the cargo into the acceptor compartment. Normally, cargos (containing important molecules such as neurotrasmitters or receptors) are transported in vesicular carriers between compartments in a regulated manner, enabling their differential protein and lipid composition. ARF proteins control the vesicles budding process (Donaldson JG, Jackson CL. 2011; Stenmark H.2009) and the tethering of vesicles to acceptor compartments and in the interaction with the cytoskeleton). These families of small guanine-nucleotide-binding (G) proteins are well suited to regulate membrane trafficking processes since their dynamic nucleotide state (GTP or GDP bound) determines their conformation and the capacity to bind to a multitude of effectors, which mediate their functions. Physiological ARF proteins function is therefore crucial inside cells, as a correct intracellular trafficking pathway ensure the communication between membrane-bound compartments of eukaryotic cells and in particular in the nervous system where cargos need to travel long distance from the soma to the axon and molecules needs to be released in a highly regulated fashion.

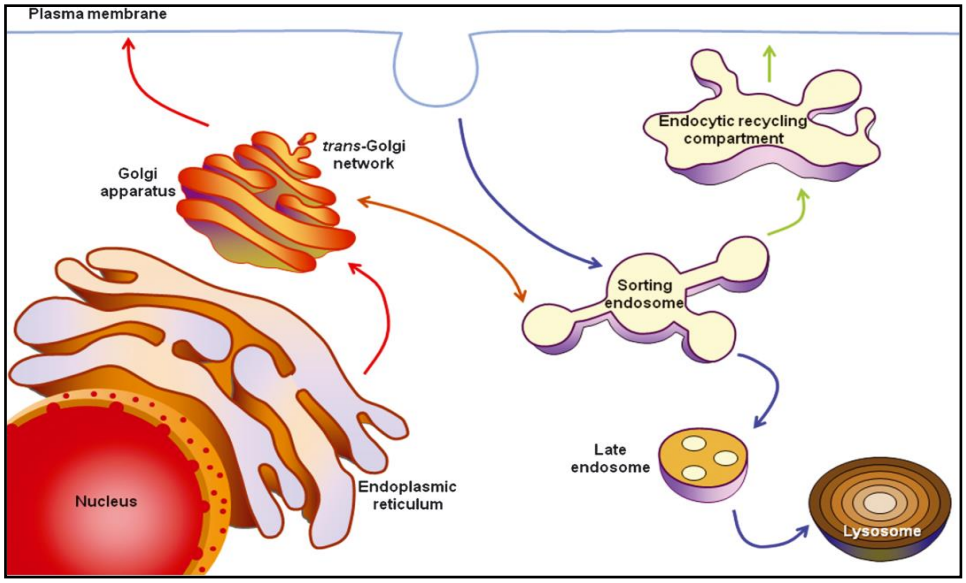


Figure 22: Intracellular trafficking pathways. Newly synthesized proteins exit the endoplasmic reticulum and are transported through the Golgi apparatus and trans-Golgi network (TGN) to the plasma membrane via the secretory pathway (red arrows). Cargo can be internalized and transported to lysosomes through early/sorting endosomes and late endosomes via the endocytic pathway (blue arrows). Cargo can also recycle through the endocytic recycling compartment (ERC) via the slow recycling pathway (green arrows). The endocytic and secretory pathways are connected by trafficking between the TGN and endosomes (brown arrow). (Seixas, *et al*. 2013).

It is therefore not surprising that in recent years, several inherited diseases affecting the nervous system have been associated with mutations in small G proteins of the Arf families. The genetic diseases caused by defects in Arfs proteins are heterogeneous and display diverse symptoms (Donaldson JG, Jackson CL. 2011; Seixas, *et al*. 2013). As described in the previous paragraph, in our lab a bioinformatic analysis has recently identified new variants in the *ARF3* gene as possible responsible for a new form of pediatric encephalopathy. In light of this, and considering the heterogeneity of the genetic diseases caused by mutations in ARF proteins in the patients, I next focused the second part of my PhD work on the functional validation and characterization of the variants identified in *ARF3*, using zebrafish as an *in vivo* model system. This was part of a bigger project involving both *in vitro* and *in vivo* studies with my colleagues, aimed to understand how an aberrant intracellular trafficking can perturb embryogenesis, particularly during neurodevelopment.

4.5 Zebrafish arf3 characterization

4.5.1 Protein sequence conservation

To analyze the conservation of amino acid residues affected by the human mutations, I performed a characterization of the *arf3* gene in zebrafish. Zebrafish *arf3* is the orthologous gene to human *ARF3* (ADP ribosylation factor 3). It is predicted to have GTP binding activity and to be involved in Golgi to plasma membrane transport and intracellular protein transport; and it predicted to localize to the Golgi apparatus and plasma membrane. ZFIN database (The Zebrafish Information Network) and Ensembl database (Ensembl genome browser 98) contains entries for two paralogs of *arf3* gene in zebrafish genome named: *arf3a* (ZFIN ID: ZDB-GENE-040801-176, ENSEMBL ID: ENSDARG00000070539.7, chromosome 23) and *arf3b* (ZFIN ID: ZDB-GENE-030616-356, ENSEMBL ID: ENSDARG00000036998, chromosome 6).

In order to determine the conserved domains of ARF3 protein in human and zebrafish, I aligned and compared the amino acid sequences between arf3a and arf3b zebrafish protein and ARF3 human protein respectively (Figure 23), using the software *Clustal Omega* (1.2.4) Multiple Sequence Alignment with default parameters. The analysis showed that the three mutated amino acid residues in human protein ARF3 (Lys127Glu**,** Leu12Val /p.Asp67Val), were preserved in the two zebrafish paralogs.

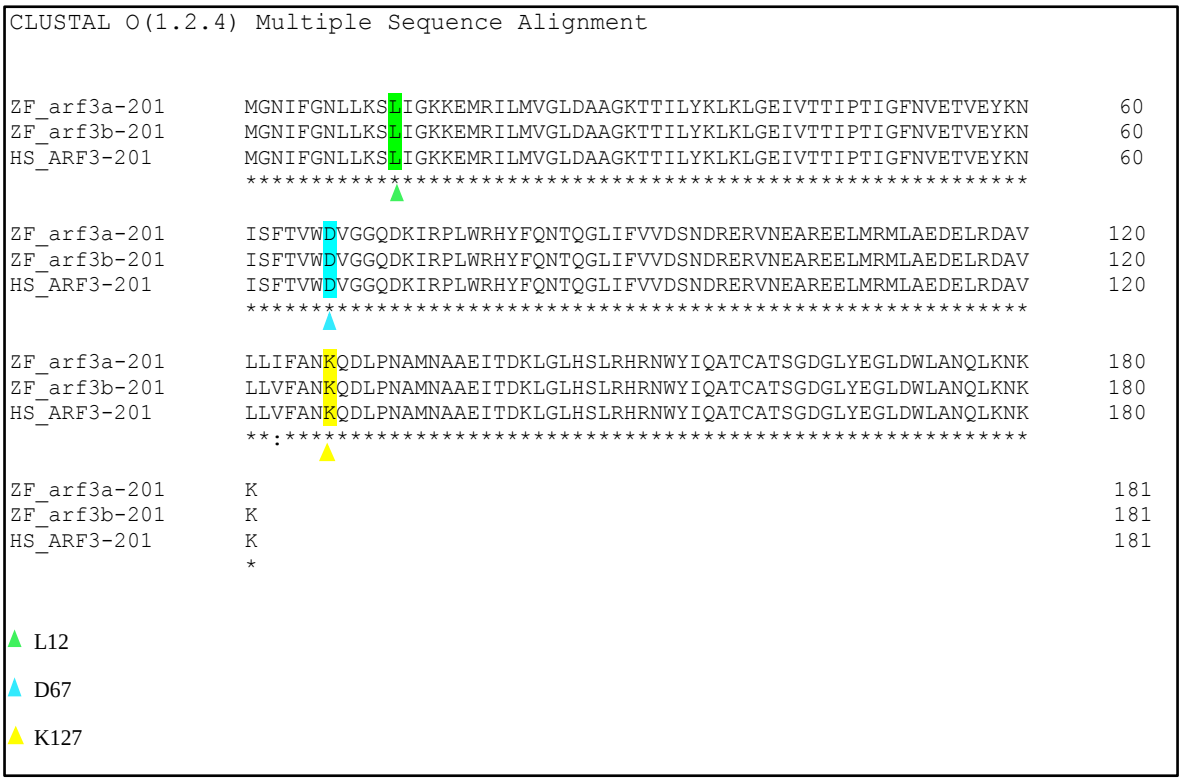


Figure 23: Multiple species alignment of *ARF3* orthologs of human and zebrafish. Colored amino acid residues, highlighted by triangles, indicate the amino acid residues affected by mutation and conserved between orthologs genes. Alignment performed using Clustal Omega (1.2.4) Multiple Sequence Alignment.

4.5.2 Phenotype characterization in the zebrafish ARF3 mutants

Mutants generation was conceived and carried out by Dr. Antonella Lauri and Dr. Giulia Fasano, using microinjection of mutated human ARF3 mRNA produced in vitro for $ARF3^{WT}$-myc, $ARF3^{K127E}$-myc and $ARF3^{L12V/D67V}$-myc). Mutagenesis on the CDS was performed by Dr. Valentina Muto 50ng/µl and 25ng/µl of the corresponding mRNA was injected in biological and technical duplicates in one cell stage embryos.

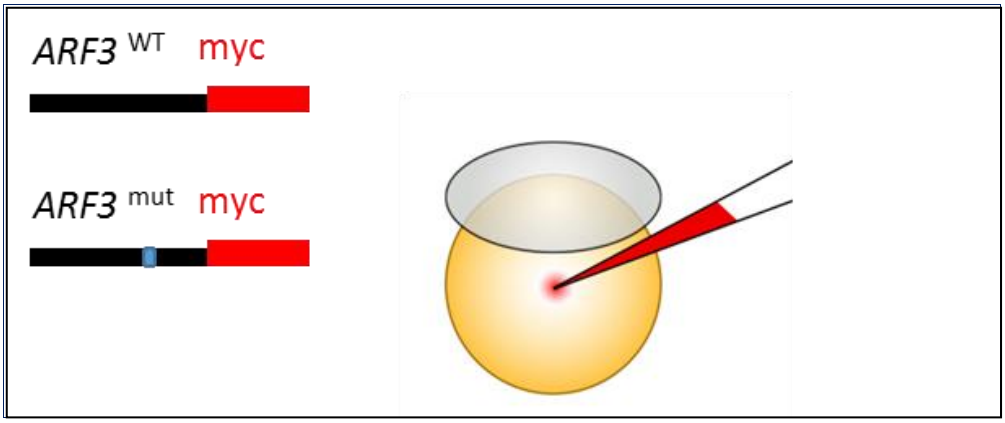


Figure 24. General scheme of microinjection into zebrafish embryos at 1-cell stage embryos. Schematic drawing of the general injection procedure. mRNA encoding for *ARF3*-myc^{WT} , *ARF3* ^{K127E} – myc and *ARF3*$^{L12V/D67V}$-myc was injected into one-cell stage zebrafish embryos.

I focused on the analysis of the impact of overexpression of mutated *ARF3* cds on zebrafish embryos physiology at the development stage of 24 hpf and 48 hpf using not-injected embryos as a control. The embryos were examined using multidimensional fluorescence stereomicroscope (Leica M205FA). The first parameter analyzed was the embryos survival rate that was checked at 24 hpf for each injection experiment. We plotted the data on a graph of survival *vs*. time (hpf). From our analysis, it was revealed that *ARF3*^{K127E}-myc embryos showed a significantly lower survival and a stronger effect compared with *ARF3*$^{L12V/D67V}$-myc embryos.

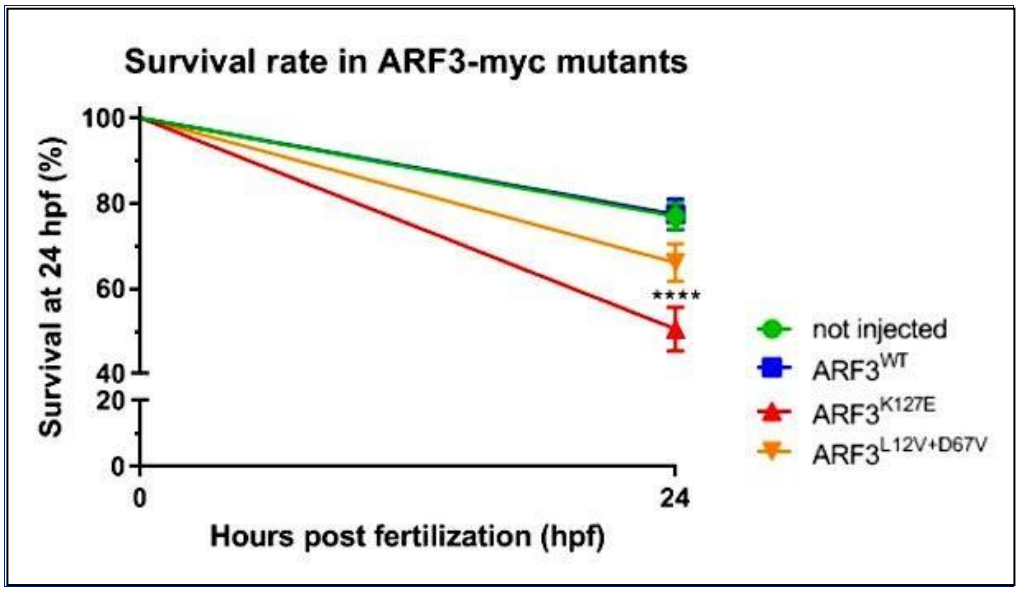
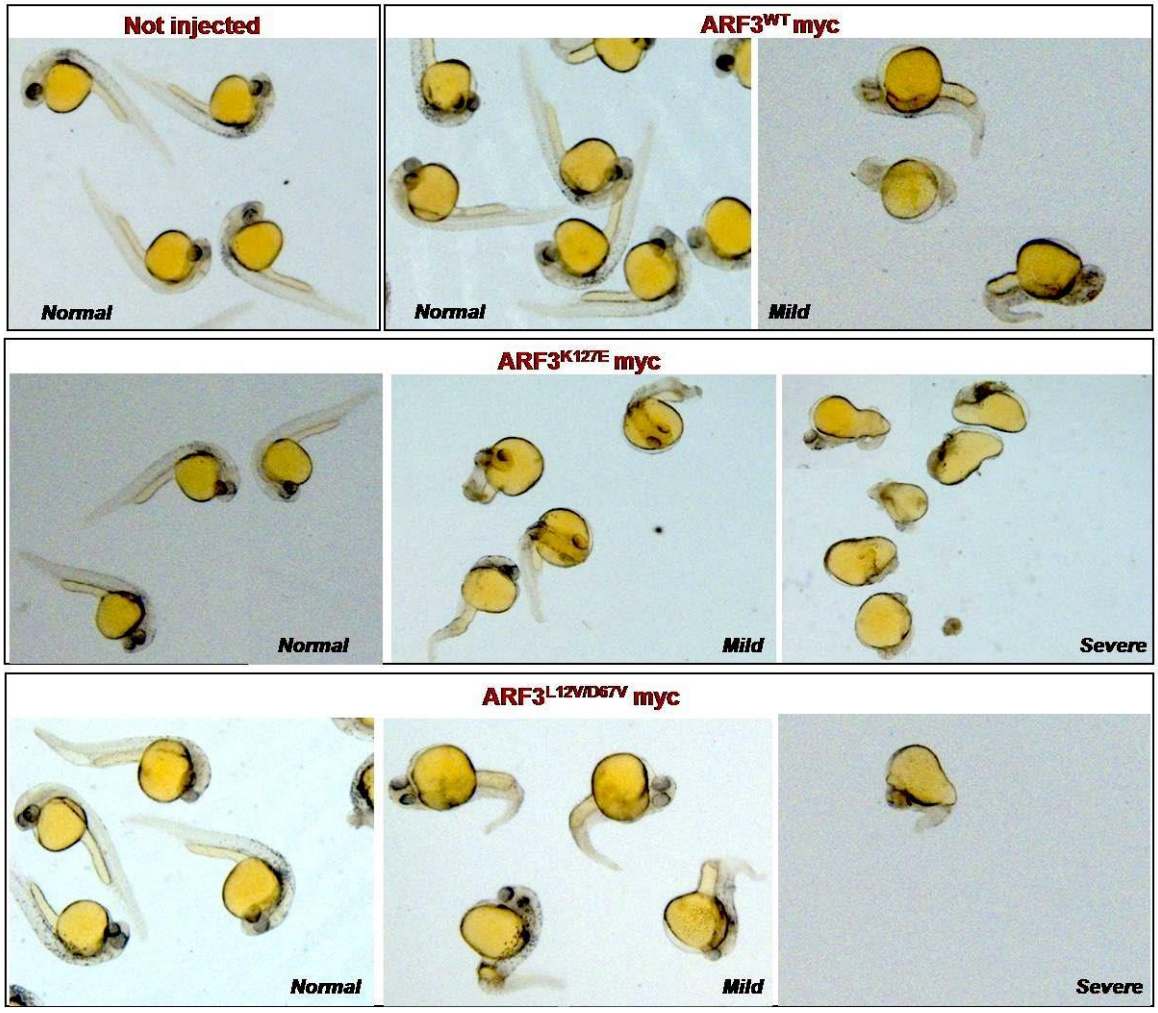


Figure 25: survival percentages in *ARF3*-myc mutants at 24 hpf. The survival rates for the not-injected and the control injected (*ARF3*^{WT}-myc) embryos showed no significant differences in survival rate, *ARF3*^{K127E}-myc embryos had a significantly lower survival rate at 24 hpf (P value<0,0001). In *ARF3*$^{L12V/D67V}$-myc embryos, instead, was observed a milder effect of survival rate compared with other mutant embryos. Analysis performed with GraphPad Prism 8 using Log-rank (Mantel-Cox) test. Here the 50ng/µl concentration of mRNA was tested.

Next, together with Dr. A. Lauri and Dr. G. Fasano, I observed and characterized gross phenotype of embryos in order to quantify alteration of the body axis formation and the effect of *ARF3* mutations on zebrafish morphology. Three different embryo's phenotypes were identified described as follows: *normal* = all the embryos with normal body length and axis, no visible body curvature, as in not-injected controls, *mild* = all the embryos that presented mild shortening of body axis and microcephaly, *severe* = all the embryos with defective body axis elongation, pronounced microcephaly and, in some rare cases, missing head (Figure 26 A). The penetrance analysis showed a statistically significant alteration of body axis formation observed in both *ARF3* mutants. In particular, *ARF3*^{K127E}-myc mutant showed a higher incidence of severe phenotype than *ARF3*$^{L12V/D67V}$-myc, this is visible already by 24 hpf in *ARF3*^{K127E}-myc mutants, worsening at 48 hpf. *ARF3*$^{L12V/D67V}$-myc mutants, instead, showed a clear phenotype starting only from 48h of development (Figure 26 B-E).

A


Not injected
Normal
ARF3WT myc
Normal
Mild
ARF3K127E myc
Normal
Mild
Severe
ARF3L12V/D67V myc
Normal
Mild
Severe

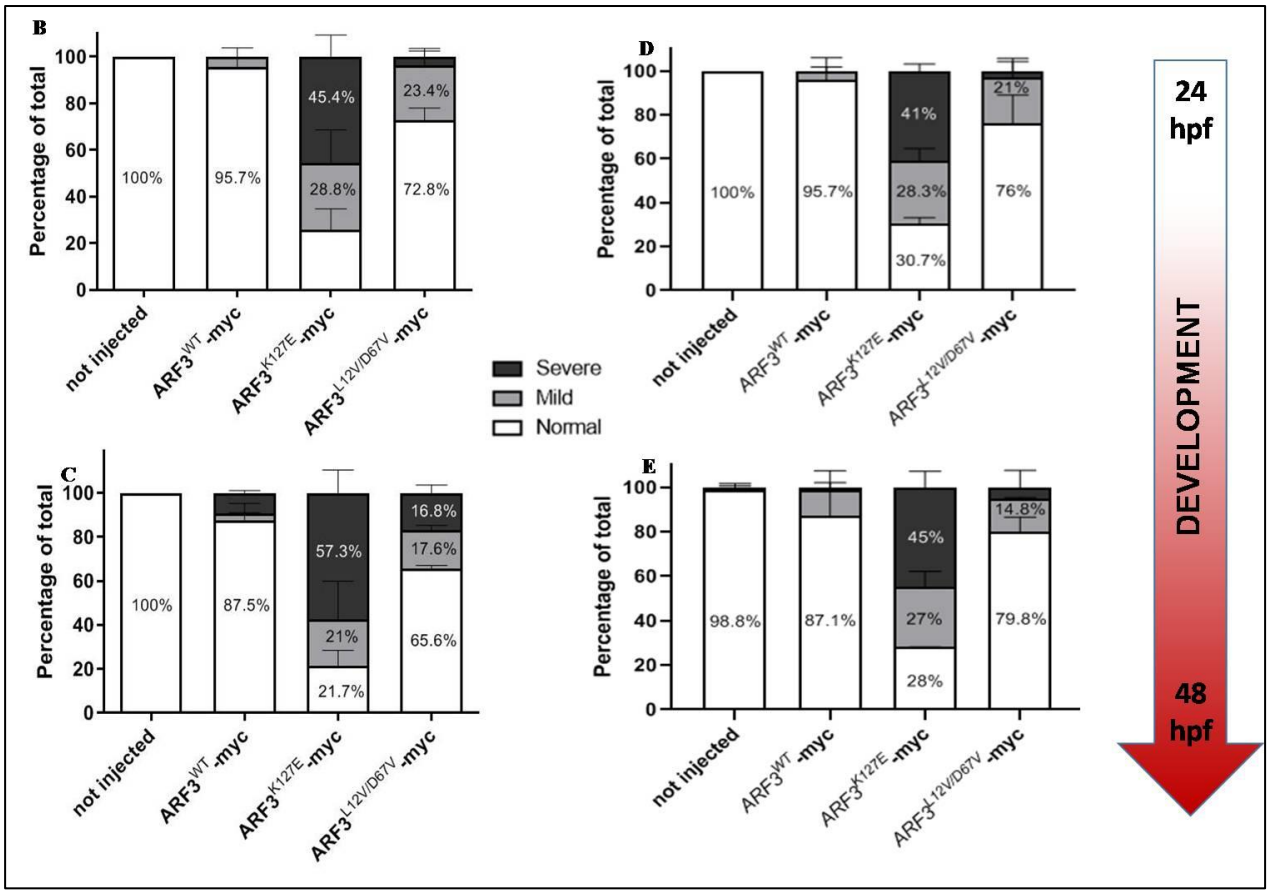


Figure 26: A. Images of representative *ARF3* mutant embryos compared with not-injected embryos and embryos injected with *ARF3*^{WT} as control. Mutant embryos showed three different phenotypes: normal, mild and severe as explained in the text. **B. C** Phenotype penetrance analysis in *ARF3* mutants (50ng) at 24 and 48 hpf respectively. Analysis are correlated to two experimental replicates. **D-E.** Phenotype penetrance analysis in *ARF3*-myc mutants (25ng) at 24 and 48 hpf respectively. Analysis are correlated to two experimental replicates. The data obtained show that *ARF3*^{K127E}-myc mutants present a higher incidence of severe phenotype than ARF3$^{L12V/D67V}$-myc, this is visible already by 24hpf in *ARF3*^{K127E}-myc, worsening at 48hpf. *ARF3*$^{L12V/D67V}$-myc mutants, instead, show a clear phenotype starting from 48h of development. Phenotype worsens visibility with development (24hpf- 48hpf) when a dose-dependent effect was observed. This is true for both concentrations of mRNA. All analysis and graphs are performed with GraphPad Prism 8.

4.5.3 In-depth body axis and brain phenotype analysis

Next, we asked what morphological characteristics could be investigated in detail in the mutants, to understand the deeper alterations involved in the gross phenotype observed. Therefore, I focused the analysis on: 1) the curvature of the body axis of the embryos and 2) the cephalic development, in particular on the possible presence of microcephaly in *ARF3* mutants, considering the clinical characteristics of the patients studied. For 1) in the first step, I analyzed the abnormalities of the body axis taking into consideration two different parameters: short stature and increase of body curvature. A significant

alteration of body axis formation in *ARF3* mutants was observed only in the mutants (Figure 27).

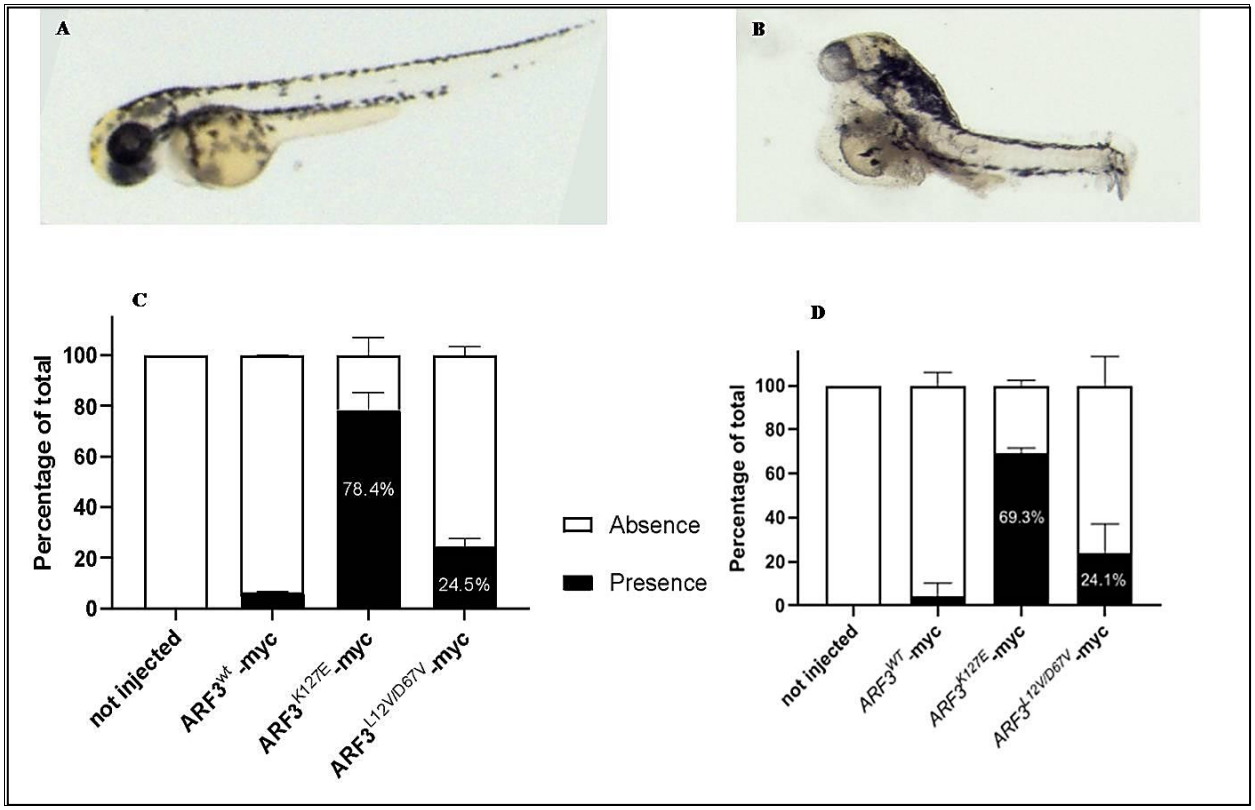


Figure 27: Body curvature in *ARF3*myc mutants. A. Image of representative ARF3 ^{WT} -myc embryo. **B**. Image of representative *ARF3*^{K127} -myc embryo. Severe defects of the body axis are visible already at 24 h of development in the mutants. **C-D**. Body curvature penetrance analysis in *ARF3*-myc mutants at two different concentrations of mRNA injected (respectively at 50ng and 25 ng).Significant alterations of the body axis formation are clear, in particular *ARF3*^{K127} -myc mutants show a more severe phenotype.

2) In order to broaden our understanding of the effect of the *ARF3* mutations, I also investigated the possible impact on cephalic development, analyzing the presence of microcephaly in the mutants. To answer this question, mutant embryos were subjected to an analysis of head size, where Dr. Giulia Fasano considered as a *cephalic area* the anatomical region from the beginning of the head to otic vesicle. This area was quantified using the measure tool calling Fiji plug-in. The analysis confirmed our hypothesis and microcephaly was observed in both mutants with a more severe evidence in *ARF3*^{K127E} -myc than *ARF3*$^{L12V/D67V}$-myc (Figure 28), according to the data described so far.

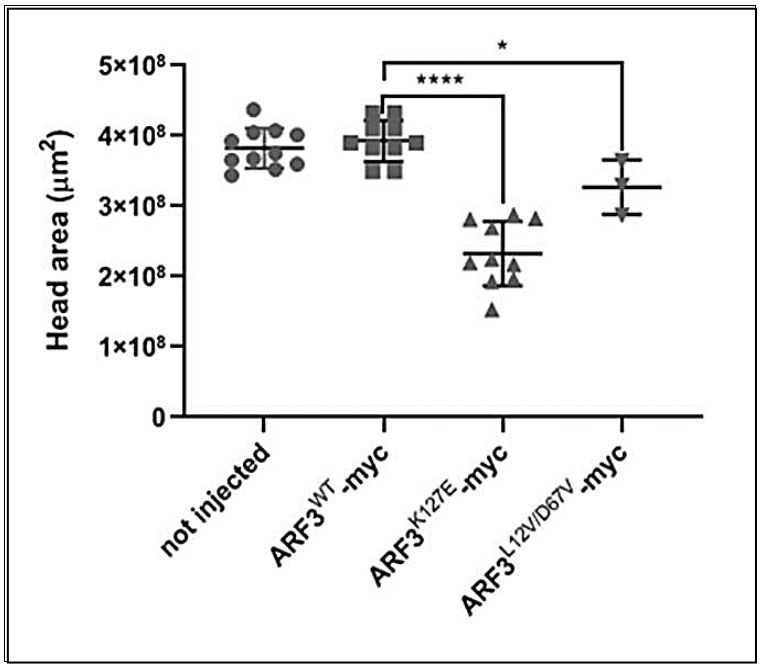


Figure 28: Microcephaly ARF3myc (48hpf) 25 ng. Data obtained and plot with Dr. Giulia Fasano and Dr. Antonella Lauri. The graph shows that reduced head area is present in both *ARF3* mutants. Multiple comparison analysis including the *ARF3*^{WT}-myc control show that *ARF3*^{K127E} –myc mutants has a more severe phenotype than *ARF3*$^{L12V/D67V}$-myc. Asterisks indicate statistically significant differences compared to control (*ARF3*^{WT}) based on one-way ANOVA test with Bonferroni *post hoc* corrections ($^{****}p<0,0001$, $^{*}p<0,05$).

4.5.4 Defects in Convergence and Extension movements analyzed by *in situ* hybridization of *krox20* and *myoD*

Once the defects of the body axis formation in embryos injected with *ARF3*^{K127E}-myc and *ARF3*$^{L12V/D67V}$-myc were confirmed, I investigated further such body establishment malformations. Precisely, I looked into defects of the process of convergence and extension is crucial for body axis formation both in anterior-posterior (AP) and medio-lateral (ML) (Adams. *and* Kimmel. 2004) through the analysis of the expression of *krox-20* and *myoD* mRNA localization during early zebrafish development, by mRNA in situ hybridization (ISH). In the stained embryos I quantified axial morphogenesis defects in *ARF3*^{WT} versus *ARF3* mutants at 15 hours post-fertilization. We chose this developmental stage because it corresponds nearly to the end of the gastrulation process and the initial formation of visible somites (about 14-somite stage). Under Dr. Antonella Lauri guidance, I chose these two genes because they represent excellent markers to detect any axial morphogenesis and convergence extension defects. *Krox20* is indeed expressed during an early phase of development in two rhombomeres, r3 and

r5, and it is very important for hindbrain metameric organization and for cephalic AP landmark (for *Krox20* function see Chapter1, paragraph 1.7, (Aragon and Pujades, 2009; Marin and Charnay, 2000; Charlotte Labalette, *et al*., 2011). *MyoD* is a major regulator in the myogenesis process and it is expressed in somites (for *myoD* function see Chapter1, paragraph 1.7 Hammond CL, *et al*., 2006; Groves JA, *et al*., 2005), important marker for trunk AP elongation and ML morphogenesis

- Krox20 and myoD probes generation

The first step I performed was the synthesis of ISH probe for *Krox20* and *myoD* by PCR amplification using cDNA of 15 hpf embryos (PCR condition in Chapter3, paragraph 3.7.4). Primer sets are given in Table 15.

Amplicon	Forward primer 5'– 3'	Reverse primer 5'– 3'	Size (bp)
ZFkrox20	GTGCACCCTCTTGCCGATAG	GGACGCAGGATTGGCCTGAG	797
ZFmyoD	CCTTGCTTCAACACCAACGA	AAATGGTTTCCTGAGCCTGC	757

Table 15. List of primers used for amplification of *krox20* and *myoD*.

The PCR products were analyzed by agarose gel electrophoresis (Figure 29).

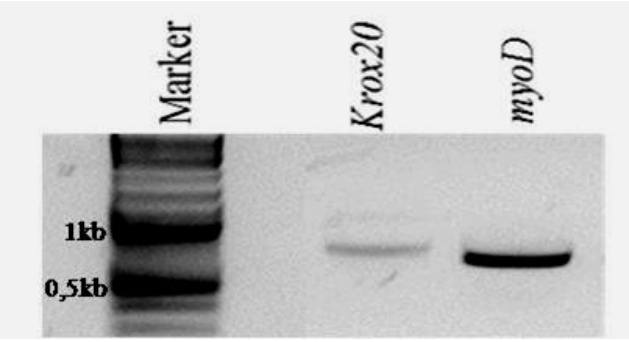


Figure 29: PCR amplification of *krox20* and *myoD* zebrafish

After PCR amplification, I cloned the correct amplicons into pGEM-T easy vector and verified by sequencing of the clones I obtained. Next, I synthesized antisense RNA probes *in vitro* complementary to *krox20* and *myoD* using digoxigenin RNA labeling nucleotide mix (in according to manufacture instructions explained in Chapter 3, methods, paragraph 3.7.5) and I analyzed, by agarose gel electrophoresis, the quality of probes, shown in figure 30.

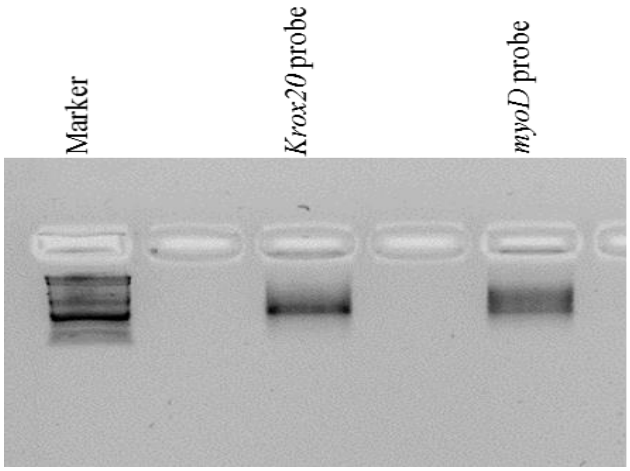


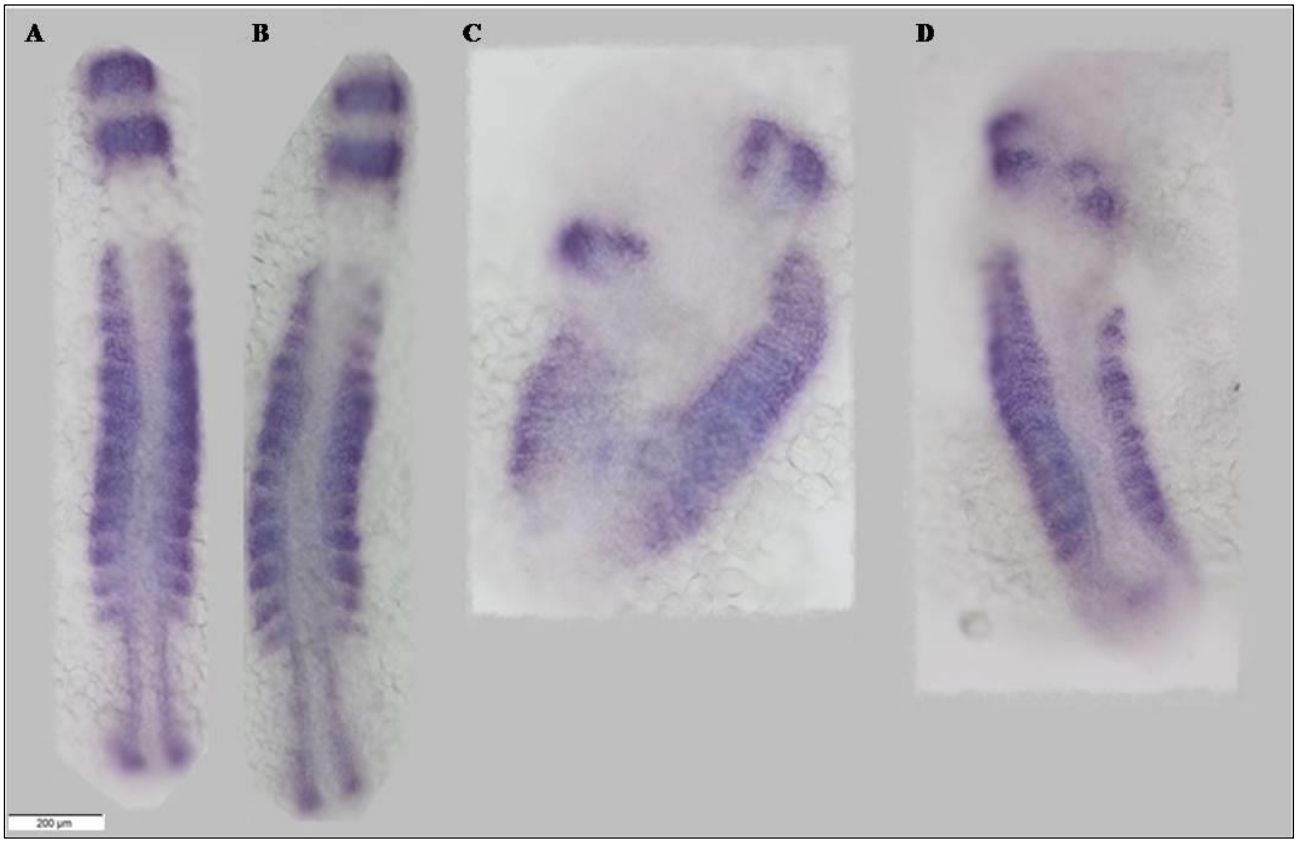
Figure 30: RNA probes of *krox20* and *myoD.* Integrity of the produced mRNA can be observed on the agarose gel. The size is not visible given the fast run of the gel (to avoid RNA degradation).

With highly pure RNA probes I performed whole mount ISH in according to the protocol of Thisse & Thisse, and the protocol kindly taught me by Dr. Antonella Lauri and Giulia Fasano (see paragraph 3.7.6). The hybridization sites on tissues were revealed by a chromogenic reaction with digoxigenin and next I analyzed the expression patterns.

- Results of *krox20* and *myoD* ISH

I carried out in situ hybridization on non-injected and $ARF3^{WT}$ -myc embryos for control, and on mutants embryos, $ARF3^{K127E}$-myc and $ARF3^{L12V/D67V}$-myc.

In order to investigate the impact of *ARF3* mutations on zebrafish morphology, on convergence and extension defects, I defined three different classes of phenotypic outcome. In particular I could describe: as "*normal"* all the embryos that showed a normal body length and axis formation with normal expression of *krox20* and *myoD*, as "*mild"* all the embryos with mild shortening of body axis, visible in the reduction of *myoD* expression in the trunk and with mild increase of medio-lateral axis, as "*severe"* all the embryos showing clear defects in the body axis in AP, severe reduction of the extent of *myoD* expression, increase of ML axis extension and/or lack of convergence process in the formation of axis (Figure31.A-D). I performed penetrance analysis that showed an evident alteration in both *ARF3* mutants, in particular, *ARF3*^{K127E}-myc mutants showed a higher incidence of severe phenotype than *ARF3*$^{L12V/D67V}$-myc (Figure 31.E).



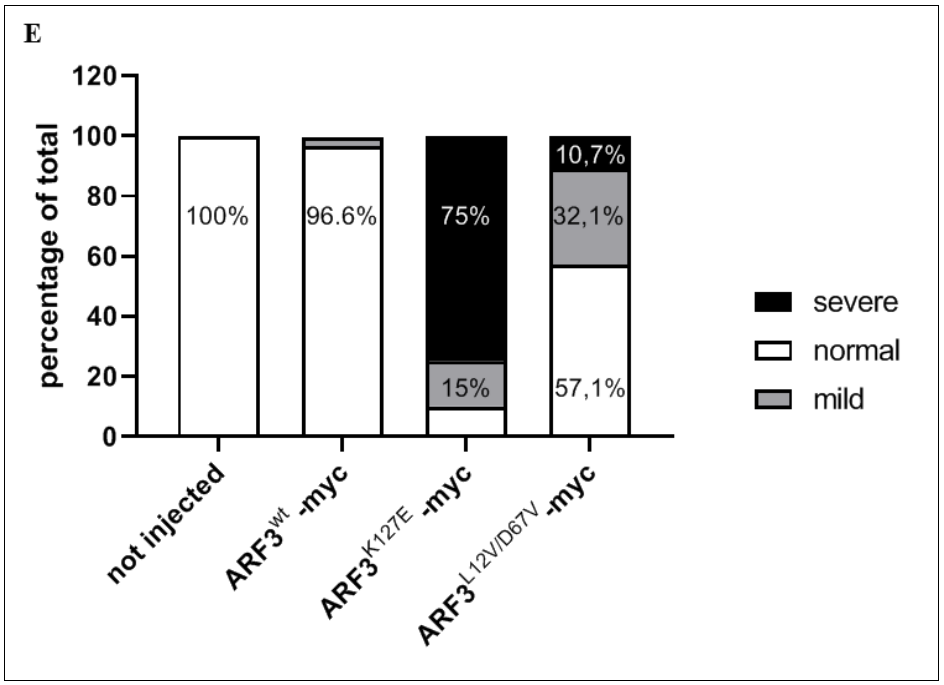


Figure 31: Analysis of body defects in convergence and extension ISH for *krox20* and *myoD* on 15hpf embryos. **A-D.** A dorsal view of embryos co-staining with *krox20* (expressed in r3 and r5) and *myoD* (expressed in somites). Images are obtained using OLYMPUS TH4-200 microscope with 10x objective. (A) Representative images of not-injected embryo; (B) image of *ARF3*^{WT} -myc embryo; (C) image of *ARF3*K127E-myc; (D) image of *ARF3*$^{L12V/D67V}$-myc injected with 25 ng of mRNA. (E) Penetrance analysis of convergence and extension defects in embryos co-staining by visualizing *krox20* and *myoD* transcripts.

I further performed morphometric analyses of *myoD* and *krox20* expression in *ARF3*^{WT}-myc *vs*. *ARF3*^{K127E}-myc and *ARF3*$^{L12V/D67V}$-myc using Olympus CellSens Standard program. For each embryo (n = 10 for each experimental group) I took in consideration four measures for the analysis. In details, I measured:

1. The anterior-posterior (AP) length axis by the length of the embryonic tissue from the start of *krox20* staining until the end of *myoD* staining;
2. The extension of the medio-lateral (ML) axis by the width of the third and fifth rhombomere stained with *krox20.*
3. The extension of the ML of the forming somites by the dimensions of the anterior somite, middle somite and posterior somite stained with *myoD.*
4. The number of visible somites stained with *myoD* within the trunk of the embryo.

Figure 32 shows a schematic expression pattern of *krox20* and *myoD* and all the measures that I have performed (numbered as above).

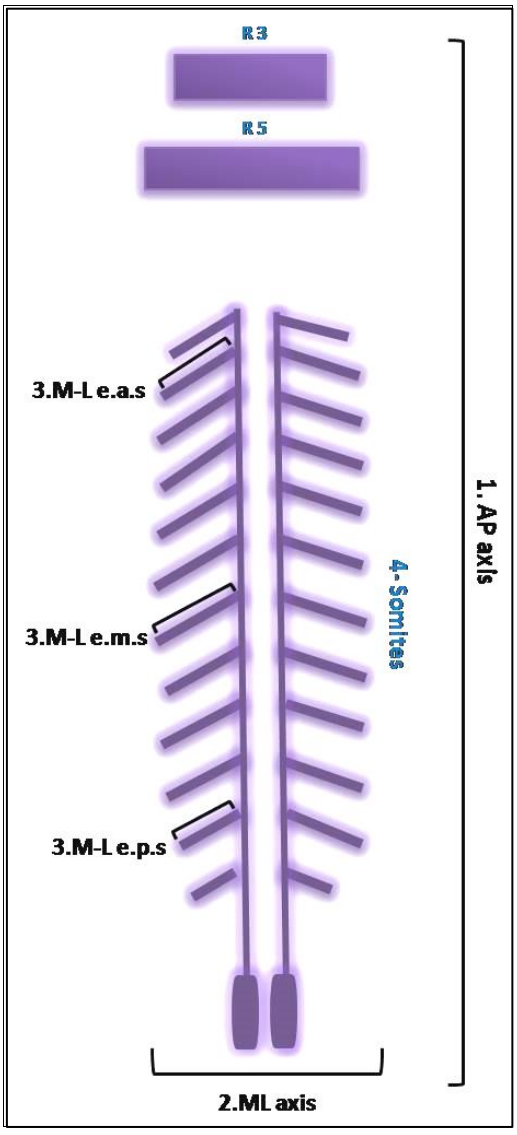


Figure 32: schematic drawing of the distinct territories where *krox20* and *myoD* are expressed at 15hpf. All territories are indicated in purple and represent the staining with BM Purple chromogenic substrate: **r3**=rhombomere 3; **r5**=rhombomere 5, stained with *krox20*. **4.** Somitic mesoderm stained with *myoD*. In the image are also indicated the measures performed for morphometric analyses, in details: **1. AP axis**= anterior-posterior axis length; **2. ML axis=** medio-lateral width axis; **3. M-L e.a.s**= medio-lateral extension of anterior somite; **3. M-L e.m.s=** medio-lateral extension of middle somite; **3. M-L e.p.s=** medio-lateral extension of posterior somite. The measures were performed using Olympus CellSens Standard program.

1/2. Measure of body axis.

I observed that the AP length in both *ARF3* mutant embryos was significantly reduced (p. value<0.0001) and, in particular, likewise the other phenotypes observed so far, stronger effect was observed in *ARF3*^{K127E}-myc compared to *ARF3*$^{L12V/D67V}$-myc (Figure 33.A). In addition, the size of the ML axis of the *ARF3* mutant embryos significantly increased, also in this case, *ARF3*^{K127E}-myc mutants showed an increase of width of ML axis more severe than *ARF3*$^{L12V/D67V}$-myc mutants (p. value<0.0001) (Figure 33.B).

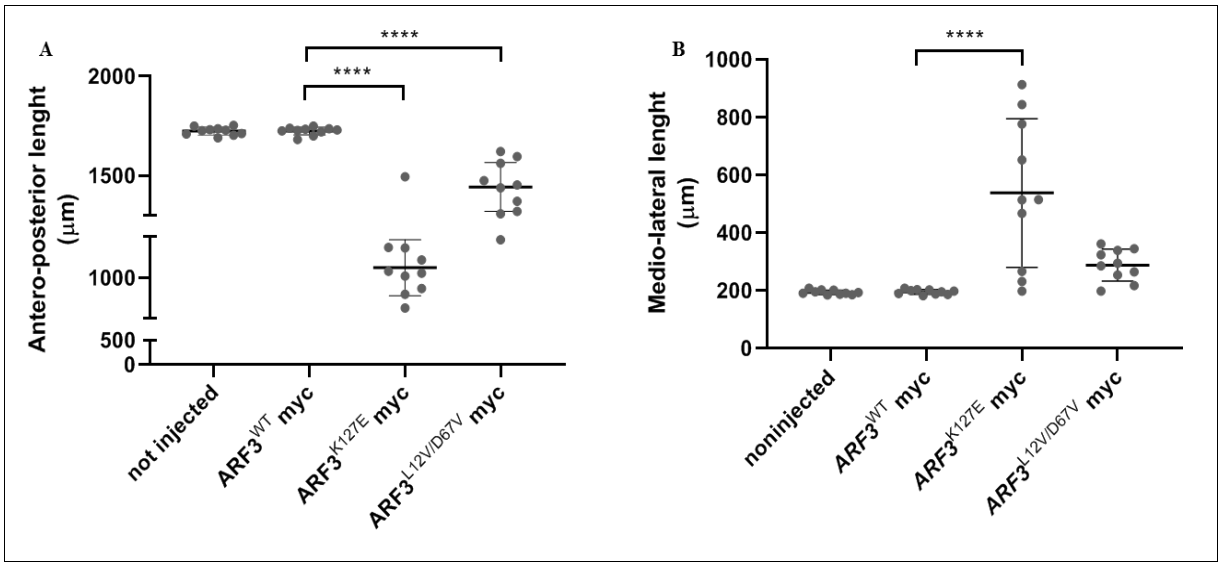


Figure 33: Graphs showing all the data points for AP/ML establishment analysis. **A.** Analysis of AP length of wt and mutant embryos (50ng). Asterisks indicate statistically significant differences compared to control (ARF3^{WT}myc) based on one-way ANOVA test with Sidak's *post hoc* corrections (****p<0.0001). **B.** Analysis of ML width of wt and mutant embryos (50ng). Asterisks indicate statistically significant differences compared to control (ARF3^{WT}myc) based on one-way ANOVA test with Sidak *post hoc* corrections (****p<0.0001).

3. ML extension of somites.

Once determined a significant impact of *ARF3* mutations on the formation of body axis, I asked whether this translated also in a defect in somites extension (also indicating a ML problem). I took into consideration three representative somites: anterior, middle and posterior somites, and I calculated the *"average ML somite extension"* in my embryos population. I observed that *ARF3* mutants, in particular *ARF3*^{K127E}-myc, also

showed abnormal somites with an increase of ML extension compared with the *ARF*^{WT}-myc control (P. value = 0.0001). (Figure 34).

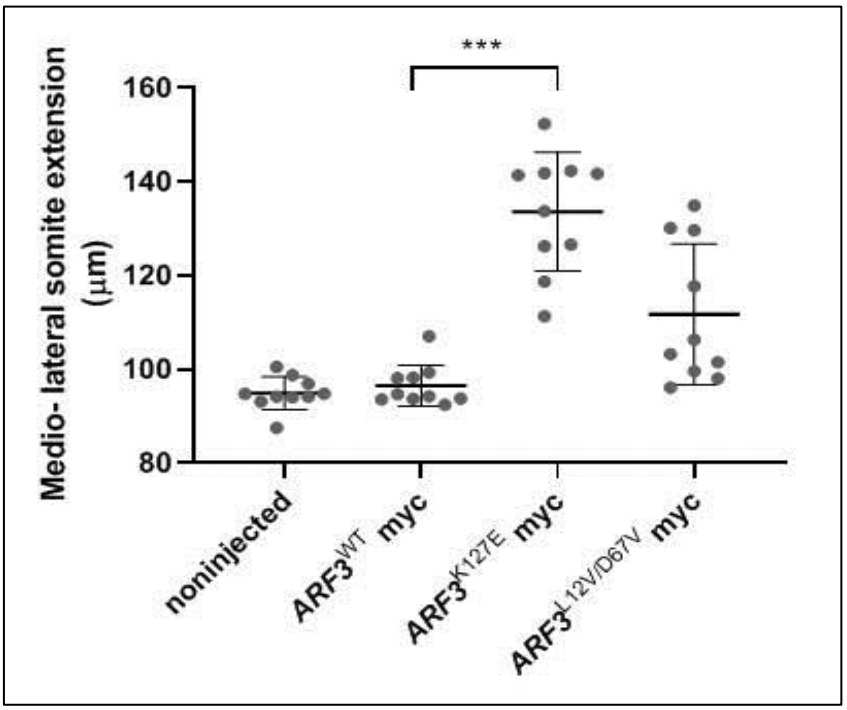


Figure 34: Graphs showing all the data points for ML somite extension. The values obtained represent the mean of the measurements of the anterior, middle and posterior somite, evaluated for each embryo analyzed: not injected, *ARF3*^{WT}-myc, *ARF3*^{K127E}-myc and *ARF3*$^{L12V/D67V}$-myc embryos. Asterisks indicate statistically significant differences compared to control (*ARF3*^{WT}-myc) based on Kruskal-Wallis test with Dunn's *post hoc* corrections ($^{***}p<0.001$).

4. Number of visible somites.

I also investigated the total number of visible somites in the embryos. For this purpose, I counted the number of *myoD+* somites present in both sides of the embryo's body (right and left) and then calculated an average of the number for each side of the body of each embryo. In this case, I observed also a variation in the number of somites, above all in *ARF3*^{K127E} -myc embryos and, mildly in the other mutant. Interestingly, isome embryos presented *myoD+*somitic territories, only on one side of the body, right (dx) or left (sx). Therefore I have analyzed each side separately and compared across conditions also separately. For this region, I considered all somites presented in the embryo making a

comparison between *ARF3*^{WT}-myc with *ARF3*^{K127E}-myc (p. value < 0.0001) and *ARF3*$^{L12V/D67V}$-myc. (P. value =0.0042). (Figure 35).

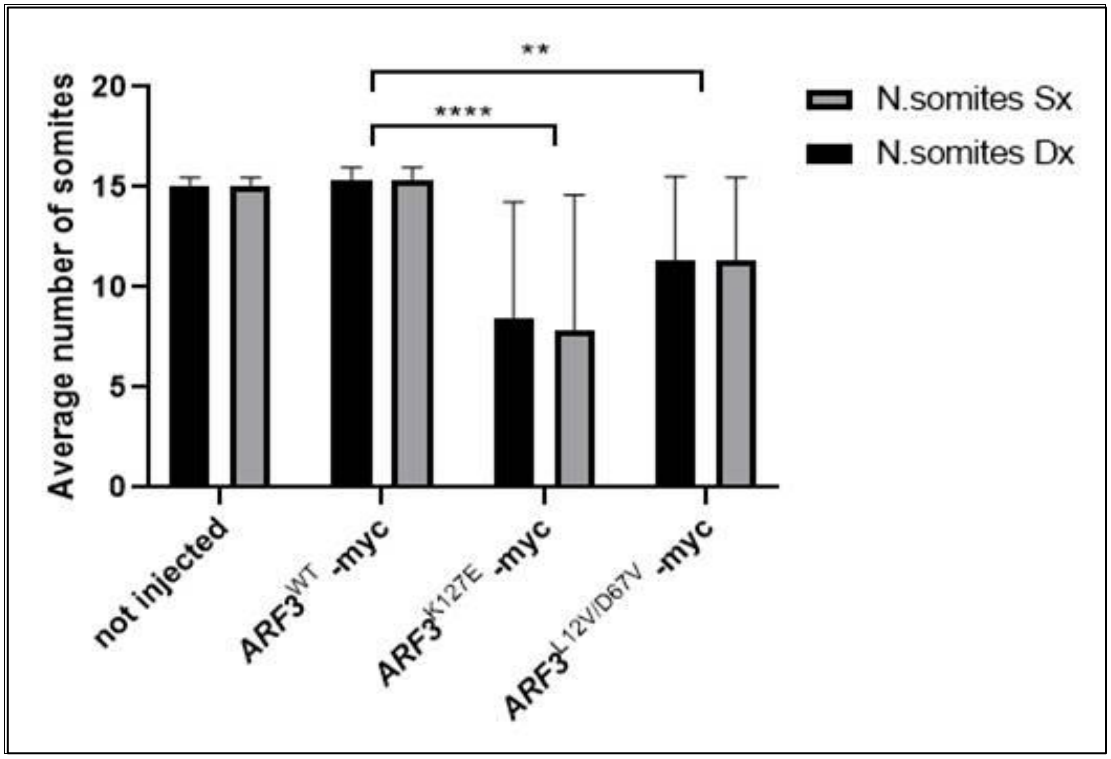


Figure 35: Average number of visible *myoD* + somites along the developing trunk. The values obtained represent the average number of somite in both side of embryos (Dx=right side; Sx=left side). Asterisks indicate statistically significant differences compared to control (*ARF3*^{WT}-myc) based on two-way ANOVA test with Dunn's *post hoc* corrections ($^{****}p<0.0001$; $^{**}p<0.01$). n=10 for each group.

In summary, all the data presented in this section confirmed that *ARF3* mutations in zebrafish embryos have a strong effect on the formation and development of body axis. It is clear that mutant embryos present an alteration of the body axis formation, showing by short stature and increased body curvature. These results brought us to hypothesize deeper defects in the process of convergence and extension, which I confirmed by analyzing the expression of *krox20* and *myoD* at the end of gastrulation. An enlargement of the medio-lateral axis as compared to a shortening of anterior- posterior axis was evident in these early mutant embryos, suggestive of a failure of the underlying convergence and extension movements.

5. Discussion

The overall purpose of this PhD thesis was to obtain more insights into the genetic causes of some undiagnosed rare pediatric conditions mainly affecting the nervous system, through [1] use of modern genomic sequencing technologies that allowed us during the study to identify a number of novel molecular genetics causes of a new forms of epileptic encephalopathies (*KCNK4*,*CLTC*, *ARF3*); [2] contribution to the phenotypic spectrum and genotype-phenotype correlations of one novel and unknown disease gene (*ARF3*) using zebrafish as an *in vivo* model to understand the impact on morphogenetic and developmental programs as well as physiology. In particular, we obtained relevant data on the alteration of cellular processes that coordinate the development of a correct body axis and nervous system

5.1 New perspectives in healthcare: from whole exome sequencing to *in vivo* validation

Rare genetic diseases, for which circa 80% are genetic (EURORDIS. 2007) are highly debilitating, which often dramatically diminish life quality and expectancy (EURORDIS. 2005). In the past three years, the "OPBG Undiagnosed Diseases Program" was activated as a project dedicated to give a molecular diagnosis to orphan patients, relying on WES as first-line approach. In this context, the *"Molecular Genetics and Functional Genomics"* Unit led by Dr. Marco Tartaglia acts as one of main research reference cores. Here, a preliminary internal estimation of the cost revealed a clear gain on the diagnosis costs, with only 3,000 euro needed for a trio-based WES analysis. An optimized workflow, composed by several standardized steps of analysis, was designed in the lab to handle many types of clinical features, starting from a multidisciplinary clinical selection of cases and samples collection for WES to the validation of results, is able, potentially, to achieve the diagnosis in a few weeks. Indeed, clinical re-evaluation of the case is made based on WES results, which can, as instance, lead to a new explanation for the phenotype and to provide a diagnosis to families. Among the 141 probands with a positive finding, in 13 disorders the pathogenic variants (11 *de novo*) were identified in novel genes not previously known to

underlie a Mendelian phenotype, explaining a new syndrome, or found to be implicated in a known but unexplained phenotype.

In this context, my PhD project was focused first at the identification, through the use of exome sequencing, of new genes responsible, when mutated, for the onset of rare neurodevelopment disorders, in particular of different forms of pediatric, isolated and syndromic encephalopathies. We were able to describe a news forms of encephalopathies caused by mutations in genes encoding channels or subunits of ion channel (***KCNK4***; Bauer., *et al*., 2018) and encephalopathies caused by mutations in genes encoding proteins involved in the control of cytoskeletal dynamics and intracellular trafficking (***CLTC***; Manti., *et al*., 2018; and ***ARF3***; manuscript in preparation).

5.1.1 KCNK4, CLTC: two cases of previously unknown disease-genes

KCNK4 is an example of a diagnostic category that can be defined as novel phenotype and novel disease gene. Our findings provide evidence that gain-of-function mutations in KCNK4 cause a distinctive neurodevelopmental syndrome, for which we propose the acronym FHEIG (facial dysmorphism, hypertrichosis, epilepsy, intellectual disability/developmental delay, and gingival overgrowth). Although it will be necessary to recruit other subjects to delineate the clinical heterogeneity, associated with activating of *KCNK4* mutations, thanks to our results, we can still say that a strong association of clinical features and distinctive facies leads us to support the idea that this condition represents a novel nosologic entity. Based on the phenotypic analysis of our patients, the main traits that define this disorder have a phenotypic sharing with activating mutations in the voltage-gated K+ channel KCNH1(MIM: 603305), (Simons, *et al*., 2015; Kortu¨m, *et al*., 2015) and disorders caused by mutations in *ABCC9* (MIM: 601439) and *KCNJ8* (MIM: 600935), two genes encoding ATP-sensitive potassium channel subunits with defective function in Cantu` syndrome (Harakalova, *et al*., 2012; Cooper, *et al*., 2014).Similarly to what is observed in these cases, the subjects with *de novo KCNK4* mutations show variable developmental delay and cognitive impairment, disorganized pattern of development and severe Intellectual disability. However, in contrast to other K2P channels, KCNK4 appears to be almost exclusively expressed in

neuronal cells of the central and peripheral nervous system and the retina. So, the discovery of its pleiotropic impact is unexpected and particular (Bauer., *et al*., 2018).

CLTC is an example of a diagnostic category that can be defined as known phenotype and novel disease gene. We can say this because the clinical features of patients with heterozygous *de novo* mutations in CLTC are known in the literature. In particular, facial dysmorphism, brain malformations, musculoskeletal defects, intellectual disability, and epileptic encephalopathies (EE) are the most common features (Hamdan. *et al*., 2017). Furthermore, it is interesting to note that, an interesting genotype-phenotype correlation can be found in CLTC defect: patients carrying variants in the first segment of clathrin light chain binding domain, exhibit refractory epilepsy; while patients with truncating variants on the protein C-terminus show hypotonia, developmental delay/intellectual disability (Hamdan. *et al*., 2017). In light of this knowledge, we describe a new clinical condition associated with a de novo variant in the CLTC gene, which encodes the clathrin heavy polypeptide. She presents with early-onset developmental delay, subsequent slow intellectual decline, and mood disorder and parkinsonism in adulthood. The finding of mildly elevated blood Phe is compatible with biallelic variants affecting PAH gene, while the unusual depletion of neurotransmitters in cerebrospinal fluid represented an unexplained biochemical feature for us since Phe levels in the blood have no negative effect on brain development and functioning and therefore are unable to explain our phenotype. According to our analysis, we propose that a defective CLTC protein, as such, may result in the depletion of biogenic amines in the brain by altering their synaptic turnover. What led us to hypothesize this defect was the role that the CLTC protein can play in synaptic trafficking. Indeed, it is known in literature that there is a link between proteins that participate in synaptic vesicle recycling and trafficking. (Abeliovich and Gilter., 2016). Clathrins play a key role in endocytosis, intracellular recycling, and trafficking as well as the generation of presynaptic vesicles (Robinson. 2015). The majority of synaptic vesicles appears to recycle through a pathway requiring clathrin, which is involved in the first steps of endocytic vesicle formation at the level of the synaptic terminal. Consistently, *in vitro* and *in vivo* studies, performed in our laboratory have provided evidence that a deficiency in clathrin function causes a progressive decline in synaptic transmission resulting in neurotransmitter depletion in the brain due to aberrant synaptic vesicles formation/turnover (Manti., *et al*., 2018).

5.1.2 ARF3: in-depth understanding of the impact of specific mutations of a previously unknown disease gene on organism physiology

Often more possible disease-causing genes are found in a single patient. How to decide if this result is important? Expression of the gene in the relevant tissues can be considered as additional evidence, as is network analysis. However, making strong statements for private disease genes is, and will be, extremely difficult. Indeed, for the progress of rare diseases studies is necessary to develop experimental models in which candidate variants can be validated functionally and disease mechanisms can be explored. *In vitro* approaches are insufficient to demonstrate pathogenicity at the whole organism level and *in vivo* models are necessary to confirm causality, and investigate the mechanism (Gui H, *et al.,* 2017; Zhang Y. *et al*., 2017). This is true especially for neurological disorders where pathologically relevant tissue is often inaccessible, or for conditions where interactions between cells organs systems are required exist.
Zebrafish combines many genetic and physiological advantages with the high-throughput possibilities and easy experimental handling useful to study disease pathologies and open up to personalized drug screening approaches in the future.
Therefore, in the second half of my PhD work, I have also contributed to define new genotype-phenotype correlations of some of the mutations found in *ARF3* employing zebrafish models.

Starting from the idea that the disease outcome can be modeled reliably in zebrafish embryo, especially for key phenotypic features such as microcephaly, cerebral malformation, developmental delay and skeletal abnormalities, the team here (Dr. Giulia Fasano, Dr. Antonella Lauri and I) used this model system to examine the consequences of h*ARF3* mutations function *in vivo*. Noteworthy, the observed phenotype was resembling the patient's clinical features: the mutant larvae, h*ARF3*^{K127E} and h*ARF3*$^{L12V/D67V}$ displayed a clear phenotype characterized by 1. Defects in body axis formation resulting in short stature and increased body curvature as well as 2. Microcephaly. Following the results, we investigated the analysis of the defects of the

body axis and microcephaly in depth, asking ourselves what could explain this phenotype. We hypothesized a failure in convergence and extension movements that I confirmed via analysis of specific transcripts. In a large part of my study, I could contribute to the cellular mechanism investigation finding specific evidence that corroborated our initial hypothesis. Indeed, my analysis of mRNA expression by *in situ* hybridization of *krox20* and *myoD* reveals an enlargement of the medio-lateral axis compared to a shortening of the anterior-posterior axis in mutant ARF3 embryos.

5.2 Towards detailed investigation of body axis perturbation and microcephaly caused by ARF3 mutations in vertebrates

By analyzing the body curvature and early embryogenesis events, and the head area, we demonstrated that zebrafish mutants had defects in the body axis phenotype and a smaller head size. Thus, the zebrafish mutants' phenotype recapitulated the human axial skeletal defects and microcephaly phenotype. In the light of these results, we asked ourselves what could be the link between the aberrant function of *ARF3* in mutant embryos and the phenotype we observe? To this date, this remains an open question that is currently under examination and with Dr. Antonella Lauri's input we are also currently looking at live morphogenetic movements during early gastrulation time points. Also, in vitro efforts (currently carried out by Dr. Valentina Muto) will help us clarify the nature of the putative intracellular trafficking defects caused by ARF3 mutations. From studies present in the literature, it is known that during vertebrate gastrulation, internalized mesodermal cells undergo convergence and extension movements in a region-specific way (Myers *et al*., 2002). Non-canonical polarity planar cell (PCP) pathways is an important driver during these early phases; regulating cell-oriented movements. An important component of PCP signaling is the transmembrane Frizzled (Fz) receptor, which recruits the Disheveled (Dsh) adapter protein and activates multiple downstream pathways to regulate cytoskeletal structures independent of β-catenin transcriptional activity. Given that acquisition and maintenance of this necessary cell polarity depends on the ability to perform intracellular trafficking and guarantee plasma membrane proteins correct positioning (Frizzled/Dishevelled/Diego and Van Gogh/Prickle PCP complexes, Goodrich and Strutt 2011; Seiferta and Mlodzik 2007;

Wang and Nathans 2007; Singh and Mlodzik 2012), we think it would be interesting to look into this in detail in mutant embryos. The alteration of the mutually exclusive localizations of these complexes should be tested. Interestingly, it has also been shown that in zebrafish mutants, which are deficient in convergence and extension movement, the PCP pathway gene regulates directional migration of a specific population of motor neurons in the hindbrain, as well as morphogenesis of the neuroepithelium (Hironiri and Hitoshi, 2009). Few studies show a connection between the PCP mechanism and the cellular trafficking process. (Carvajal-Gonzalez *et al*., 2015). Thus, with the possible role that ARF3 plays in humans in the control of intracellular trafficking (including vesicular formation and endosomal recycling process of membrane proteins, Donaldson and Jackson, 2011) we would like to test next whether *ARF3* mutations could generate these type of cellular defects depending on altered cellular trafficking, ultimately explaining the convergence and extension impairment that we observed.

Taken together, the results of my work provide new genetic information on previously unknown diseases affecting the nervous system and new mechanistic insights regarding the possible role of ARF3 in body axis establishment and brain development. Further research would be needed to understand the molecular mechanisms and to clarify the pathways underlying the convergence and extension phenomena as well as the causes of microcephaly.

Preliminary results are starting to show the extend of this microcephaly ("brain volume") on the developing nervous system. In the future, we will determine the level of neurogenesis and differentiation as well as neuronal death, which might lead to microcephaly; together with possible defects in cell polarity during these processes. In particular, we will investigate, through real-time PCR analysis and whole-mount immunofluorescence, the expression of the majors' marker of neurogenesis in zebrafish, *sox2*, *elavl3*, *neuroD*, and *neuroG*, in different developmental stages. Understanding these mechanisms will enrich our knowledge on the molecular and cellular phenomena crucial for embryo development and broaden our understanding of genetically-caused microcephaly and axial defects

6. Conclusions

In summary, my PhD work shows the importance of detailed genetic and mechanistic investigation of rare diseases relying on WES as a front-line approach and in vivo validation. A growing number of candidate disease-associated variants are being rapidly identified for many rare disorders. However, a genetic confirmation result is not always easy to obtain. There may be several variables that come into play, and, often, we can have a situation of uncertainty on which is the causative gene/variant for a certain disease. Above all, those cases which we found novel candidate genes, but the absence of phenotype associations or the impossibility of undertaking functional studies (e.g. for unavailability of patient's cells, or in case of genes only expressed at early developmental stages) could preclude providing a definite diagnosis. It is likely that in the next years WES will probably continue to be the main diagnostic tool to study Mendelian traits, considering the advantages it entails (cost savings, minor data storage compared to Whole-genome sequencing and ease of data interpretation). Nevertheless, the diagnostic yield on diseases that remain still undiagnosed despite WES, could be improved by the development of new bioinformatics tools and the expansion of the genetic analysis on of genomic data, such as via whole-genome sequencing (WGS, which gives us the opportunity to study variants in non-coding regions such as promoters, enhancers or other regulatory regions), RNA-seq and analysis of the methylome to better understand complex human diseases.

In this context of the need for new tools, the use of zebrafish as a model system for validating the candidate genes/mechanisms found is also proving to be convenient. Technologies for zebrafish research are evolving rapidly and will open new opportunities towards improving differential diagnosis, as well as our understanding of the molecular mechanisms involved in diseases, needed for pharmaceutical drugs development.

APPENDIX

Identification and functional validation of a new class of SHP2 inhibitors

During the first year of my Ph.D. program, in the frame of a multicenter study conducted in collaboration with Prof. Lorenzo Stella (Università di Roma "Tor Vergata") and Dr. Simone Martinelli (Istituto Superiore di Sanità), aimed at identifying a new class of inhibitors of SHP2, I contributed to perform the biochemical studies. SHP2, encoded by *PTPN11*, is a protein tyrosine phosphatase involved in developmental disorders (*i.e.*, Noonan syndrome and LEOPARD syndrome) and leukemia (Mohi and Neel, 2007; Roberts *et al*., 2013). As a positive modulator of RAS signaling, SHP2 is an ideal target for the development of anti-tumor drugs, particularly for the treatment of juvenile myelomonocytic leukemia (JMML), for which no pharmacological therapy is currently available, as well as post-natal issues of Noonan syndrome (*e.g.*, short stature, hypertrophic cardiomyopathy, cancer susceptibility).

Molecules targeting SHP2 that have been developed so far act by inhibiting the catalytic site of the phosphatase. This approach, however, has proved to be vain. The approach used in the project founded by the Italian Association for Cancer Research (IG19171 to L. Stella) was based on a different strategy. Specifically, aim of the project was to developing peptide mimics targeting the N-SH2 domain of SHP2 and competing with the natural interactors of SHP2 in the cell, in order to inhibit aberrant protein-protein interactions. In the frame of this project, I purified the recombinant wild-type protein and a selected panel of mutants associated with Noonan syndrome or leukemia, and the corresponding isolated phosphatase domain. These proteins have been used to evaluate experimentally the affinity of the best leading peptides identified by *in silico* modelling. I also performed *in vitro* phosphatase assays using para-nitro-phenyl-phosphate as a universal substrate for protein tyrosine phosphatases.

REFERENCES

Adamson, K. I., Sheridan, E., & Grierson, A. J. (2018a). Use of zebrafish models to investigate rare human disease. *Journal of Medical Genetics*, *55*(10), 641–654. https://doi.org/10.1136/jmedgenet-2018-105358

Adler, P. N. (2012). The frizzled/stan pathway and planar cell polarity in the Drosophila wing. *Current Topics in Developmental Biology*, Vol. 101, pp. 1–31. https://doi.org/10.1016/b978-0-12-394592-1.00001-6

Alexandre, P., Reugels, A. M., Barker, D., Blanc, E., & Clarke, J. D. W. (2010). Neurons derive from the more apical daughter in asymmetric divisions in the zebrafish neural tube. *Nature Neuroscience*, *13*(6), 673–679. https://doi.org/10.1038/nn.2547

Aragon, F., & Pujades, C. (2009). FGF signaling controls caudal hindbrain specification through Ras-ERK1/2 pathway. *BMC Developmental Biology*, *9*(1). https://doi.org/10.1186/1471-213X-9-61

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N., & Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes and Development*, *17*(1), 126–140. https://doi.org/10.1101/gad.224503

Balikova, I., Robson, A. G., Holder, G. E., Ostergaard, P., Mansour, S., & Moore, A. T. (2016a). Ocular manifestations of microcephaly with or without chorioretinopathy, lymphedema or intellectual disability (MCLID) syndrome associated with mutations in KIF11. *Acta Ophthalmologica*, *94*(1), 92–98. https://doi.org/10.1111/aos.12759

Balikova, I., Robson, A. G., Holder, G. E., Ostergaard, P., Mansour, S., & Moore, A. T. (2016b). Ocular manifestations of microcephaly with or without chorioretinopathy, lymphedema or intellectual disability (MCLID) syndrome associated with mutations in KIF11. *Acta Ophthalmologica*, *94*(1), 92–98. https://doi.org/10.1111/aos.12759

Bauer, C. K., Calligari, P., Radio, F. C., Caputo, V., Dentici, M. L., Falah, N., Tartaglia, M. (2018). Mutations in KCNK4 that Affect Gating Cause a Recognizable Neurodevelopmental Syndrome. *American Journal of Human Genetics*, *103*(4), 621–630. https://doi.org/10.1016/j.ajhg.2018.09.001

Bayly, R., & Axelrod, J. D. (2011). Pointing in the right direction: New developments in the field of planar cell polarity. *Nature Reviews Genetics*, *12*(6), 385–391. https://doi.org/10.1038/nrg2956

Ben-Ari, Y., & Holmes, G. L. (2006). Effects of seizures on developmental processes in the immature brain. *Lancet Neurology*, Vol. 5, pp. 1055–1063. https://doi.org/10.1016/S1474-4422(06)70626-3

Bénazéraf, B., & Pourquié, O. (2013). Formation and Segmentation of the Vertebrate Body Axis. *Annual Review of Cell and Developmental Biology*, *29*(1), 1–26. https://doi.org/10.1146/annurev-cellbio-101011-155703

Berg, A. T., Langfitt, J. T., Testa, F. M., Levy, S. R., DiMario, F., Westerveld, M., & Kulas, J. (2008). Global cognitive function in children with epilepsy: A community-based study. *Epilepsia*, *49*(4), 608–614. https://doi.org/10.1111/j.1528-1167.2007.01461.x

Bonafé, L., Blau, N., Burlina, A. P., Romstad, A., Güttler, F., & Burlina, A. B. (2001). Treatable neurotransmitter deficiency in mild phenylketonuria. *Neurology*, *57*(5), 908–911. https://doi.org/10.1212/WNL.57.5.908

Bonsignorio, D., Perego, L., Del Giacco, L., & Cotelli, F. (1996). Structure and macromolecular composition of the zebrafish egg chorion. *Zygote*, *4*(2), 101–108. https://doi.org/10.1017/S0967199400002975

Bland, J. Martin, and Douglas G. Altman. 1999. "Measuring Agreement in Method Comparison Studies." *Statistical Methods in Medical Research*. https://doi.org/10.1191/096228099673819272.

Branam, A. M., Hoffman, G. G., Pelegri, F., & Greenspan, D. S. (2010). Zebrafish chordin-like and chordin are functionally redundant in regulating patterning of the dorsoventral axis. *Developmental Biology*, *341*(2), 444–458. https://doi.org/10.1016/j.ydbio.2010.03.001

Brand, M. (1998). Genetic analysis of early development of a vertebrate nervous system in the zebrafish, Danio rerio. *Zoology*, Vol. 101, pp. 345–364.

Brand, M., Heisenberg, C. P., Jiang, Y. J., Beuchle, D., Lun, K., Furutani-Seiki, M., … Nüsslein-Volhard, C. (1996). Mutations in zebrafish genes affecting the formation of the boundary between midbrain and hindbrain. *Development*, *123*, 179–190.

Brooks-Kayal, A. (2011). Molecular mechanisms of cognitive and behavioral comorbidities of epilepsy in children. *Epilepsia*, *52*(SUPPL. 1), 13–20. https://doi.org/10.1111/j.1528-1167.2010.02906.x

Burke, E. A., Frucht, S. J., Thompson, K., Wolfe, L. A., Yokoyama, T., Bertoni, M., … Malicdan, M. C. V. (2018). Biallelic mutations in mitochondrial tryptophanyl-tRNA synthetase cause Levodopa-responsive infantile-onset Parkinsonism. *Clinical Genetics*, *93*(3), 712–718. https://doi.org/10.1111/cge.13172

C.-C., S., M., C., J.P., S., & M., G. (2000). Potassium channels: Molecular defects, diseases, and therapeutic opportunities. *Pharmacological Reviews*.

Carlos Amor, J., Harrison, D. H., Kahn, R. A., & Ringe, D. (1994). Structure of the human ADP-ribosylation factor 1 complexed with GDP. *Nature*, *372*(6507), 704–708. https://doi.org/10.1038/372704a0

Carvajal-Gonzalez, J. M., Balmer, S., Mendoza, M., Dussert, A., Collu, G., Roman, A. C., … Mlodzik, M. (2015). The clathrin adaptor AP-1 complex and Arf1 regulate planar cell polarity in vivo. *Nature Communications*, *6*. https://doi.org/10.1038/ncomms7751

Cavenagh, M. M., Whitney, J. A., Carroll, K., Zhang, C. J., Boman, A. L., Rosenwald, A. G., … Kahn, R. A. (1996). Intracellular distribution of Arf proteins in mammalian cells: Arf6 is uniquely localized to the plasma membrane. *Journal of Biological Chemistry*, *271*(36), 21767–21774. https://doi.org/10.1074/jbc.271.36.21767

Chal, J., & Pourquié, O. (2017). Making muscle: Skeletal myogenesis in vivo and in vitro. *Development (Cambridge)*, Vol. 144, pp. 2104–2122. https://doi.org/10.1242/dev.151035

Chan, G., Kalaitzidis, D., & Neel, B. G. (2008). The tyrosine phosphatase Shp2 (PTPN11) in cancer. *Cancer and Metastasis Reviews*, Vol. 27, pp. 179–192. https://doi.org/10.1007/s10555-008-9126-y

Chan, R. J., & Feng, G. S. (2007). PTPN11 is the first identified proto-oncogene that encodes a tyrosine phosphatase. *Blood*, Vol. 109, pp. 862–867. https://doi.org/10.1182/blood-2006-07-028829

Cheng, Y. C., Amoyel, M., Qiu, X., Jiang, Y. J., Xu, Q., & Wilkinson, D. G. (2004). Notch activation regulates the segregation and differentiation of rhombomere boundary cells in the zebrafish hindbrain. *Developmental Cell*, *6*(4), 539–550. https://doi.org/10.1016/S1534-5807(04)00097-8

Chiamvimonvat, N., Chen-Izu, Y., Clancy, C. E., Deschenes, I., Dobrev, D., Heijman, J., … Nerbonne, J. M. (2017). Potassium currents in the heart: functional roles in repolarization, arrhythmia and therapeutics. *Journal of Physiology*, *595*(7), 2229–2252. https://doi.org/10.1113/JP272883

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., & Rutter, W. J. (1979). Isolation of Biologically Active Ribonucleic Acid from Sources Enriched in Ribonuclease. *Biochemistry*, *18*(24), 5294–5299. https://doi.org/10.1021/bi00591a005

Cho, K. W. Y., & De Robertis, E. M. (1990). Differential activation of Xenopus homeo box genes by mesoderm-inducing growth factors and retinoic acid. *Genes and Development*, *4*(11), 1910–1916. https://doi.org/10.1101/gad.4.11.1910

Choi, S. C., & Han, J. K. (2002). Xenopus Cdc42 regulates convergent extension movements during gastrulation through Wnt/Ca2+ signaling pathway. *Developmental Biology*, *244*(2), 342–357. https://doi.org/10.1006/dbio.2002.0602

Chong, J. X., Caputo, V., Phelps, I. G., Stella, L., Worgan, L., Dempsey, J. C., … Doherty, D. (2016). Recessive Inactivating Mutations in TBCK, Encoding a Rab GTPase-Activating Protein, Cause Severe Infantile Syndromic Encephalopathy. *American Journal of Human Genetics*, *98*(4), 772–781. https://doi.org/10.1016/j.ajhg.2016.01.016

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., … Ruden, D. M. (2012). 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*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695

Ciruna, B., Jenny, A., Lee, D., Mlodzik, M., & Schier, A. F. (2006). Planar cell polarity

signalling couples cell division and morphogenesis during neurulation. *Nature*, *439*(7073), 220–224. https://doi.org/10.1038/nature04375

Claros, M. G., Bautista, R., Guerrero-Fernández, D., Benzerki, H., Seoane, P., & Fernández-Pozo, N. (2014). Why assembling plant genome sequences is so challenging. In *The Role of Bioinformatics in Agriculture* (pp. 27–54). https://doi.org/10.1201/b16568

Cnossen, W. R., Te Morsche, R. H. M., Hoischen, A., Gilissen, C., Chrispijn, M., Venselaar, H., … Drenth, J. P. H. (2014). Whole-exome sequencing reveals LRP5 mutations and canonical Wnt signaling associated with hepatic cystogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(14), 5343–5348. https://doi.org/10.1073/pnas.1309438111

Cooper, Paige E., Heiko Reutter, Joachim Woelfle, Hartmut Engels, Dorothy K. Grange, Gijs van Haaften, Bregje W. van Bon, Alexander Hoischen, and Colin G. Nichols. 2014. "Cantú Syndrome Resulting from Activating Mutation in the KCNJ8 Gene." *Human Mutation*. https://doi.org/10.1002/humu.22555.

Cunningham, M. L., Mac Auley, A., & Mirkes, P. E. (1994). From gastrulation to neurulation: Transition in retinoic acid sensitivity identifies distinct stages of neural patterning in the rat. *Developmental Dynamics*, *200*(3), 227–241. https://doi.org/10.1002/aja.1002000305

Depristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–501. https://doi.org/10.1038/ng.806

Dharmat, R., Liu, W., Ge, Z., Sun, Z., Yang, L., Li, Y., … Chen, R. (2017). IFT81 as a candidate gene for nonsyndromic retinal degeneration. *Investigative Ophthalmology and Visual Science*, *58*(5), 2483–2490. https://doi.org/10.1167/iovs.16-19133

Diekmann, Y., Seixas, E., Gouw, M., Tavares-Cadete, F., Seabra, M. C., & Pereira-Leal, J. B. (2011). Thousands of Rab GTPases for the cell biologist. *PLoS Computational Biology*, *7*(10). https://doi.org/10.1371/journal.pcbi.1002217

Doldur-Balli, F., Ozel, M. N., Gulsuner, S., Tekinay, A. B., Ozcelik, T., Konu, O., & Adams, M. M. (2015). Characterization of a novel zebrafish (Danio rerio) gene, wdr81, associated with cerebellar ataxia, mental retardation and dysequilibrium syndrome (CAMRQ). *BMC Neuroscience*, *16*(1). https://doi.org/10.1186/s12868-015-0229-4

Domingos, P. M., Itasaki, N., Jones, C. M., Mercurio, S., Sargent, M. G., Smith, J. C., & Krumlauf, R. (2001). The Wnt/β-catenin pathway posteriorizes neural tissue in Xenopus by an indirect mechanism requiring FGF signalling. *Developmental Biology*, *239*(1), 148–160. https://doi.org/10.1006/dbio.2001.0431

Donaldson, J. G., & Jackson, C. L. (2011). ARF family G proteins and their regulators: Roles in membrane transport, development and disease. *Nature Reviews Molecular Cell Biology*, Vol. 12, pp. 362–375. https://doi.org/10.1038/nrm3117

Dong, Z., Yang, N., Yeo, S. Y., Chitnis, A., & Guo, S. (2012). Intralineage Directional Notch Signaling Regulates Self-Renewal and Differentiation of Asymmetrically Dividing

Radial Glia. *Neuron*, *74*(1), 65–78. https://doi.org/10.1016/j.neuron.2012.01.031

Doniach, T., & Musci, T. J. (1995). Induction of anteroposterior neural pattern in Xenopus: evidence for a quantitative mechanism. *Mechanisms of Development*, *53*(3), 403–413. https://doi.org/10.1016/0925-4773(95)00457-2

Dow, L. E., & Lowe, S. W. (2012). Life in the fast lane: Mammalian disease models in the genomics era. *Cell*, Vol. 148, pp. 1099–1109. https://doi.org/10.1016/j.cell.2012.02.023

Esmaeeli Nieh, S., & Sherr, E. H. (2014). Epileptic Encephalopathies: New Genes and New Pathways. *Neurotherapeutics*, Vol. 11, pp. 796–806. https://doi.org/10.1007/s13311-014-0301-2

EURORDIS Rare Diseases Europe. (2007). What Is a Rare Disease? *Rare Diseases Europe*, 14–15.

Faheem, M., Naseer, M. I. mra., Rasool, M., Chaudhary, A. G., Kumosani, T. A., Ilyas, A. M. uhamma., … Saleh Jamal, H. (2015). Molecular genetics of human primary microcephaly: an overview. *BMC Medical Genomics*, Vol. 8, p. S4. https://doi.org/10.1186/1755-8794-8-S1-S4

Feingold, M., & Bartoshesky, L. (1992). Microcephaly, lymphedema, and chorioretinal dysplasia: A distinct syndrome? *American Journal of Medical Genetics*, *43*(6), 1030–1031. https://doi.org/10.1002/ajmg.1320430623

Fink, M., Lesage, F., Duprat, F., Heurteaux, C., Reyes, R., Fosset, M., & Lazdunski, M. (1998). A neuronal two P domain K+ channel stimulated by arachidonic acid and polyunsaturated fatty acids. *EMBO Journal*, *17*(12), 3297–3308. https://doi.org/10.1093/emboj/17.12.3297

Fisch, R. O., Ketterling, W. C., Schacht, L. E., & Letson, R. D. (1973). Ocular abnormalities of a child associated with familial microcephaly. *American Journal of Ophthalmology*, *76*(2), 260–264. https://doi.org/10.1016/0002-9394(73)90171-2

Flex, E., Niceta, M., Cecchetti, S., Thiffault, I., Au, M. G., Capuano, A., … Tartaglia, M. (2016). Biallelic Mutations in TBCD, Encoding the Tubulin Folding Cofactor D, Perturb Microtubule Dynamics and Cause Early-Onset Encephalopathy. *American Journal of Human Genetics*, *99*(4), 962–973. https://doi.org/10.1016/j.ajhg.2016.08.003

Friedrich, R. W., & Korsching, S. I. (1997). Combinatorial and chemotopic odorant coding in the zebrafish olfactory bulb visualized by optical imaging. *Neuron*, *18*(5), 737–752. https://doi.org/10.1016/S0896-6273(00)80314-1

Gass, J. D. M., & Taney, B. S. (1994). Flecked Retina Associated With Café au Lait Spots, Microcephaly, Epilepsy, Short Stature, and Ring 17 Chromosome. *Archives of Ophthalmology*, Vol. 112, pp. 738–739. https://doi.org/10.1001/archopht.1994.01090180036013

Ghasemi, Asghar, and Saleh Zahediasl. 2012. “Normality Tests for Statistical Analysis: A Guide for Non-Statisticians.” *International Journal of Endocrinology and Metabolism*.

https://doi.org/10.5812/ijem.3505.

Ghosh, P., Maurer, J. M., & Sagerström, C. G. (2018). Analysis of novel caudal hindbrain genes reveals different regulatory logic for gene expression in rhombomere 4 versus 5/6 in embryonic zebrafish. *Neural Development*, *13*(1). https://doi.org/10.1186/s13064-018-0112-y

Glickman, N. S., Kimmel, C. B., Jones, M. A., & Adams, R. J. (2003). Shaping the zebrafish notochord. *Development*, Vol. 130, pp. 873–887. https://doi.org/10.1242/dev.00314

Goodrich, L. V., & Strutt, D. (2011). Principles of planar polarity in animal development. *Development*, Vol. 138, pp. 1877–1892. https://doi.org/10.1242/dev.054080

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, Vol. 17, pp. 333–351. https://doi.org/10.1038/nrg.2016.49

Groves, J. A., Hammond, C. L., & Hughes, S. M. (2005). Fgf8 drives myogenic progression of a novel lateral fast muscle fibre population in zebrafish. *Development*, *132*(19), 4211–4222. https://doi.org/10.1242/dev.01958

Gui, H., Schriemer, D., Cheng, W. W., Chauhan, R. K., Antiňolo, G., Berrios, C., … Hofstra, R. M. W. (2017). Whole exome sequencing coupled with unbiased functional analysis reveals new Hirschsprung disease genes. *Genome Biology*, *18*(1). https://doi.org/10.1186/s13059-017-1174-6

Hamdan, F. F., Myers, C. T., Cossette, P., Lemay, P., Spiegelman, D., Laporte, A. D., … Michaud, J. L. (2017). High Rate of Recurrent De Novo Mutations in Developmental and Epileptic Encephalopathies. *American Journal of Human Genetics*, *101*(5), 664–685. https://doi.org/10.1016/j.ajhg.2017.09.008

Hammond, C. L., Hinits, Y., Osborn, D. P. S., Minchin, J. E. N., Tettamanti, G., & Hughes, S. M. (2007). Signals and myogenic regulatory factors restrict pax3 and pax7 expression to dermomyotome-like tissue in zebrafish. *Developmental Biology*, *302*(2), 504–521. https://doi.org/10.1016/j.ydbio.2006.10.009

Hanai, A., Ohgi, M., Yagi, C., Ueda, T., Shin, H. W., & Nakayama, K. (2015). Class i Arfs (Arf1 and Arf3) and Arf6 are localized to the Flemming body and play important roles in cytokinesis. *Journal of Biochemistry*, *159*(2), 201–208. https://doi.org/10.1093/jb/mvv088

Harakalova, Magdalena, Jeske J.T. Van Harssel, Paulien A. Terhal, Stef Van Lieshout, Karen Duran, Ivo Renkens, David J. Amor, et al. 2012. "Dominant Missense Mutations in ABCC9 Cause Cantúsyndrome." *Nature Genetics*. https://doi.org/10.1038/ng.2324.

Heisenberg, C. P., Tada, M., Rauch, G. J., Saúde, L., Concha, M. L., Geisler, R., … Wilson, S. W. (2000). Silberblick/Wnt11 mediates convergent extension movements during zebrafish gastrulation. *Nature*, *405*(6782), 76–81. https://doi.org/10.1038/35011068

Hillman, E. M., Voleti, V., Patel, K., Li, W., Yu, H., Perez-Campos, C., … Galwaduge, P. T.

(2018). High-speed 3D imaging of cellular activity in the brain using axially-extended beams and light sheets. *Current Opinion in Neurobiology*, Vol. 50, pp. 190–200. https://doi.org/10.1016/j.conb.2018.03.007

Holley, S. A. (2006). Anterior-posterior differences in vertebrate segments: Specification of trunk and tail somites in the zebrafish blastula. *Genes and Development*, Vol. 20, pp. 1831–1837. https://doi.org/10.1101/gad.1453706

Holowacz, T., & Sokol, S. (1999). FGF is required for posterior neural patterning but not for neural induction. *Developmental Biology*, *205*(2), 296–308. https://doi.org/10.1006/dbio.1998.9108

Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., … Stemple, D. L. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature*, *496*(7446), 498–503. https://doi.org/10.1038/nature12111

Ijaz, S., & Hoffman, E. J. (2016). Zebrafish: A Translational Model System for Studying Neuropsychiatric Disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *55*(9), 746–748. https://doi.org/10.1016/j.jaac.2016.06.008

Jagadeesh, K. A., Wenger, A. M., Berger, M. J., Guturu, H., Stenson, P. D., Cooper, D. N., … Bejerano, G. (2016). M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nature Genetics*, *48*(12), 1581–1586. https://doi.org/10.1038/ng.3703

Jarmas, A. L., Weaver, D. D., Ellis, F. D., & Davis, A. (1981). Microcephaly, Microphthalmia, Falciform Retinal Folds, and Blindness: A New Syndrome. *American Journal of Diseases of Children*, *135*(10), 930–933. https://doi.org/10.1001/archpedi.1981.02130340036013

Jentsch, T. J. (2000). Neuronal KCNQ potassium channels: Physislogy and role in disease. *Nature Reviews Neuroscience*, *1*(1), 21–30. https://doi.org/10.1038/35036198

Jones, C., & Chen, P. (2007). Planar cell polarity signaling in vertebrates. *BioEssays*, Vol. 29, pp. 120–132. https://doi.org/10.1002/bies.20526

Karlstrom, R. O., Trowe, T., & Bonhoeffer, F. (1997). Genetic analysis of axon guidance and mapping in the zebrafish. *Trends in Neurosciences*, *20*(1), 3–8. https://doi.org/10.1016/S0166-2236(96)40005-4

Kasprowicz, J., Kuenen, S., Miskiewicz, K., Habets, R. L. P., Smitz, L., & Verstreken, P. (2008). Inactivation of clathrin heavy chain inhibits synaptic recycling but allows bulk membrane uptake. *Journal of Cell Biology*, *182*(5), 1007–1016. https://doi.org/10.1083/jcb.200804162

Keller, R., Davidson, L., Edlund, A., Elul, T., Ezin, M., Shook, D., & Skoglund, P. (2000). Mechanisms of convergence and extension by cell intercalation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *355*(1399), 897–922. https://doi.org/10.1098/rstb.2000.0626

Kiecker, C., & Lumsden, A. (2012). The Role of Organizers in Patterning the Nervous System. *Annual Review of Neuroscience*, *35*(1), 347–367. https://doi.org/10.1146/annurev-neuro-062111-150543

Kimelman, D., & Martin, B. L. (2012). Anterior-posterior patterning in early development: Three strategies. *Wiley Interdisciplinary Reviews: Developmental Biology*, *1*(2), 253–266. https://doi.org/10.1002/wdev.25

Kimmel, C. B. (1989). Genetics and early development of zebrafish. *Trends in Genetics*, Vol. 5, pp. 283–288. https://doi.org/10.1016/0168-9525(89)90103-0

Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., & Schilling, T. F. (1995). Stages of embryonic development of the zebrafish. *Developmental Dynamics*, *203*(3), 253–310. https://doi.org/10.1002/aja.1002030302

Kimmel, C. B., & Warga, R. M. (1986). Tissue-specific cell lineages originate in the gastrula of the zebrafish. *Science*, *231*(4736), 365–368. https://doi.org/10.1126/science.231.4736.365

Kimmel, C. B., Warga, R. M., & Kane, D. A. (1994). Cell cycles and clonal strings during formation of the zebrafish central nervous system. *Development*, *120*(2), 265–276.

Kircher, M., Witten, D. M., Jain, P., O'roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*, *46*(3), 310–315. https://doi.org/10.1038/ng.2892

Kita, E. M., Scott, E. K., & Goodhill, G. J. (2015). Topographic wiring of the retinotectal connection in zebrafish. *Developmental Neurobiology*, *75*(6), 542–556. https://doi.org/10.1002/dneu.22256

Kitazawa, T., & Rijli, F. M. (2018). Integrating into the Rhombomere Community across the Border. *Developmental Cell*, Vol. 45, pp. 546–548. https://doi.org/10.1016/j.devcel.2018.05.019

Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K., & Mardis, E. R. (2013). XThe next-generation sequencing revolution and its impact on genomics. *Cell*, Vol. 155, p. 27. https://doi.org/10.1016/j.cell.2013.09.006

Köppen, M., García Fernández, B., Carvalho, L., Jacinto, A., & Heisenberg, C. P. (2006). Coordinated cell-shape changes control epithelial movement in zebrafish and Drosophila. *Development*, *133*(14), 2671–2681. https://doi.org/10.1242/dev.02439

Kortüm, F., Caputo, V., Bauer, C. K., Stella, L., Ciolfi, A., Alawi, M., … Kutsche, K. (2015). Mutations in KCNH1 and ATP6V1B2 cause Zimmermann-Laband syndrome. *Nature Genetics*, *47*(6), 661–667. https://doi.org/10.1038/ng.3282

Korzh, V. (2018). Development of brain ventricular system. *Cellular and Molecular Life Sciences*, Vol. 75, pp. 375–383. https://doi.org/10.1007/s00018-017-2605-y

Kozol, R. A., Abrams, A. J., James, D. M., Buglo, E., Yan, Q., & Dallman, J. E. (2016). Function over form: Modeling groups of inherited neurological conditions in zebrafish.

Frontiers in Molecular Neuroscience, Vol. 9. https://doi.org/10.3389/fnmol.2016.00055

Kudoh, T., Wilson, S. W., & Dawid, I. B. (2002). Distinct roles for Fgf, Wnt and retinoic acid in posteriorizing the neural ectoderm. *Development*, Vol. 129, pp. 4335–4346.

Labalette, C., Wassef, M. A., Desmarquet-Trin Dinh, C., Bouchoucha, Y. X., Le Men, J., Charnay, P., & Gilardi-Hebenstreit, P. (2015). Molecular dissection of segment formation in the developing hindbrain. *Development (Cambridge)*, *142*(1), 185–195. https://doi.org/10.1242/dev.109652

Lee, C. M., Haun, R. S., Tsai, S. C., Moss, J., & Vaughan, M. (1992). Characterization of the human gene encoding ADP-ribosylation factor 1, a guanine nucleotide-binding activator of cholera toxin. *Journal of Biological Chemistry*, *267*(13), 9028–9034.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Preprint ArXiv*. https://doi.org/arXiv:1303.3997 [q-bio.GN]

Li, Q., & Wang, K. (2017). InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *American Journal of Human Genetics*, *100*(2), 267–280. https://doi.org/10.1016/j.ajhg.2017.01.004

Lin, B., Wang, J., & Cheng, Y. (2010). Recent Patents and Advances in the Next-Generation Sequencing Technologies. *Recent Patents on Biomedical Engineeringe*, *1*(1), 60–67. https://doi.org/10.2174/1874764710801010060

Lin, C. Y., Yung, R. F., Lee, H. C., Chen, W. T., Chen, Y. H., & Tsai, H. J. (2006). Myogenic regulatory factors Myf5 and Myod function distinctly during craniofacial myogenesis of zebrafish. *Developmental Biology*, *299*(2), 594–608. https://doi.org/10.1016/j.ydbio.2006.08.042

Lin, Y. J., Bovetto, S., Carver, J. M., & Giordano, T. (1996). Cloning of the cDNA for the human NMDA receptor NR2C subunit and its expression in the central nervous system and periphery. *Molecular Brain Research*, *43*(1–2), 57–64. https://doi.org/10.1016/S0169-328X(96)00146-5

Liu, J. S. (2011). Molecular genetics of neuronal migration disorders. *Current Neurology and Neuroscience Reports*, Vol. 11, pp. 171–178. https://doi.org/10.1007/s11910-010-0176-5

Liu, J., Zhou, Y., Qi, X., Chen, J., Chen, W., Qiu, G., … Wu, N. (2017). CRISPR/Cas9 in zebrafish: an efficient combination for human genetic diseases modeling. *Human Genetics*, Vol. 136. https://doi.org/10.1007/s00439-016-1739-6

Liu, X., Jian, X., & Boerwinkle, E. (2013). dbNSFP v2.0: A database of human non-synonymous SNVs and their functional predictions and annotations. *Human Mutation*, *34*(9). https://doi.org/10.1002/humu.22376

Liu, X., Wu, C., Li, C., & Boerwinkle, E. (2016). dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Human Mutation*, *37*(3), 235–241. https://doi.org/10.1002/humu.22932

Lohmann, K., & Klein, C. (2014). Next Generation Sequencing and the Future of Genetic

Diagnosis. *Neurotherapeutics*, Vol. 11, pp. 699–707. https://doi.org/10.1007/s13311-014-0288-8

Lu, H. D., Chen, G., Cai, J., & Roe, A. W. (2017). Intrinsic signal optical imaging of visual brain activity: Tracking of fast cortical dynamics. *NeuroImage*, *148*, 160–168. https://doi.org/10.1016/j.neuroimage.2017.01.006

Lumsden, A. (2004). Segmentation and compartition in the early avian hindbrain. *Mechanisms of Development*, Vol. 121, pp. 1081–1088. https://doi.org/10.1016/j.mod.2004.04.018

Lumsden, A., & Keynes, R. (1989). Segmental patterns of neuronal development in the chick hindbrain. *Nature*, *337*(6206), 424–428. https://doi.org/10.1038/337424a0

Lumsden, A., & Krumlauf, R. (1996). Patterning the vertebrate neuraxis. *Science*, Vol. 274, pp. 1109–1115. https://doi.org/10.1126/science.274.5290.1109

MacRae, C. A., & Peterson, R. T. (2015). Zebrafish as tools for drug discovery. *Nature Reviews Drug Discovery*, Vol. 14, pp. 721–731. https://doi.org/10.1038/nrd4627

Manti, F., Nardecchia, F., Barresi, S., Venditti, M., Pizzi, S., Hamdan, F. F., … Leuzzi, V. (2019). Neurotransmitter trafficking defect in a patient with clathrin (CLTC) variation presenting with intellectual disability and early-onset parkinsonism. *Parkinsonism and Related Disorders*, *61*, 207–210. https://doi.org/10.1016/j.parkreldis.2018.10.012

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

Marin, F., & Charnay, P. (2000). Hindbrain patterning: FGFs regulate Krox20 and mafB/kr expression in the otic/preotic region. *Development*, *127*(22), 4925–4935.

Martinelli, S., Torreri, P., Tinti, M., Stella, L., Bocchinfuso, G., Flex, E., … Tartaglia, M. (2008). Diverse driving forces underlie the invariant occurrence of the T42A, E139D, I282V and T468M SHP2 amino acid substitutions causing Noonan and LEOPARD syndromes. *Human Molecular Genetics*, *17*(13), 2018–2029. https://doi.org/10.1093/hmg/ddn099

McCammon, J. M., & Sive, H. (2015). Challenges in understanding psychiatric disorders and developing therapeutics: A role for zebrafish. *DMM Disease Models and Mechanisms*, *8*(7), 647–656. https://doi.org/10.1242/dmm.019620

Meena, Narendra, Praveen Mathur, Krishna Medicherla, and Prashanth Suravajhala. 2018. "A Bioinformatics Pipeline for Whole Exome Sequencing: Overview of the Processing and Steps from Raw Data to Downstream Analysis." *BIO-PROTOCOL*. https://doi.org/10.21769/bioprotoc.2805.

Metzker, M. L. (2005). Emerging technologies in DNA sequencing. *Genome Research*, Vol. 15, pp. 1767–1776. https://doi.org/10.1101/gr.3770505

Metzker, M. L. (2010). Sequencing technologies the next generation. *Nature Reviews Genetics*, Vol. 11, pp. 31–46. https://doi.org/10.1038/nrg2626

Miller-Bertoglio, V. E., Fisher, S., Sánchez, A., Mullins, M. C., & Halpern, M. E. (1997). Differential regulation of Chordin expression domains in mutant zebrafish. *Developmental Biology*, *192*(2), 537–550. https://doi.org/10.1006/dbio.1997.8788

Minoux, M., & Rijli, F. M. (2010). Molecular mechanisms of cranial neural crest cell migration and patterning in craniofacial development. *Development*, Vol. 137, pp. 2605–2621. https://doi.org/10.1242/dev.040048

Miyake, N., Fukai, R., Ohba, C., Chihara, T., Miura, M., Shimizu, H.,et al. (2016). Biallelic TBCD Mutations Cause Early-Onset Neurodegenerative Encephalopathy. *American Journal of Human Genetics*, *99*(4), 950–961. https://doi.org/10.1016/j.ajhg.2016.08.005

Miyasaka, N., Morimoto, K., Tsubokawa, T., Higashijima, S. I., Okamoto, H., & Yoshihara, Y. (2009). From the olfactory bulb to higher brain Centers: Genetic visualization of secondary olfactory pathways in zebrafish. *Journal of Neuroscience*, *29*(15), 4756–4767. https://doi.org/10.1523/JNEUROSCI.0118-09.2009

Moens, C. B., & Prince, V. E. (2002). Constructing the hindbrain: Insights from the zebrafish. *Developmental Dynamics*, Vol. 224, pp. 1–17. https://doi.org/10.1002/dvdy.10086

Moulton, M. J., & Letsou, A. (2016). Modeling congenital disease and inborn errors of development in Drosophila melanogaster. *DMM Disease Models and Mechanisms*, Vol. 9, pp. 253–269. https://doi.org/10.1242/dmm.023564

Mueller, T., & Wullimann, M. F. (2009). An evolutionary interpretation of teleostean forebrain anatomy. *Brain, Behavior and Evolution*, *74*(1), 30–42. https://doi.org/10.1159/000229011

Muto, V., Flex, E., Kupchinsky, Z., Primiano, G., Galehdari, H., Dehghani, M. et al. (2018). Biallelic SQSTM1 mutations in early-onset, variably progressive neurodegeneration. *Neurology*, *91*(4), E319–E330. https://doi.org/10.1212/WNL.0000000000005869

Myers, D. C., Sepich, D. S., & Solnica-Krezel, L. (2002). Convergence and extension in vertebrate gastrulae: Cell movements according to or in search of identity? *Trends in Genetics*, Vol. 18, pp. 447–455. https://doi.org/10.1016/S0168-9525(02)02725-7

Myklatun, A., Lauri, A., Eder, S. H. K., Cappetta, M., Shcherbakov, D., Wurst, W., et al. (2018). Zebrafish and medaka offer insights into the neurobehavioral correlates of vertebrate magnetoreception. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-03090-6

Nasevicius, A., & Ekker, S. C. (2000). Effective targeted gene "knockdown" in zebrafish. *Nature Genetics*, *26*(2), 216–220. https://doi.org/10.1038/79951

Nashabat, M., Al Qahtani, X. S., Almakdob, S., Altwaijri, W., Ba-Armah, D. M., Hundallah, K., et al. (2019). The landscape of early infantile epileptic encephalopathy in a consanguineous population. *Seizure*, *69*, 154–172. https://doi.org/10.1016/j.seizure.2019.04.018

Niceta, M., Stellacci, E., Gripp, K. W., Zampino, G., Kousi, M., Anselmi, M., et al. (2015).

Mutations impairing GSK3-mediated MAF phosphorylation cause cataract, deafness, intellectual disability, seizures, and a down syndrome-like facies. *American Journal of Human Genetics*, *96*(5), 816–825. https://doi.org/10.1016/j.ajhg.2015.03.001

Novorol, C., Burkhardt, J., Wood, K. J., Iqbal, A., Roque, C., Coutts, N.,et al. (2013). Microcephaly models in the developing zebrafish retinal neuroepithelium point to an underlying defect in metaphase progression. *Open Biology*, *3*(10), 130065. https://doi.org/10.1098/rsob.130065

Oxtoby, E., & Jowett, T. (1993). Cloning of the zebrafish krox-20 gene (krx-20) and its expression during hindbrain development. *Nucleic Acids Research*, *21*(5), 1087–1095. https://doi.org/10.1093/nar/21.5.1087

Pacheco-Rodriguez, G., Moss, J., & Vaughan, M. (2013). Arf Family. In *Encyclopedia of Biological Chemistry: Second Edition* (pp. 134–136). https://doi.org/10.1016/B978-0-12-378630-2.00354-6

Palmeirim, I., Henrique, D., Ish-Horowicz, D., & Pourquié, O. (1997). Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell*, *91*(5), 639–648. https://doi.org/10.1016/S0092-8674(00)80451-1

Papageorgiou, E., Pilat, A., Proudlock, F., Lee, H., Purohit, R., Sheth, V., et al. (2018). Retinal and optic nerve changes in microcephaly. *Neurology*, *91*(6), e571–e585. https://doi.org/10.1212/WNL.0000000000005950

Papan, C., & Campos-Ortega, J. A. (1999). Region-specific cell clones in the developing spinal cord of the zebrafish. *Development Genes and Evolution*, *209*(3), 135–144. https://doi.org/10.1007/s004270050237

Peng, Y., & Axelrod, J. D. (2012). Asymmetric protein localization in planar cell polarity: mechanisms, puzzles, and challenges. *Current Topics in Developmental Biology*, Vol. 101, pp. 33–53.

Placzek, M., & Briscoe, J. (2018). Sonic hedgehog in vertebrate neural tube development. *International Journal of Developmental Biology*, *62*(1–3), 221–230. https://doi.org/10.1387/ijdb.170293jb

Portugues, R., & Engert, F. (2009). The neural basis of visual behaviors in the larval zebrafish. *Current Opinion in Neurobiology*, Vol. 19, pp. 644–647. https://doi.org/10.1016/j.conb.2009.10.007

Pourquié, O. (1999). 3 Segmentation of the Paraxial Mesoderm and Vertebrate Somitogenesis. *Current Topics in Developmental Biology*, *47*(C), 81–105. https://doi.org/10.1016/S0070-2153(08)60722-X

Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., et al. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, *13*(1). https://doi.org/10.1186/1471-2164-13-341

Robinson, M. S. (2015). Forty Years of Clathrin-coated Vesicles. *Traffic*, Vol. 16, pp. 1210–1238. https://doi.org/10.1111/tra.12335

Rossi, G., & Messina, G. (2014). Comparative myogenesis in teleosts and mammals. *Cellular and Molecular Life Sciences*, Vol. 71, pp. 3081–3099. https://doi.org/10.1007/s00018-014-1604-5

Ruffalo, M., Laframboise, T., & Koyutürk, M. (2011). Comparative analysis of algorithms for next-generation sequencing read alignment. *Bioinformatics*, *27*(20), 2790–2796. https://doi.org/10.1093/bioinformatics/btr477

Sakai, C., Ijaz, S., & Hoffman, E. J. (2018). Zebrafish Models of Neurodevelopmental Disorders: Past, Present, and Future. *Frontiers in Molecular Neuroscience*, Vol. 11. https://doi.org/10.3389/fnmol.2018.00294

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Sasai, Y. (1998). Identifying the missing links: Genes that connect neural induction and primary neurogenesis in vertebrate embryos. *Neuron*, Vol. 21, pp. 455–458. https://doi.org/10.1016/S0896-6273(00)80554-1

Sasai, Y., & De Robertis, E. M. (1997). Ectodermal patterning in vertebrate embryos. *Developmental Biology*, Vol. 182, pp. 5–20. https://doi.org/10.1006/dbio.1996.8445

Saude, L., Wooley, K., Martin, P., Driever, W., & Stemple, D. L. (2000). Axis-inducing activities and cell fates of the zebrafish organizer. *Development*, *127*(16), 3407–3417.

Schartl, M. (2014). Beyond the zebrafish: Diverse fish species for modeling human disease. *DMM Disease Models and Mechanisms*, Vol. 7, pp. 181–192. https://doi.org/10.1242/dmm.012245

Scheffer, I. E., Berkovic, S., Capovilla, G., Connolly, M. B., French, J., Guilhoto, L., et al. (2018). ILAE classification of the epilepsies: position paper of the ILAE Commission for Classification and Terminology. *Zeitschrift Fur Epileptologie*, Vol. 31, pp. 296–306. https://doi.org/10.1007/s10309-018-0218-6

Schilling, T. F. (2008). Anterior-posterior patterning and segmentation of the vertebrate head. *Integrative and Comparative Biology*, Vol. 48, pp. 658–667. https://doi.org/10.1093/icb/icn081

Schmidt, R., Strähle, U., & Scholpp, S. (2013). Neurogenesis in zebrafish - from embryo to adult. *Neural Development*, Vol. 8. https://doi.org/10.1186/1749-8104-8-3

Seifert, J. R. K., & Mlodzik, M. (2007). Frizzled/PCP signalling: A conserved mechanism regulating cell polarity and directed motility. *Nature Reviews Genetics*, Vol. 8, pp. 126–138. https://doi.org/10.1038/nrg2042

Seixas, E., Barros, M., Seabra, M. C., & Barral, D. C. (2013). Rab and Arf proteins in genetic diseases. *Traffic*, Vol. 14, pp. 871–885. https://doi.org/10.1111/tra.12072

Sepich, D. S., Calmelet, C., Kiskowski, M., & Solnica-Krezel, L. (2005). Initiation of convergence and extension movements of lateral mesoderm during zebrafish gastrulation. *Developmental Dynamics*, *234*(2), 279–292. https://doi.org/10.1002/dvdy.20507

Sepúlveda, F. V., Cid, L. P., Teulon, J., & Niemeyer, M. I. (2015). Molecular aspects of structure, gating, and physiology of pH-sensitive background K2P and Kir K+-transport channels. *Physiological Reviews*, *95*(1), 179–217. https://doi.org/10.1152/physrev.00016.2014

Shieh, C. C., Coghlan, M., Sullivan, J. P., & Gopalakrishnan, M. (2000). Potassium channels: Molecular defects, diseases, and therapeutic opportunities. *Pharmacological Reviews*, Vol. 52, pp. 557–593.

Silverman, G. A., Luke, C. J., Bhatia, S. R., Long, O. S., Vetica, A. C., Perlmutter, D. H., & Pak, S. C. (2009). Modeling molecular and cellular aspects of human disease using the nematode Caenorhabditis elegans. *Pediatric Research*, Vol. 65, pp. 10–18. https://doi.org/10.1203/PDR.0b013e31819009b0

Simons, Cas, Lachlan D. Rash, Joanna Crawford, Linlin Ma, Ben Cristofori-Armstrong, David Miller, Kelin Ru, et al. 2015. "Mutations in the Voltage-Gated Potassium Channel Gene KCNH1 Cause Temple-Baraitser Syndrome and Epilepsy." *Nature Genetics*. https://doi.org/10.1038/ng.3153.

Singh, J., & Mlodzik, M. (2012). Planar cell polarity signaling: Coordination of cellular orientation across tissues. *Wiley Interdisciplinary Reviews: Developmental Biology*, Vol. 1, pp. 479–499. https://doi.org/10.1002/wdev.32

Sive, H. L., Draper, B. W., Harland, R. M., & Weintraub, H. (1990). Identification of a retinoic acid-sensitive period during primary axis formation in Xenopus laevis. *Genes and Development*, *4*(6), 932–942. https://doi.org/10.1101/gad.4.6.932

Snider, D. R., & Clegg, E. D. (1975). Alteration of phospholipids in porcine spermatozoa during in vivo uterus and oviduct incubation. *Journal of Animal Science*, *40*(2), 269–274. https://doi.org/10.2527/jas1975.402269x

Spemann, H., & Mangold, H. (2001). Induction of embryonic primordia by implantation of organizers from a different species. *International Journal of Developmental Biology*, *45*(1 SPEC. ISS. 1), 13–38. https://doi.org/10.1387/ijdb.11291841

Spence, R., Gerlach, G., Lawrence, C., & Smith, C. (2008). The behaviour and ecology of the zebrafish, Danio rerio. *Biological Reviews*, Vol. 83, pp. 13–34. https://doi.org/10.1111/j.1469-185X.2007.00030.x

Stenmark, H. (2009). Rab GTPases as coordinators of vesicle traffic. *Nature Reviews Molecular Cell Biology*, Vol. 10, pp. 513–525. https://doi.org/10.1038/nrm2728

Streit, A., Berliner, A. J., Papanayotou, C., Slrulnik, A., & Stern, C. D. (2000). Initiation of neural induction by FGF signalling before gastrulation. *Nature*, *406*(6791), 74–78. https://doi.org/10.1038/35017617

Tang, Ke, Guangdun Peng, Yunbo Qiao, Lu Song, and Naihe Jing. 2015. "Intrinsic Regulations in Neural Fate Commitment." *Development Growth and Differentiation*. https://doi.org/10.1111/dgd.12204.

Tartaglia, M., Mehler, E. L., Goldberg, R., Zampino, G., Brunner, H. G., Kremer, H., et al. (2001). Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2, cause Noonan syndrome. *Nature Genetics*, *29*(4), 465–468. https://doi.org/10.1038/ng772

Tartaglia, M., Niemeyer, C. M., Fragale, A., Song, X., Buechner, J., Jung, A. et al. (2003). Somatic mutations in PTPN11 in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. *Nature Genetics*, *34*(2), 148–150. https://doi.org/10.1038/ng1156

Tawk, M., Araya, C., Lyons, D. A., Reugels, A. M., Girdler, G. C., Bayley, P. R. et al. (2007). A mirror-symmetric cell division that orchestrates neuroepithelial morphogenesis. *Nature*, *446*(7137), 797–800. https://doi.org/10.1038/nature05722

Teame, Tsegay, Zhen Zhang, Chao Ran, Hongling Zhang, Yalin Yang, Qianwen Ding, Minxu Xie, et al. 2019. "The Use of Zebrafish (Danio Rerio) as Biomedical Models." *Animal Frontiers*. https://doi.org/10.1093/af/vfz020.

Terlau, H., & Kirchhoff, F. (2006). Ion Channels/Excitable Membranes. In *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine* (pp. 913–916). https://doi.org/10.1007/3-540-29623-9_5640

Tian, C., Zhu, R., Zhu, L., Qiu, T., Cao, Z., & Kang, T. (2014). Potassium channels: Structures, diseases, and modulators. *Chemical Biology and Drug Design*, *83*(1), 1–26. https://doi.org/10.1111/cbdd.12237

Torkamani, A., Scott-Van Zeeland, A. A., Topol, E. J., & Schork, N. J. (2011). Annotating individual human genomes. *Genomics*, Vol. 98, pp. 233–241. https://doi.org/10.1016/j.ygeno.2011.07.006

Tsai, S. C., Haun, R. S., Tsuchiya, M., Moss, J., & Vaughan, M. (1991). Isolation and characterization of the human gene for ADP-ribosylation factor 3, a 20-kDa guanine nucleotide-binding protein activator of cholera toxin. *Journal of Biological Chemistry*, *266*(34), 23053–23059.

Tuchman, R., & Cuccaro, M. (2011). Epilepsy and autism: Neurodevelopmental perspective. *Current Neurology and Neuroscience Reports*, *11*(4), 428–434. https://doi.org/10.1007/s11910-011-0195-x

Urrego, D., Tomczak, A. P., Zahed, F., Stühmer, W., & Pardo, L. A. (2014). Potassium channels in cell cycle and cell proliferation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, Vol. 369. https://doi.org/10.1098/rstb.2013.0094

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., et al. (2013). From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, (SUPL.43).

https://doi.org/10.1002/0471250953.bi1110s43

Varga, M., Ralbovszki, D., Balogh, E., Hamar, R., Keszthelyi, M., & Tory, K. (2018). Zebrafish Models of Rare Hereditary Pediatric Diseases. *Diseases*, *6*(2), 43. https://doi.org/10.3390/diseases6020043

Vaz, R., Hofmeister, W., & Lindstrand, A. (2019). Zebrafish models of neurodevelopmental disorders: Limitations and benefits of current tools and techniques. *International Journal of Molecular Sciences*, Vol. 20. https://doi.org/10.3390/ijms20061296

Veldman, M. B., & Lin, S. (2008). Zebrafish as a developmental model organism for pediatric research. *Pediatric Research*, Vol. 64, pp. 470–476. https://doi.org/10.1203/PDR.0b013e318186e609

Wada, H., & Okamoto, H. (2009). Roles of noncanonical Wnt/PCP pathway genes in neuronal migration and neurulation in zebrafish. *Zebrafish*, *6*(1), 3–8. https://doi.org/10.1089/zeb.2008.0557

Wallingford, J. B., Fraser, S. E., & Harland, R. M. (2002). Convergent extension: The molecular control of polarized cell movement during embryonic development. *Developmental Cell*, Vol. 2, pp. 695–706. https://doi.org/10.1016/S1534-5807(02)00197-1

Wang, Y., & Nathans, J. (2007). Tissue/planar cell polarity in vertebrates: New insights and new questions. *Development*, Vol. 134, pp. 647–658. https://doi.org/10.1242/dev.02772

Warburg, M. (1976). Heterogeneity of congenital retinal non attachment, falciform folds and retinal dysplasia. A guide to genetic counselling. *Human Heredity*, *26*(2), 137–148. https://doi.org/10.1159/000152795

Warburg, M., & Heuer, H. E. (1994). Chorioretinal dysplasia-microcephaly-mental retardation syndrome [2]. *American Journal of Medical Genetics*, Vol. 52, p. 117. https://doi.org/10.1002/ajmg.1320520124

Weinberg, E. S., Allende, M. L., Kelly, C. S., Abdelhamid, A., Murakami, T., Andermann, P., et al. (1996). Developmental regulation of zebrafish MyoD in wild-type, no tail and spadetail embryos. *Development*, *122*(1), 271–280.

Weller, J., Steinhäuser, C., & Seifert, G. (2016). PH-Sensitive K+ Currents and Properties of K2P Channels in Murine Hippocampal Astrocytes. *Advances in Protein Chemistry and Structural Biology*, *103*, 263–294. https://doi.org/10.1016/bs.apcsb.2015.10.005

Wessely, O., Agius, E., Oelgeschläger, M., Pera, E. M., & De Robertis, E. M. (2001). Neural induction in the absence of mesoderm: β-Catenin-dependent expression of secreted BMP antagonists at the blastula stage in Xenopus. *Developmental Biology*, *234*(1), 161–173. https://doi.org/10.1006/dbio.2001.0258

White, R. J., & Schilling, T. F. (2008). How degrading: Cyp26s in hindbrain development. *Developmental Dynamics*, Vol. 237, pp. 2775–2790. https://doi.org/10.1002/dvdy.21695

Wilson, P. A., Lagna, G., Suzuki, A., & Hemmati-Brivanlou, A. (1997). Concentration-dependent patterning of the Xenopus ectoderm by BMP4 and its signal transducer

Smad1. *Development*, *124*(16), 3177–3184.

Wilson, S. I., Rydström, A., Trimborn, T., Willert, K., Musse, R., Jessell, T. M., & Edlund, T. (2001). The status of Wnt signalling regulates neural and epidermal fates in the chick embryo. *Nature*, *411*(6835), 325–330. https://doi.org/10.1038/35077115

Wolf, B., Lesnaw, J. A., & Reichmann, M. E. (1970). A Mechanism of the Irreversible Inactivation of Bovine Pancreatic Ribonuclease by Diethylpyrocarbonate: A General Reaction of Diethylpyrocarbonate with Proteins. *European Journal of Biochemistry*, *13*(3), 519–525. https://doi.org/10.1111/j.1432-1033.1970.tb00955.x

Woo, K., & Fraser, S. E. (1995). Order and coherence in the fate map of the zebrafish nervous system. *Development*, *121*(8), 2595–2609.

Wright, C. F., FitzPatrick, D. R., & Firth, H. V. (2018). Paediatric genomics: Diagnosing rare disease in children. *Nature Reviews Genetics*, Vol. 19, pp. 253–268. https://doi.org/10.1038/nrg.2017.116

Y., L. C. (2012). EURORDIS-The European organisation for rare diseases. *Haemophilia*, *18*, 6. https://doi.org/10.1111/j.1365-2516.2012.02820.x

Yin, Chen, Li, X., & Du, J. (2019). Optic tectal superficial interneurons detect motion in larval zebrafish. *Protein and Cell*, *10*(4), 238–248. https://doi.org/10.1007/s13238-018-0587-7

Yin, Chunyue, Kiskowski, M., Pouille, P. A., Farge, E., & Solnica-Krezel, L. (2008). Cooperation of polarized cell intercalations drives convergence and extension of presomitic mesoderm during zebrafish gastrulation. *Journal of Cell Biology*, *180*(1), 221–232. https://doi.org/10.1083/jcb.200704150

Young, I. D., Fielder, A. R., & Simpson, K. (1987). Microcephaly, microphthalmos, and retinal folds: Report of a family. *Journal of Medical Genetics*, *24*(3), 172–174. https://doi.org/10.1136/jmg.24.3.172

Zalik, S. E., Lewandowski, E., Kam, Z., & Geiger, B. (1999). Cell adhesion and the actin cytoskeleton of the enveloping layer in the zebrafish embryo during epiboly. *Biochemistry and Cell Biology*, *77*(6), 527–542. https://doi.org/10.1139/o99-058

Zhang, Yibo, Zhang, Z., & Ge, W. (2018). An efficient platform for generating somatic point mutations with germline transmission in the zebrafish by CRISPR/Cas9-mediated gene editing. *Journal of Biological Chemistry*, *293*(17), 6611–6622. https://doi.org/10.1074/jbc.RA117.001080

Zhang, Yihan, Huang, H., Zhao, G., Yokoyama, T., Vega, H., Huang, Y., et al. (2017). ATP6V1H Deficiency Impairs Bone Development through Activation of MMP9 and MMP13. *PLoS Genetics*, *13*(2). https://doi.org/10.1371/journal.pgen.1006481