



**HAL**  
open science

# Exploration of interaction between common and rare variants in genetic susceptibility to holoprosencephaly

Artem Kim

► **To cite this version:**

Artem Kim. Exploration of interaction between common and rare variants in genetic susceptibility to holoprosencephaly. Human health and pathology. Université de Rennes, 2020. English. NNT : 2020REN1B019 . tel-03199534

**HAL Id: tel-03199534**

**<https://theses.hal.science/tel-03199534>**

Submitted on 15 Apr 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THESE DE DOCTORAT DE

L'UNIVERSITE DE RENNES 1

ECOLE DOCTORALE N° 605

*Biologie Santé*

Spécialité : Génétique, génomique et bio-informatique

Par Artem KIM

## **Exploration de l'interaction entre variants rares et communs dans la susceptibilité génétique à holoprosencéphalie**

Thèse présentée et soutenue à Rennes, le 30/09/2020

Unité de recherche : Institut de Génétique et Développement de Rennes (IGDR), UMR6290 CNRS, Université Rennes 1

### **Rapporteurs avant soutenance :**

**Thomas BOURGERON**

Professeur des universités  
Université de Paris, CNRS UMR3571

**Gaël Nicolas**

Maître de conférences des universités  
Université de Rouen, Inserm U1079

### **Composition du Jury :**

**Thomas BOURGERON**

Professeur des universités, Université de Paris, CNRS UMR3571

**Gaël Nicolas**

Maître de conférences des universités, Université de Rouen, Inserm U1079

**Sylvain Lehmann**

Professeur des universités, Université de Montpellier, CNRS UMR1051

Président

**Sandrine Lagarrigue**

Professeur des universités, Agrocampus Ouest, UMR13448

Directeur de thèse

**Marie de TAYRAC**

Maître de conférences des universités, Université Rennes 1, CNRS UMR6290

Co-directeur de thèse

**Véronique David**

Professeur des universités, Université Rennes 1, CNRS UMR6290

# SUMMARY

<b>INTRODUCTION</b> .....	<b>1</b>
<b>I. Mendelian and complex genetic disorders</b> .....	<b>2</b>
a. The human genome .....	2
b. Genetic disorders and mutations .....	3
c. Mendelian disease era and single-gene disorders .....	5
d. Complex non-Mendelian disorders .....	7
e. Towards the end of the Mendelian disease era .....	11
<b>II. Holoprosencephaly</b> .....	<b>13</b>
a. Definition.....	13
b. Clinical spectrum .....	13
c. Etiology.....	17
d. Diagnostic and molecular testing .....	23
<b>III. Molecular and genetic basis of HPE</b> .....	<b>25</b>
a. Forebrain and facial development.....	25
b. Molecular pathways of HPE .....	32
c. Genetic mechanisms of HPE .....	43
d. Non-mendelian HPE - Article I.....	50
<b>IV. Next Generation Sequencing in medical genomics</b> .....	<b>51</b>
a. Main principles .....	52
b. Applications.....	53
c. DNA-Seq and identification of disease-causing variants.....	55
d. Novel strategies and future challenges for clinical variant interpretation.....	62
<b>RESULTS</b> .....	<b>70</b>
<b>I. Implication of hypomorphic variants and oligogenic inheritance in HPE.</b> .....	<b>70</b>
a. Background and objectives .....	70
b. Methods .....	70
c. Results - Article II .....	78
<b>II. Pathogenic impact of synonymous variants in <i>SHH</i> gene</b> .....	<b>82</b>
a. Background.....	82
b. Strategy .....	87
c. Results - Article III .....	90
<b>III. Autosomal recessive holoprosencephaly in consanguineous families</b> .....	<b>94</b>
a. Background.....	94
b. Strategy .....	94
c. Results .....	100
<b>DISCUSSION</b> .....	<b>108</b>

<b>I. Oligogenic inheritance in complex genetic disorders.....</b>	<b>108</b>
<b>II. Novel strategies for investigating the clinical impact of hypomorphic variants .....</b>	<b>110</b>
<b>III. Autosomal recessive inheritance in HPE .....</b>	<b>113</b>
<b>IV. Perspectives .....</b>	<b>113</b>
<b>Conclusion.....</b>	<b>117</b>
<b>REFERENCES.....</b>	<b>120</b>
<b>ACKNOWLEDGEMENTS AND PERSONAL NOTES .....</b>	<b>144</b>

# ***INTRODUCTION***

# I. Mendelian and complex genetic disorders

## a. The human genome

The haploid human genome is composed of roughly 3 billion pairs of nucleotides. While a small amount of genetic information is found in the mitochondria (mitochondrial genome), the major part of human DNA is condensed within the nucleus as 23 pairs of chromosomes, one inherited from each parent. The nuclear genome encodes approximately 30,000 protein-coding genes consisting of coding (exonic) and non-coding (intronic) regions. The exonic regions are spliced and translated into ~100,000 different proteins containing ~6,000,000 amino acids, but compose less than 3 % of the human genome (Bohlander, 2013). The function is much less understood for the remaining, untranslated part of the genome. Sometimes referred to as the genomic 'dark matter', the non-coding part of the genome has been shown to play key roles in coordinating critical biological processes (Blaxter, 2010). More than 70% of the human genome is transcribed into RNA that does not encode proteins (ncRNA) but is involved in various regulatory processes of gene expression including splicing (lncRNA), post-transcriptional regulation (miRNA) and translational machinery (rRNA, tRNA) among others (Carcini et al., 2005; Romero-Barios et al., 2018). The non-coding genome also contains over 14,000 pseudogenes - copies of functional genes with coding-sequence alterations. Although considered as evolutionary relics, some pseudogenes have been shown to regulate their protein-coding counterparts (Jackson et al., 2018). In 2007, the authors of the Encyclopedia of DNA Elements (ENCODE) project claimed that a "biochemical function" could be assigned to 80% of the human genome (ENCODE Project Consortium, 2007; 2013). This statement was heavily debated afterwards, based on a loose definition of "function" by the authors and by invoking the so-called "Onion test": briefly, the genome size of the onion plant, *Allium cepa*, is approximately five times larger than the humans (Palazzo and Gregory, 2014). If most eukaryotic DNA is functional at the organism level, then why does an onion require five times more of

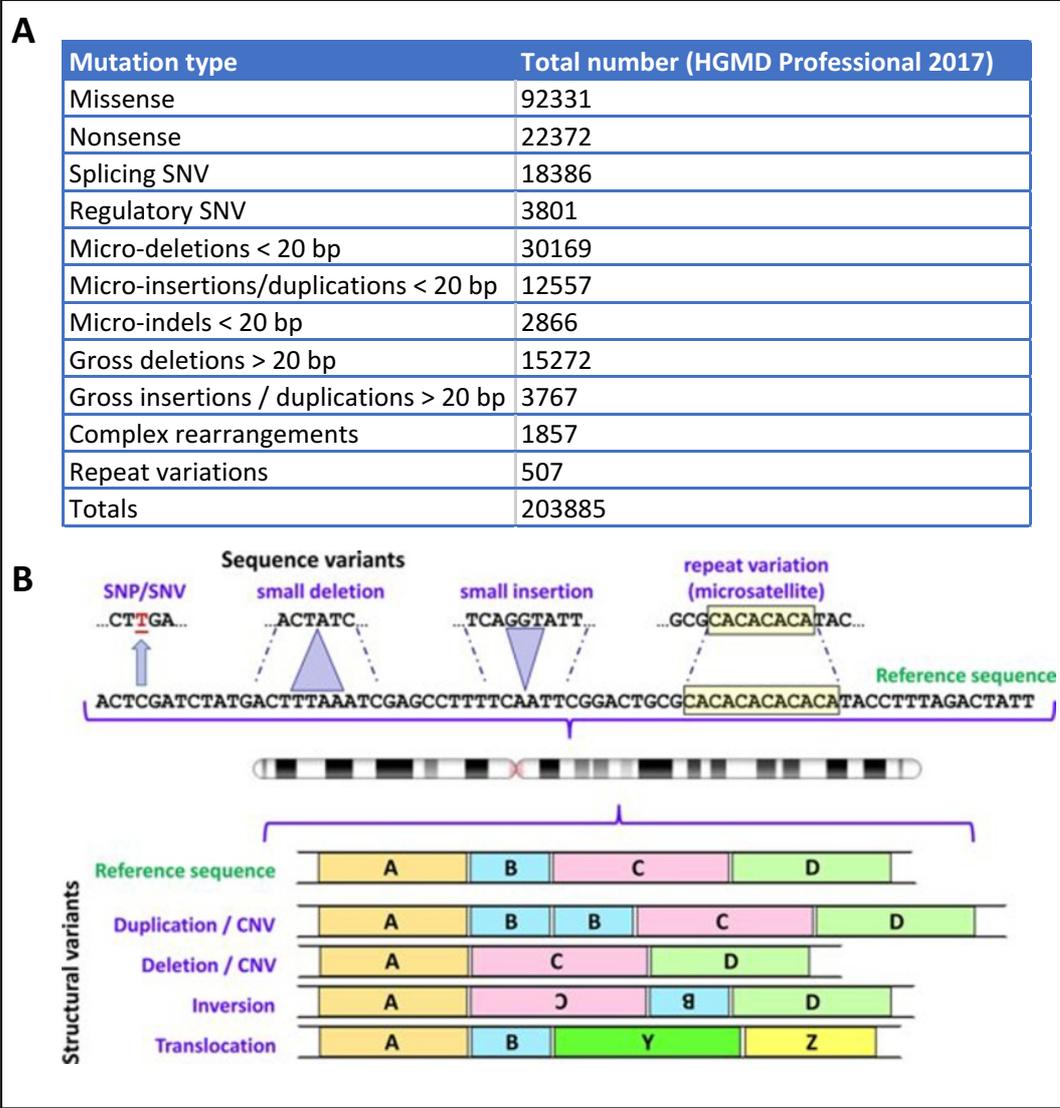
it than a human? Thus, the question of what proportion of the human genome is functional remains heavily debated.

## **b. Genetic disorders and mutations**

Genetic disorders are recognized to be one of the major categories of human diseases and result from cellular dysfunction caused by one or more alterations of the genome (mutations). If disease-causing mutations occur in the individual's germ cells (constitutional/germline mutations), genetic diseases are hereditary and can be transmitted to descendants. Many genetic disorders can also occur sporadically, caused by a *de novo* mutation in one of the parental gametes, resulting in a genetic abnormality present in the affected child but absent in either parent's genome.

A classic definition of mutation is 'any heritable change to the DNA sequence'. Although such changes may be disease-causing, many alterations of the genome have no consequences on the individual's phenotypes. Consequently, in modern medical genetics, disease-causing alterations of the genome are called 'mutations', while alterations with neutral/unknown effect are referred to as 'variants' (*i.e.*, variations from the sequence of the reference genome) (Karki et al., 2015). Variants can be further classified as benign/polymorphic if they do not cause a disease phenotype, or as pathogenic/disease-causing when presenting sufficient evidence for their disease implication (Richards et al., 2015). In case of unknown effect, variants are classified as 'variants of uncertain significance' (VUS). Genetic variants are now known to result in more than 6,000 genetic diseases, many of which are fatal or causing extremely severe conditions (Hamosh et al., 2002). Over the past decade, novel technologies of DNA analysis have led to remarkable progress in identification of genetic factors underlying human disorders (Durmaz et al., 2015). Numerous efforts were made to collate all known genetic anomalies identified in clinical context. For example, the Human Gene Mutation Database (HGMD), established in 1996, represents one of the largest databases of inherited disease-related genetic variants (Figure 1A). As of 2017, HGMD contained more than 203,000 variants identified in over 8,000 genes manually curated

from over 2,600 peer-reviewed scientific journals (Cooper et al. 2010; Stenson et al. 2017).



**Figure 1. (A) Summary statistics of the HGMD database (2017 release).** Data is retrieved from Stenson et al., 2017. **(B) Types of variants found in human genomes.** For depiction of structural variants, A, B, C and D correspond to large segments of DNA; Y and Z represent segments of DNA originating from a different chromosome. Jackson et al., 2018.

Single Nucleotide Variants (SNVs), historically termed Single Nucleotide Polymorphisms (SNPs), correspond to a substitution of a single base pair and represent the most common type of genetic variation in the human genome, occurring

approximately once every 100-300 bases (Sherry et al., 2001). SNVs also represent the majority of disease-related variants recorded in the HGMD (67%). Of these, missense variants resulting in substitution of a single amino acid are the most common, followed by nonsense variants resulting in a premature stop codon producing a truncated, usually non-functional protein. Substitutions in splicing and regulatory regions account for 10 % of the HGMD database. Insertions and deletions of short genomic sequences (< 1000 bp), commonly referred to as Indels, constitute the second most abundant class of genetic variation (The 1000 Genomes Project Consortium, 2015). Structural variants correspond to large (> 1000 bp) deletions/duplications (copy number variants or CNVs) and complex rearrangements, such as inversions and translocations (Figure 1B).

For a long time, studies of genetic diseases have distinguished disorders of Mendelian and complex inheritance (Badano and Katsanis, 2002). Mendelian disorders are considered rare and have predictable inheritance patterns as they usually result from a single highly-penetrant pathogenic mutation in a disease-associated gene. In contrast, complex disorders are considered common in the general population and result from interaction of multiple genetic and environmental factors. Although the existence of functional links between Mendelian and complex diseases has now been illustrated, this simplified dichotomy was useful for elucidating the genetic basis of human disorders in the past.

### **c. Mendelian disease era and single-gene disorders**

In 1865, Gregor Mendel laid the foundation for our modern comprehension of genetic disorders (Weiling, 1991). In his famous work on the inheritance of different features in pea plants, Mendel discovered the fundamental laws of genotype-phenotype relationships by tracking the segregation of parental genes and their appearance in the offspring as dominant and recessive traits. Since this milestone discovery, the emphasis in human genetics has long been focused on studies of single gene disorders

inherited in a dominant or recessive manner - the period termed as the Mendelian disease era (Katsanis, 2016). Such studies have led to important insights into the genetic basis underlying phenotypic variation in human.

Mendelian disorders, also called monogenic, are the most well-characterized type of genetic disease (Kennedy, 2005). As their name suggests, the disease phenotype in such disorders results from a mutation at a single genetic locus and its inheritance follows a recognizable Mendelian pattern. Several types of Mendelian inheritance patterns can be distinguished (Table 1), depending on the location of the responsible mutation (autosomes or sex chromosomes) and whether the disease results from alterations of one or two gene copies (heterozygous / homozygous mutations).

Inheritance	Characteristics	Disease Examples
Autosomal Dominant	Disease is caused by an autosomal heterozygous mutation, which can be inherited from an affected parent or occur <i>de novo</i> .	Huntington's disease, neurofibromatosis
Autosomal Recessive	Disease is caused by an autosomal homozygous mutation. Both parents of an affected person present the mutation at heterozygous state (healthy carriers).	Tay-Sachs disease, sickle cell anemia, cystic fibrosis,
X-linked Dominant	The disease-causing mutation is present at heterozygous state on the X chromosome. Males generally present more severe symptoms. No male-to-male transmission.	Fragile X syndrome, ornithine transcarbamylase deficiency
X-linked Recessive	For males, one altered gene copy is sufficient to cause the disease. For females, both alleles need to be altered for disease manifestation. Males are more frequently affected; affected males present in each generation.	Hemophilia A, Duchenne muscular dystrophy
Mitochondrial	The disease is caused by a mutation in the mitochondrial genome. Maternal transmission. Can appear in every generation.	Leber's hereditary optic neuropathy, Kearns-Sayre syndrome

**Table 1. Common types of Mendelian inheritance.** Adapted from Jackson et al., 2018

Initial attempts to identify genes underlying Mendelian disorders primarily relied on studies of multigenerational affected families (Botstein and Risch, 2003). Specifically, linkage mapping techniques are applied to family genotyping data in order to identify particular SNP markers that are present at identical state in the affected family members and transmitted from a common ancestor. Such markers are also called identical by descent (IBD). Genomic regions located between two IBD markers in affected individuals are considered as candidate regions segregating with the disease. The candidate regions are subsequently sequenced and analyzed to identify the gene involved in the disease pathogenesis. Linkage mapping has proven extremely successful in discovering genes and variants underlying about one-half to one-third (~3,000) Mendelian traits (McKusik, 2007; Kaiser, 2010). However, several factors can limit the power of this approach: the disease rarity and a small number of available families, incomplete penetrance (*i.e.*, not all carriers of the disease-causing mutation exhibit expected phenotype) and locus heterogeneity (*i.e.*, mutations in different genes can lead to the same disease phenotype) (Anthonarakis and Beckman, 2006).

Since 2005, the advent of Next Generation Sequencing (NGS), described in part IV of the Introduction, has greatly contributed towards identifying novel genetic factors underlying Mendelian disorders. Novel studies rapidly demonstrated the utility of NGS in clinical context, as it became possible to identify pathogenic mutations cost- and time-efficiently, using only the sequencing data of several unrelated patients or a single affected family (Gilissen et al., 2010; Hoischen et al., 2010; Lalonde et al., 2010; Ng et al., 2010a, 2010b). NGS has rapidly become a powerful approach to identify disease-causing genes in cases where the traditional linkage mapping has failed, and is now routinely used in clinical setting for the diagnosis of Mendelian disorders (Bamshad et al., 2011).

#### **d. Complex non-Mendelian disorders**

Although studies of Mendelian inheritance have contributed greatly to our understanding of genetic mechanisms underlying human disease, many disorders present familial recurrence and a clear genetic component without following a

recognizable Mendelian pattern. Such non-Mendelian disorders are called complex or multifactorial (van Heyningen and Yeyati, 2004). Common examples include neurodevelopmental disorders such as Autism Spectrum Disorder (ASD), bipolar disorder and epilepsy; autoimmune diseases; type II diabetes; and congenital malformations including Neural Tube Defects (NTDs) and ciliopathies. The genetic basis of such disorders remains largely unknown and the molecular diagnosis is challenging due to absence of clear genotype-phenotype correlation.

### *GWAS*

Unlike Mendelian disorders, studies of complex traits relied on population-based approaches rather than family-based investigations (Baron, 2001). In particular, genome wide association studies (GWAS) apply a case-control approach to identify particular SNP genotypes (alleles) present at significantly higher frequencies in patients (cases) as compared to unaffected individuals (controls). Such markers significantly associated with the disease (also called risk variants or risk alleles) are believed to either directly participate in disease pathogenesis or to be linked to disease-causing variants via linkage disequilibrium (LD) - the nonrandom association between different alleles driven by evolutionary forces such as genetic drift, natural selection and recombination events (Slatkin, 2008). Over the years, GWAS studies have identified more than 1,200 risk alleles for complex traits, but their precise contributions to the disease phenotype or links to the disease-causing variants remain mostly unclear (Richards et al., 2015). Although GWAS have brought important insights into genes and pathways underlying complex disorders, the identification of precise disease-causing factors remains challenging in clinical context (Visscher et al., 2012). Studies of complex diseases often lead to a long-term, and often unsuccessful, search for a causative factor, also known as the 'diagnostic odyssey' (Baynam, 2016; Lappe et al., 2018). In order to resolve these enigmatic cases, unconventional patterns of inheritance involving multiple genetic and environmental factors are considered.

### *Oligogenic/Polygenic inheritance*

As all genetic variants are transmitted *de facto* in a Mendelian manner, the distinction between Mendelian and complex inheritance can sometimes appear elusive. While Mendelian disorders are caused by an alteration of a single gene, complex disorders can require a combined input of multiple genetic variants affecting several genes at once (Kousi and Katsanis, 2015). Such inheritance pattern, also termed oligogenic or polygenic, has been illustrated in different genetic pathologies. Various cases of Hirschsprung disease (HSCR) involve alterations of at least three genes required for the disease pathogenesis (Gabriel et al., 2002; Brooks et al., 2004; Borrego et al., 2013). Evidence for oligogenic inheritance has been reported in Parkinson's disease (Lubbe et al., 2016). In developmental disorders, ciliopathies are characterized by complex relationships between numerous disease genes and are believed to involve oligogenic inheritance (Davis and Katsanis, 2012). The earliest example of such patterns in ciliopathies is the report of triallelic inheritance (*i.e.*, involving variants in 3 different genes) in Bardet-Biedl syndrome (BBS), initially considered as a classical autosomal recessive disorder (Katsanis, 2001).

### *Implication of common variants*

While Mendelian pathologies are caused by a rare mutation with a high deleterious impact on cell function, the etiology of complex disorders can involve genetic variants that are hypomorphic, *i.e.*, of a relatively low deleterious effect (Shastri and Shmidt, 2016; Nava et al., 2015). Assuming that one hypomorphic variant is not sufficient to induce the non-Mendelian disease phenotype, they can be relatively common in the general population. For example, a variant in *RET* conferring susceptibility to HSCR has a minor allele frequency (MAF) of 45% and 25% in Asian and European populations, respectively (Emison et al., 2005). In age-related macular degeneration, common variants in *CFH* have been associated with an increased disease risk (Maller et al., 2006; Hageman et al., 2005). In Crohn's disease, a common variant in *IRGM* has been shown to alter gene expression and participate in the disease pathogenesis (Brest et al., 2011). The idea that both common variants of small effect and rare variants of large effect

contribute to the genetic architecture of complex disorders is now largely accepted (Gibson, 2012), thus further complicating the distinction between disease-related and truly benign variants. Unmasking common variants involved in disease pathogenesis represents a major challenge in modern medical genetics.

#### *Environmental factors*

Furthermore, the genetic risk of a complex disorder can be modulated by environmental factors. For example, smoking is considered as a strong risk factor for Crohn's disease, while malnutrition during pregnancy has been shown to increase the risk for the child to be affected by mental retardation (Aldhous and Satsangi, 2010; Raina et al., 2016). Some environmental factors remain difficult to interpret, such as birth seasonality in psychiatric disorders: cases of bipolar disorder and schizophrenia have been shown to preferentially occur in winter-spring births, but the underlying cause of that association remains elusive (Torrey et al., 1997).

#### *Modifier effect*

Polygenic/oligogenic inheritance observed in complex disorders has often been referred to as 'modifier effect', however recent studies begin to distinguish these two concepts (Kousi and Katsanis, 2015). While oligogenic inheritance refers to a combination of several variants which are strictly necessary for the disease manifestation, modifier effect corresponds to the presence of a 'secondary' mutation which modulates the impact of the 'main' pathogenic variant, sufficient itself to induce the disease (Versbraegen et al., 2019). Although genetic modifiers are not disease-causing themselves, they can influence the phenotypic output of another pathogenic variant, either by modifying the expression of the affected gene or by altering the function of the resulting protein (Nadeau, 2001; Slavotinek and Biesecker, 2003). The modifier variant can be located in the same gene as the disease-causing variant. In Erythropoietic protoporphyria, a common polymorphic variant in *FECH* modulates the effect of a rare pathogenic mutation located in the same gene (Gouya et al., 2002). Modifier variants can also be present at a distinct location. For example, analysis of 228 patients affected by BBS, has revealed that a variant in *CCDC28B* modulates the

effect of a recurrent pathogenic mutation in *BBS1* and is associated with more severe clinical phenotypes (Badano et al., 2006). Similarly, a common microsatellite in *D4Z4* modulates the penetrance of pathogenic *SMCHD1* variants in facioscapulohumeral muscular dystrophy (Lemmers et al., 2012). Modifier effects have also been extensively studied in mouse models. For example, while knockout of either *Pax3* or *Grhl3* (Curly tail) results in NTDs, the simultaneous deficiency for both genes causes more severe forms than those found in single knockout mutants, indicating that interaction between these two genes influences the resulting phenotype (Estibeiro et al., 1993).

### *Clinical variability*

The overall complex etiology of non-Mendelian disorders contributes to a wide phenotypic variability observed in affected individuals. In neurodevelopmental disorders, ASD is associated with a wide spectrum of mental and behavioral abnormalities ranging from mild to severe learning disability, various deficits in communication skills and manifestations of schizophrenia and/or epilepsy (Wozniak et al., 2017). In NTDs, clinical manifestations can vary from spina bifida (failed closure of the neural tube at the brain) and exencephaly (failed closure at the spinal cord) to more severe forms of craniorachischisis (failed closure at both the brain and the spinal regions) (Greene and Copp, 2014).

### **e. Towards the end of the Mendelian disease era**

Since Mendel's discoveries, the field of medical genetics had several fundamental concepts of genetic causality, which are now being challenged (Katsanis, 2016). For much of the Mendelian disease era, the reported disease-causing mutations have been described as both necessary and sufficient for disease pathogenesis. We are now aware that many 'disease-causing mutations' do not necessarily cause the disease phenotype in a certain proportion of its carriers (*i.e.*, exhibit reduced/incomplete penetrance), and that tolerance to pathogenic mutations can vary among individuals (the concept termed as 'genome resilience') (Chen et al., 2016). Reduced penetrance of pathogenic variants has been illustrated for various disorders normally characterized by autosomal dominant (congenital cataract, *GJA3*, Burdon et al., 2004;

retinis pigmentosa, *PRPF8*, Maubaret et al., 2011) or autosomal recessive (von Willebrand disease, *VWF*, Castaman et al., 2007; hemochromatosis, *HFE*, Beutler, 2003) inheritance. Although reduced penetrance of pathogenic mutations may be explained by non-genetic factors (age of onset, sex, epigenetic and environmental modifiers), it can also indicate that the onset of a Mendelian disease phenotype requires the presence of additional genetic variants, overlooked during initial diagnosis (Cooper et al., 2013).

As a result, the definitions of Mendelian and complex disorders are becoming less distinct. Many of the features describing complex disorders have now been reported in classical Mendelian traits. Modifier events were reported in Huntington's disease, considered to follow the traditional autosomal dominant inheritance pattern (Gusella et al., 2014). Cases of complex inheritance have now been reported in classical Mendelian traits such as cystic fibrosis (Cutting, 2010). With the accrual of genomic data and advances in discovery of disease genes, the number of disorders for which the disease phenotype can be explained by a single genetic mutation is now decreasing. Altogether, these results suggest that single gene disorders are more complex than originally thought and underline the necessity to revisit the traditional dichotomy between Mendelian and non-Mendelian inheritance. Certain Mendelian traits are now being redefined as complex disorders, requiring the joint effect of multiple genetic events. This is the case of holoprosencephaly (HPE) - a complex pathology of forebrain development, described in the next chapters.

Overall, modern studies focus on understanding the context-dependent effects of both rare and common variants in the pathogenesis of genetic disorders. By challenging the concepts of genetic causality existing since the Mendelian era, future explorations of genetic mechanisms underlying both Mendelian and complex traits will elucidate their etiology, leading to more accurate diagnosis and better patient management.

## II. Holoprosencephaly

### a. Definition

Holoprosencephaly (HPE, OMIM # #236100) is a cerebral congenital malformation, characterized by a failure of cleavage of the two cerebral hemispheres (Dubourg et al., 2007). The disease results from incomplete or absent division of the forebrain (prosencephalon) at early stages of embryonic development. This developmental defect remains one of the most common brain abnormalities, occurring in approximately 1 of 8000 live births and in 1 of 250 conceptuses (Matsunaga and Shiota, 1977; Leoncini et al., 2008).

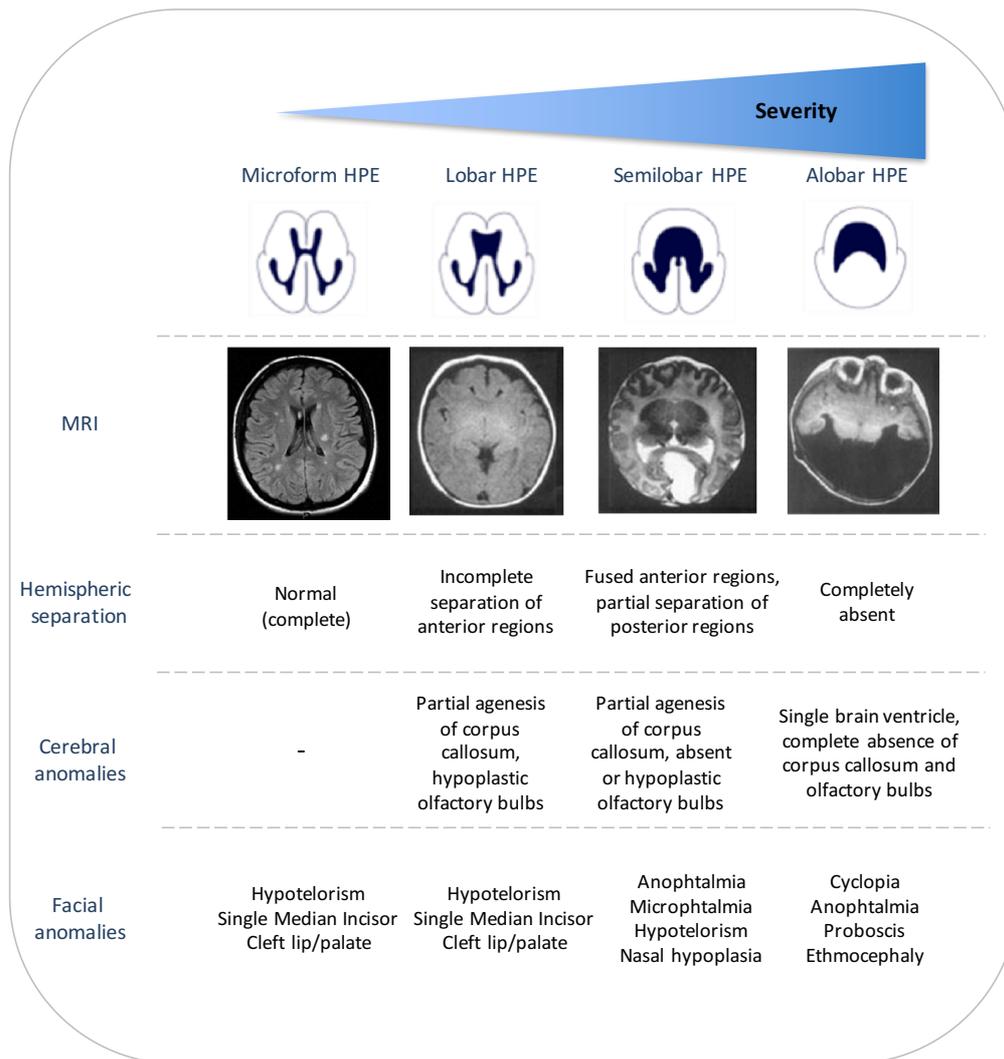
### b. Clinical spectrum

#### *HPE subtypes*

HPE encompasses a wide spectrum of midline craniofacial malformations of varying severity. Depending on the degree of separation of the cerebral hemispheres, the phenotypic continuum of HPE defects has been classified into several anatomic subtypes, in decreasing order of severity: alobar, semilobar, lobar and microform HPE (Figure 2) (Demyer et al., 1964; Tekendo-Ngongang et al., 2000).

Alobar HPE represents the most severe form of the disease and is characterized by a complete absence of hemispheric separation leading to a single large brain ventricle (monoventricle) with fused frontal lobes, thalami and basal ganglia. Clinical manifestations generally include extremely severe facial anomalies such as cyclopia, proboscis, ethmocephaly and anophthalmia. Semilobar HPE constitutes the intermediate form of the disease in terms of severity and is characterized by a partial interhemispheric cleavage. Specifically, the separation is partially present in posterior regions of the brain, while the anterior regions (cortex, basal ganglia, thalamus) remain fused. Facial aspect can vary from relatively normal to mild/severe manifestations including ocular (hypotelorism, microphthalmia), nasal (depressed nasal ridge, absent nasal septum) and mouth (cleft lip/palate) malformations. Lobar HPE is considered as

the least severe form of the three classical subtypes. The interhemispheric cleavage is normal along almost the entire brain midline, including the thalami and most of the basal ganglia. Therefore, the hemispheric separation is almost complete except for the frontal lobes and the most anterior parts of the brain. Clinical manifestations generally include mild facial anomalies such as slight hypotelorism, single median incisor and bilateral cleft lip. Of note, such mild anomalies can be overlooked in slightly affected family members (Lacbawan et al 2009, Solomon et al 2009). Finally, microform HPE is characterized by midline craniofacial anomalies found in classic HPE subtypes, but without the usual defect of interhemispheric cleavage (Tekendo-Ngongang et al., 2000). Microforms are generally identified in the relatives of patients affected by more severe HPE subtypes, illustrating the variable expressivity of this disorder.



**Figure 2. Main anatomical subtypes of HPE.** Cerebral and facial anomalies frequently associated to each subtype are described. MRI - Magnetic resonance imaging.

Over the years, additional subtypes of the HPE spectrum presenting distinct neurological features have been described. Syntelencephaly, also called the Middle Interhemispheric Fusion Variant (MIHV), represents a rare form of the disease (2 to 15 % of total cases), characterized by abnormal midline connections of the cerebral hemispheres in posterior, frontal and parietal regions (Simon et al., 2002). In contrast to classic HPE, syntelencephaly presents a normally separated basal forebrain resulting in milder symptoms and suggesting a distinct embryological origin of this particular subtype (Fernandes et al., 2007). Septopreoptic HPE, described in a small case series, represents another mild subtype of the disease, in which the midline fusion is

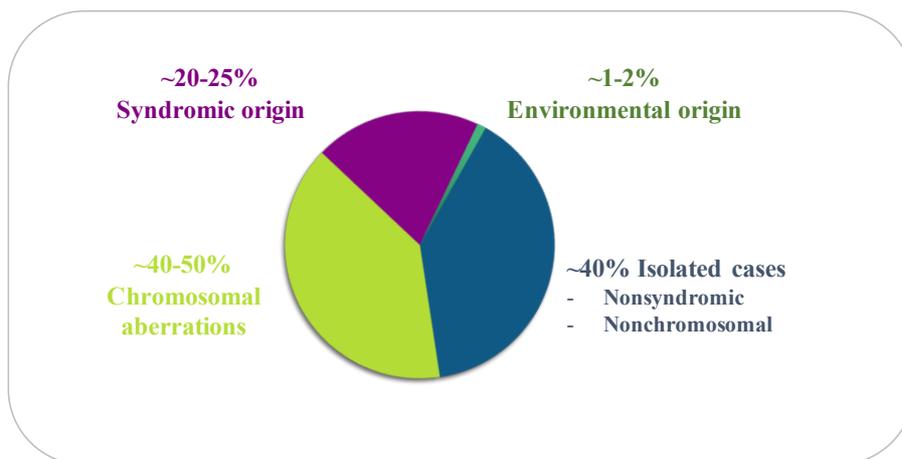
restricted to the septal and/or preoptic regions of the telencephalon (Hahn et al., 2010). Additionally, a subset of patients affected by schizencephaly (a rare congenital brain malformation characterized by deep clefts) was reported to carry HPE-causing mutations, suggesting a partially common genetic background between the two disorders and further expanding the clinical spectrum of HPE (Hehr et al., 2010).

#### *Associated clinical findings*

HPE is often associated with additional extracephalic clinical features (Martinez et al., 2018). Common manifestations include severe developmental and psychomotor delays, microcephaly and seizures, which are observed in all HPE subtypes except certain microforms. Short stature and general growth delay are common in severely affected individuals, partially caused by the associated growth hormone deficiency and/or chromosome anomalies. Central diabetes insipidus caused by pituitary dysfunction is commonly found in patients affected by nonchromosomal, nonsyndromic HPE (Lacbawan et al 2009, Solomon et al 2010). Hydrocephalus resulting in macrocephaly can be encountered in about one-sixth of HPE patients although the underlying cause remains unclear (Levey et al., 2010). Pulmonary complications (pneumopathies, lung hypoplasia, chronic respiratory disorders) and gastrointestinal dysfunctions such as anal atresia and cholestasis have also been reported. Additional extracephalic manifestations observed in HPE include abnormalities of cardiovascular (situs ambiguous, septal defect), genitourinary (micropenis, renal anomalies) and skeletal (polydactyly, clinodactily, costal/vertebretal anomalies) systems. Affected children present with feeding and swallowing difficulties, partially caused by axial hypotonia, neurological complications and side effects of medical treatments. A small portion of patients presents HPE associated with NTDs (rachischisis, spina bifida). Interestingly, many clinical features encountered in HPE subjects are also characteristic of other developmental pathologies such as ciliopathies (Meckel-Gruber syndrome, Bardet-Biedl syndrome) and NTDs, suggesting a partially common etiology between these disorders.

### c. Etiology

The etiology of HPE is complex and heterogeneous (Krauss, 2007). The high variability of clinical manifestations indicates a multisystemic nature of the disease and makes the molecular diagnosis particularly challenging. The disease can involve both genetic and environmental factors or occur as a part of a polymalformative syndrome. Known etiologies of HPE are shown in Figure 3 and described below.



**Figure 3. Etiological causes of HPE.** The estimated proportions are based on recent reports of Dubourg et al., 2018; Solomon et al., 2018.

#### *Chromosomal aberrations*

Approximately 40-50 % of total HPE cases are caused by chromosome abnormalities, which are generally found in the most severe forms of the disease (Muenke and Beachy, 2001; Solomon et al., 2010; Abe et al., 2018). Trisomy 13, trisomy 18 and triploidies were the first cytogenetic anomalies reported in HPE cases. Trisomy 13 is encountered in 40-60% of HPE fetuses and remains the most common disease etiology. Recent screen of over 50, 000 pregnancies reported that 86 % of abnormal karyotypes found in HPE correspond to trisomy 13 (Kagan et al., 2010). Trisomy 18 is much less frequent and is found in 1-2 % of HPE patients (Muenke and Beachy, 2001; Solomon et al., 2010). Recent studies report that trisomy 18 could have a higher prevalence in HPE (~7 %), but these findings are limited by small sample sizes (Rosa et al., 2017, Petracchi et al., 2011). Triploidy is also believed to account for a substantial part of HPE etiology (Solomon et al., 2010). Additionally, rare cases of HPE associated with trisomy 21 (n =

5), trisomy 22 (n = 3) and trisomy 16 (n = 1) have been reported (Kruzska and Muenke, 2018). Most cases of chromosomal abnormalities represent isolated occurrences. Several recurrences have been reported, however, such as triploidy of maternal origin (Brancati et al., 2003). Overall, the recurrence risk of HPE due to chromosomal anomalies is estimated at 1 % (Peebles, 1998).

### *Syndromic HPE*

In approximately 20 % of cases, HPE is syndromic, *i.e.*, occurring as a part of various polymalformative syndromes described below.

Chromosome 13q deletion syndrome is a rare genetic disorder characterized by partial deletions of chromosome 13 resulting in various birth defects. Depending on the size and location of the deleted region, clinical symptoms differ among patients and include facial anomalies (hypotelorism, micrognathia), eye abnormalities (anophthalmia, microphthalmia) and various developmental delays. Multiple studies report cases of chromosome 13q deletion syndromes associated with HPE phenotype (Junior et al., 2006; Gutierrez et al., 2001; Garcia-Rodriguez et al., 2015). Interestingly, the 13q deletions in such cases often encompass *ZIC2*, one of the major HPE-associated genes (described further), indicating that the HPE phenotype encountered in 13q deletion syndrome could be linked to *ZIC2* haploinsufficiency (Quelin et al., 2009).

Similar link was observed between HPE and 18p deletion syndrome. 18p syndrome corresponds to deletion of the short arm of chromosome 18 resulting in various dysmorphic features, mental retardation and growth deficiencies. 18p deletions found in patients exhibiting HPE phenotype often encompass *TGIF1*, one of the minor genes implicated in HPE (Portnoi et al., 2007; Chen et al., 2013; Yi et al., 2014).

Hartsfield syndrome is defined as a combined manifestation of HPE and ectrodactily, with or without cleft lip/palate. This disorder is caused by mutations in *FGFR1*, a major component of the FGF pathway (Simonis et al. 2013). Interestingly, a recent study of

257 HPE cases identified 6 pathogenic mutations in *FGFR1*, 4 of which were found in patients presenting HPE without ectrodactily, indicating that *FGFR1* mutations can also account for a substantial part of classic HPE spectrum (Dubourg et al., 2016).

HPE is also encountered in Smith-Lemli-Opitz syndrome (SLOS), resulting from inborn errors of cholesterol metabolism (Porter, 2008). SLOS is caused by mutations in *DHCR7*, which encodes an enzyme implicated in catalysis of reduction of 7-dehydrocholesterol and its subsequent conversion to cholesterol. SLOS is characterized by malformations at the craniofacial level (microcephaly, cleft lip/palate, ocular dysfunctions, micrognathia), intellectual disability, anomalies of cardio-vascular and gastro-intestinal systems, as well as external genitalia defects in males. Approximately 5 % of SLOS patients present with clinical signs of HPE, indicating the role of cholesterol metabolism in the disease pathogenesis (Kruzska and Muenke, 2018). Interestingly, cholesterol has also been shown to affect the activity of Sonic Hedgehog signaling (SHH), one of the major pathways implicated in HPE (Riobo, 2012).

HPE is found in approximately 2 % of patients affected by CHARGE syndrome (Colobomata, Heart defect, choanal Atresia, Growth and/or developmental Retardation, Genital hypoplasia and Ear abnormalities) (Lin et al., 1990). Approximately 50 % of CHARGE syndrome is caused by mutations in *CHD7*, which encodes a DNA-binding protein of the chromodomain helicase (CHD) family, implicated in differentiation of craniofacial tissues during embryonic development (Zentner et al., 2010).

In rare cases, HPE has been observed in patients with Pallister-Hall syndrome, characterized by hypothalamic hamartomas, pituitary dysfunction, central polydactyly and visceral malformations (Biesecker et al., 1996). Finally, HPE phenotypes have also been reported in Lambotte (Verloes et al., 1990), hydrolethalus (Bachman et al., 1990), Genoa (Camera et al., 1993), agnathia-microstomia-synotia (Wai & Chandran, 2017),

agnathia-otocephaly (Faye-Petersen et al., 2006) and Steinfeld's syndromes (Stevens et al., 2010).

#### *Non-genetic risk factors in HPE*

Although a substantial part of HPE cases result from genetic alterations, non-genetic risk factors can also contribute to the disease pathogenesis. Epidemiological studies of HPE are challenging due to its low birth prevalence. Nevertheless, environmental factors, acting alone or in combination with additional genetic alterations, are believed to account for approximately 1 % of total HPE cases.

Maternal diabetes, a known risk factor for various birth defects, is one of the most extensively studied in HPE. A relatively high prevalence of diabetes (6-9 %) is observed among mothers of HPE patients (Olsen et al., 1997). Several studies reported that the incidence of diabetes among mothers of HPE children is twice as high compared to controls (Martinez-Frias et al., 1998; Anderson et al., 2005; Correa et al., 2008). These findings are consistent with animal studies: inducing diabetes in pregnant mice has been shown to increase the risk of HPE, along with other congenital anomalies (Padmanabhan and Shafiullah, 2004).

Epidemiological and animal studies have also illustrated a link between HPE and maternal alcoholism during early stages of pregnancy (Johnson and Rasmussen, 2010). Ethanol is a known teratogen, which induces various birth defects in humans. Known teratogenic effects involve ethanol metabolism, which is known to induce oxidative stress; and ethanol-derived products, such as ethanol-derived acetaldehyde implicated in exencephaly (Bhatia et al., 2019). However, recent animal study concluded that ethanol itself, rather than its metabolism or derived products, can act as an HPE-inducing teratogen, in combination with additional genetic factors (Hong and Krauss, 2017).

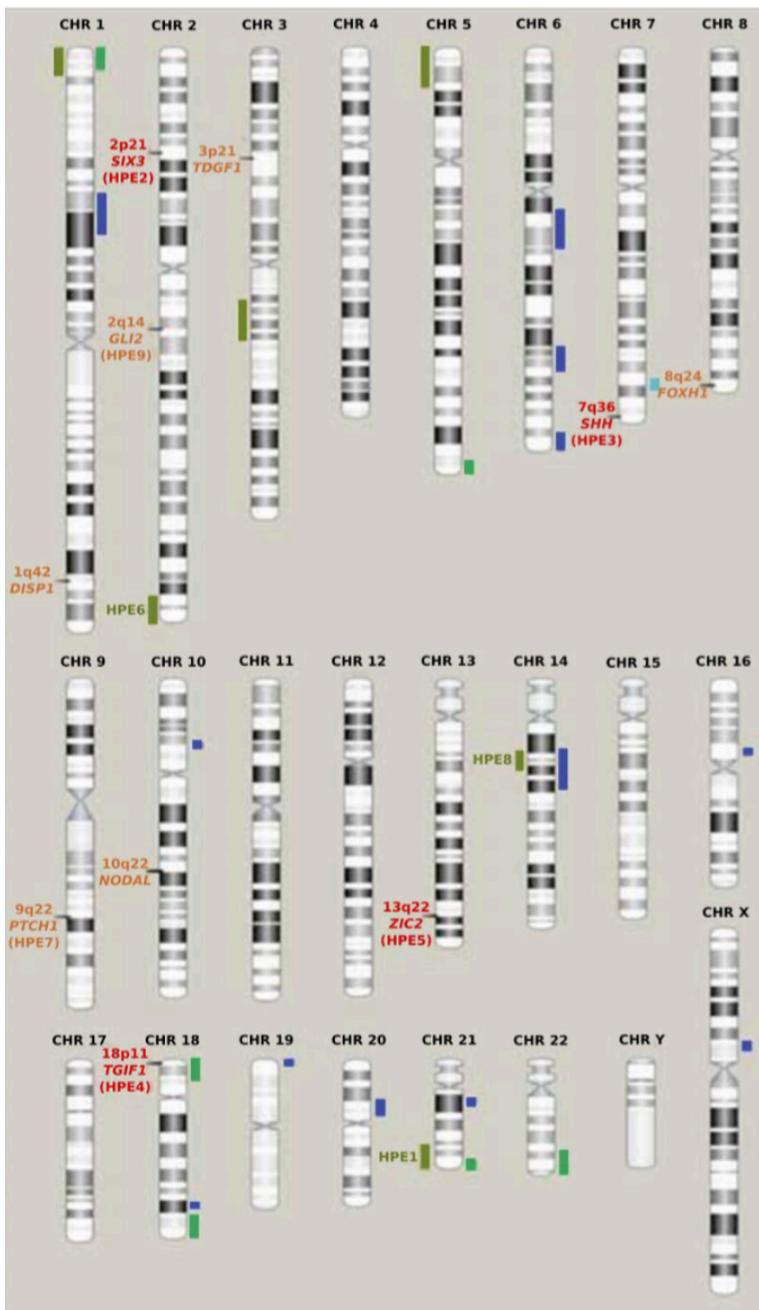
Additional risk factors reported in HPE include alterations of cholesterol biosynthesis and the use of certain medications during pregnancy such as retinoids, antiepileptics and hormone supplements (Miller et al., 2010).

### *Isolated HPE*

Nonsyndromic HPE without chromosomal aberrations and known environmental causes accounts for approximately 40 % of total disease cases (Muenke and Beachy, 2001). Such cases are referred to as 'isolated' and remain the most enigmatic due to their low diagnostic yield and unelucidated etiology. Isolated HPE is believed to be caused by genetic factors involving partial chromosomal deletions/duplications (CNVs) and/or point mutations.

CNVs account for a substantial portion of isolated HPE cases. In 2009, an impressive rate of 19 HPE patients among 111 (~17 %) presenting *de novo* CNVs was reported (Bendavid et al., 2009). Recent study confirmed these numbers by identifying clinically significant CNVs in 23/222 individuals (21 %), including several recurrent deletions identified in previous studies (Hu et al., 2018).

Detection of recurrent CNVs in HPE led to identifications of specific patterns of CNV anomalies associated with the disease (Dubourg et al., 2007; Bendavid et al., 2010) (Figure 4). Studies of the corresponding regions led to the discovery of first causal HPE genes: *SHH* at 7q36 (Belloni et al., 1996; Roessler et al., 1996), *ZIC2* at 13q32 (Brown et al., 1998), *SIX3* at 2p21 (Wallis et al., 1999) and *TGIF1* at 18p11.3 (Gripp et al., 2000). Deletions affecting one of these four genes are identified in approximately 5-8 % of HPE patients presenting a normal karyotype (Bendavid et al., 2006; Stokes et al., 2018).



**Figure 4. Cytogenetic abnormalities and HPE genes.**

On the left side: karyotypes represent deletions detected by routine karyotype.

The locations of major (red) and minor (orange) HPE genes are represented.

On the right side: green bars represent rearrangements identified by subtelomeric multiplex ligation-dependent probe amplification (MLPA), blue bars represent *de novo* (dark blue) and inherited (light blue) deletions detected by array CGH (Bendavid et al., 2010).

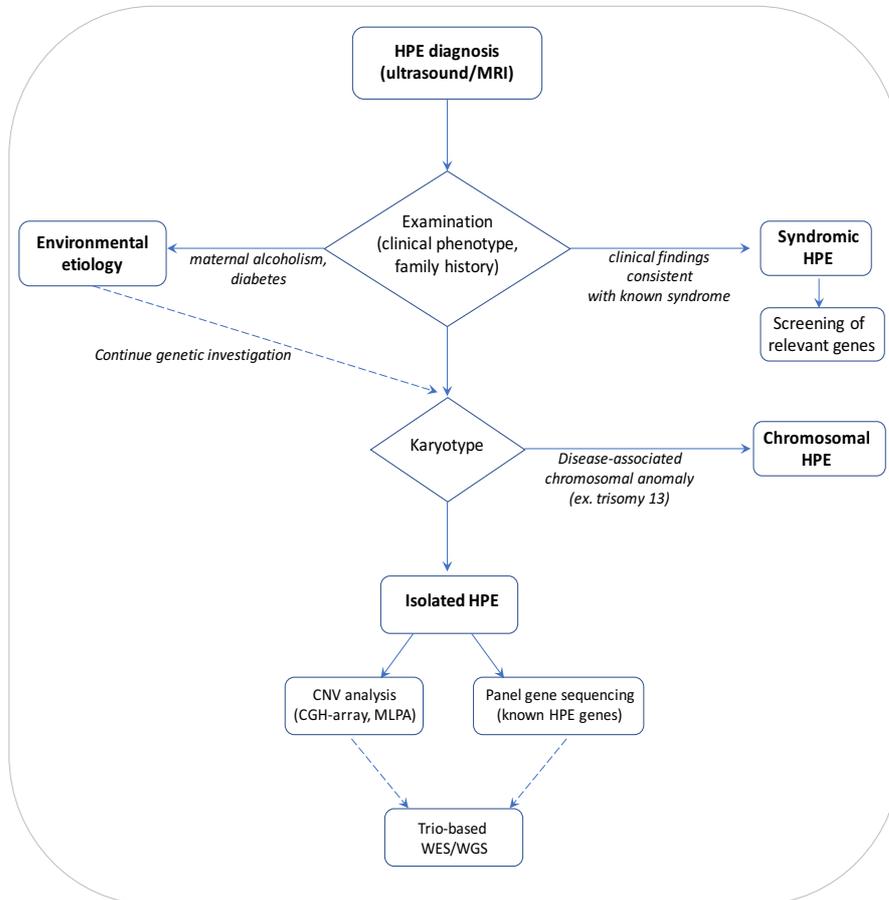
Follow-up studies revealed that pathogenic point mutations (SNVs and Indels) can also underlie a substantial part of isolated HPE cases. Approximately 25 % of patients without any CNVs/chromosomal anomalies harbor pathogenic point mutations in *SHH*, *ZIC2*, *SIX3* and *TGIF1* - known disease genes shown in Figure 4 and described in part III of the Introduction (Dubourg et al., 2007; Mercier et al., 2011).

With the advent of new technologies, detection of recurrent chromosomal anomalies by CGH-array and multiplex ligation-dependent probe amplification (MLPA), combined with the analysis of sequence variants by Next Generation Sequencing (NGS), have greatly expanded our knowledge of the genetic spectrum of isolated HPE. To date, it has been shown that at least 16 genes are implicated in the disease. All known HPE genes belong to major pathways of embryonic development (Sonic Hedgehog, Notch, Nodal, FGF), described in part III of the Introduction.

#### **d. Diagnostic and molecular testing**

The guidelines for diagnosis and testing of HPE are illustrated in Figure 5. Severe cases of alobar HPE can be detected during routine prenatal ultrasound examination (between 10<sup>th</sup> and 14<sup>th</sup> gestational week), generally based on the presence of abnormal facial morphology. Milder cases (semilobar and lobar HPE) cannot be reliably detected by ultrasound, therefore, a second-line investigation by fetal MRI is generally recommended if ultrasound studies suggest the presence of an anomaly. If the HPE diagnosis is established, family history and fetus phenotype are thoroughly examined to evaluate the possibility of environmental (maternal alcoholism, diabetes) or syndromic (consistent clinical findings) etiology. A karyotype study is performed to detect the presence of chromosomal aberrations (trisomy 13/18, triploidy) known to cause HPE. In case of negative results of the abovementioned studies, HPE is then considered as isolated. Fetal DNA samples are extracted for molecular testing, which involves microarray studies (CGH-array, MLPA) and targeted sequencing of known HPE genes to detect pathogenic CNVs and/or point mutations underlying the disease. If a pathogenic CNV/mutation is found, this analysis is extended to parents to evaluate its inheritance pattern.

In case of all the abovementioned exams bearing negative results, a trio-based Whole Exome/Genome sequencing is generally recommended to conduct a thorough analysis of genomic alterations potentially underlying the disease.



**Figure 5. Recommended genetic testing for HPE.** Adapted from Kruszka et al., 2018.

Despite numerous advances in molecular and clinical genetics, most patients affected by isolated HPE remain without an established molecular diagnosis. Resolving these enigmatic cases is critical provide better patient management in HPE. This requires further elucidating of complex molecular and genetic mechanisms underlying the disease pathogenesis.

### III. Molecular and genetic basis of HPE

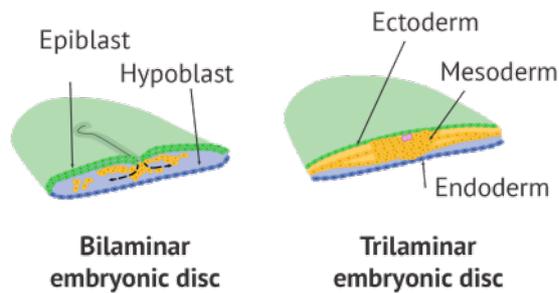
Molecular and genetic events leading to HPE have been studied extensively throughout the years. Clinical studies of HPE patients identified specific genes and pathways involved in the disease pathogenesis. Animal studies further elucidated molecular and genetic mechanisms of forebrain and craniofacial development underlying HPE.

#### a. Forebrain and facial development

The main clinical features of HPE are impaired forebrain structures and craniofacial midline defects. Therefore, understanding the disease pathogenesis requires knowledge of the mechanisms of forebrain and craniofacial development. As demonstrated in mice studies, the induction of HPE generally takes place during early embryonic development, corresponding to a period between 3<sup>rd</sup> and 4<sup>th</sup> gestational week in human (Lipinski et al., 2010; Kauvar and Muenke, 2010). The developmental and molecular events taking place during this period are briefly described below.

##### *Early stages and gastrulation*

At the end of the second week post-conception, the human embryo corresponds to an oval-shaped, two-layered structure called the bilaminar embryonic disc (Figure 6). Each of the two layers contains different, very primitive cells: the cells of the upper layer (epiblast) will eventually contribute to the formation of early embryonic structures; the cells of the lower layer (hypoblast) will contribute to the formation of extraembryonic structures.



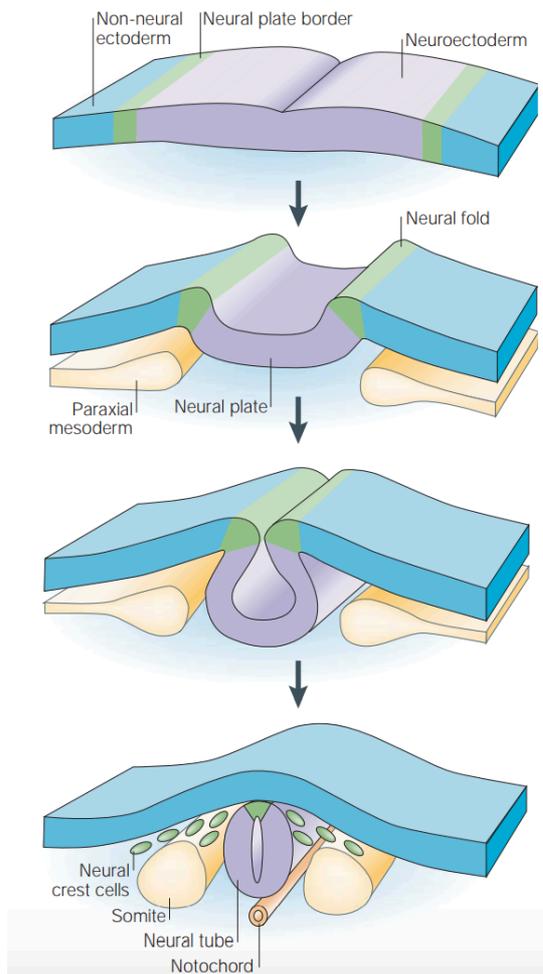
**Figure 6. Schematic representation of early embryonic development.**

www.invitro.com

During the 3<sup>rd</sup> week of embryonic development, the embryo undergoes the gastrulation phase which consists of several developmental processes including the establishment of anteroposterior/dorsoventral axes and highly organized cell migrations. These cell migrations, also called morphogenetic movements, will transform the embryo into a three-layered structure (also called the trilaminar embryonic disc) containing ectodermal and endodermal layers with the formation of an intermediate, mesodermal layer between the two (Figure 6). The endodermal cell layer will give rise to the gut and respiratory tract. The mesoderm, containing the prechordal plate and the notochord, will form bone, muscle, cartilage and the vascular system. Cells of the ectodermal layer are divided into two types of stem cells: epidermal ectodermal stem cells form structures such as skin and nails, while neuroectodermal stem cells will act as the neural progenitor cells and give rise to the central nervous system.

### *Neurulation*

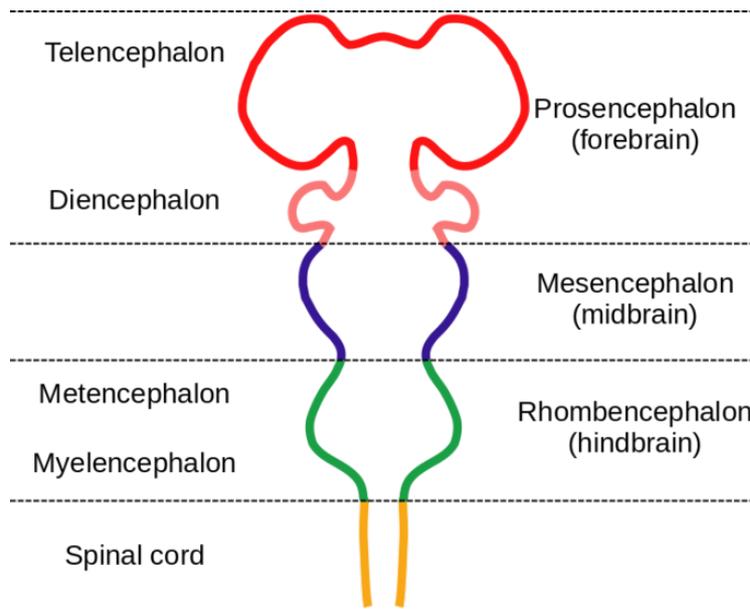
At the end of the gastrulation phase, signals received from the mesodermal prechordal plate and the notochord induce neural progenitor cells to be positioned along the midline of the ectodermal layer and form a region called the neural plate. The extremities of the neural plate, known as the neural folds, rise, fold inward and fuse to form a neural tube - the first well-defined structure of the central nervous system. This process is called the neurulation phase (Figure 7). Fusion begins in the center of the developing tube and then gradually proceeds to the most rostral and caudal regions.



**Figure 7. Neurulation phase**

(Gammil and Bronner-Fraser, 2003)

The formation of the neural tube is accompanied by a rapid growth of the embryo and major changes in the organization of the central nervous system. At the end of the neural tube formation, the rostral end divides into three main vesicles along the posterior-anterior axis: hindbrain (rhombencephalon), midbrain (mesencephalon) and forebrain (prosencephalon) (Figure 8). Forebrain is the largest, most anterior part of the brain and, by the end of the embryonic period, further divides into telencephalon and diencephalon eventually giving rise to key brain structures such as cerebrum, hypothalamus, thalamus, limbic system, pituitary gland and olfactory bulb. Forebrain is the main structure affected in HPE.



**Figure 8. Primary vesicles of the embryonic brain.**

<https://courses.lumenlearning.com/>

*Forebrain patterning and HPE pathogenesis*

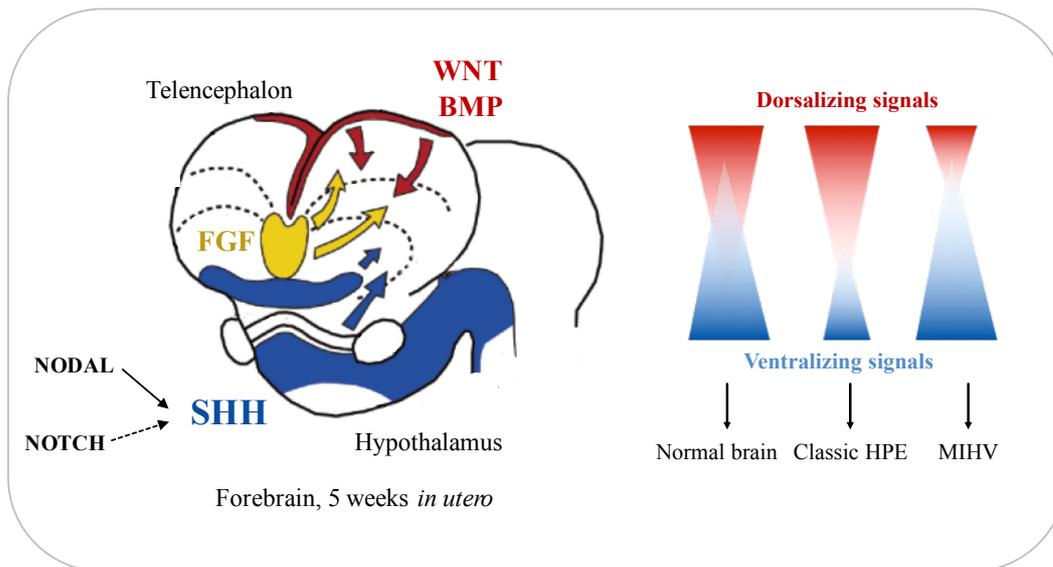
By the end of the embryonic development, forebrain corresponds to a highly organized structure with telencephalon and eyes positioned on the dorsal side, the ventrally positioned hypothalamus and the caudally located diencephalon. The correct dorsoventral positioning of forebrain components is critical for its correct function. Therefore, the critical step in the elaboration of the forebrain is its dorsoventral patterning, which involves a set of morphogens acting as dorsalizing or ventralizing signals (Wilson and Houart, 2004; Geng and Oliver, 2009).

A major morphogen that promotes the ventral patterning of the forebrain is Sonic Hedgehog (SHH). Produced in notochord and prechordal plate, SHH is then secreted into the ventral midline of the forebrain (Bertrand and Dahmane, 2006). SHH acts as ventralizing signal of forebrain patterning and contributes to the separation of the single eye field into left and right eye. The correct expression of SHH on the ventral side of the forebrain is regulated by a variety of factors including NODAL and NOTCH signaling pathways; and transcriptional activator *SIX3* (Müller et al., 2000; Dupé et al., 2011; Jeong et al., 2008). Environmental factors, such as retinoic acid or ethyl alcohol have also been shown to impact on SHH activity (Petryk et al., 2015). Moreover, mouse

studies indicate that *FGF8* and *ZIC2*, secreted in the midline, act as additional factors in concert with SHH to assure the correct patterning of the forebrain on the ventral side (Cheng et al., 2006; Strom et al., 2008).

On the dorsal side of the forebrain, several Bone Morphogenic Proteins (BMP2, BMP4-7) and members of the WNT pathway (WNT1, WNT3, WNT3a, WNT4, WNT7b), expressed in discrete and overlapping patterns, act as dorsalizing morphogens (Bond et al., 2012; Furuta et al., 1997; Harrison-Uy and Pleasure, 2012). BMPs and WNTs are secreted from the dorsal midline of the telencephalon, specifically the roofplate and the cortical hem.

SHH, WNT and BMP proteins, expressed in a precise spatio-temporal manner, establish a dorsoventral gradient controlling the correct patterning of the forebrain (Figure 9). The perturbations of this gradient lead to incorrect or absent forebrain structures and HPE pathogenesis (Fernandes and Hebert, 2008), as demonstrated by animal studies. Inactivation of *Shh* in mice leads to severe HPE-related defects including absent ventral structures and cyclopia (Chiang et al., 1996). Partial or complete inactivation of members of the SHH signaling pathway (*DISP1*, *GLI2*, *SMO*) and/or other ventralizing factors (*ZIC2*, *FGF8*) cause various manifestations of semilobar and alobar HPE in mice (Warr, 2008; Hayhurst and McConnell, 2003). Interestingly, while classic HPE is caused by a defective ventralization of the forebrain (mainly resulting from absent or insufficient SHH signaling), defective dorsalizing signals lead to the MIHV subtype of HPE. Inactivation of *BMPR1a* and *BMPR1b*, members of the BMP pathway, results in MIHV phenotype in mice (Fernandes et al., 2007; Gupta and Sen, 2016). Moreover, inhibition of BMP signaling by ectopic expression of SHH in the dorsal midline leads to MIHV (Huang et al., 2007).

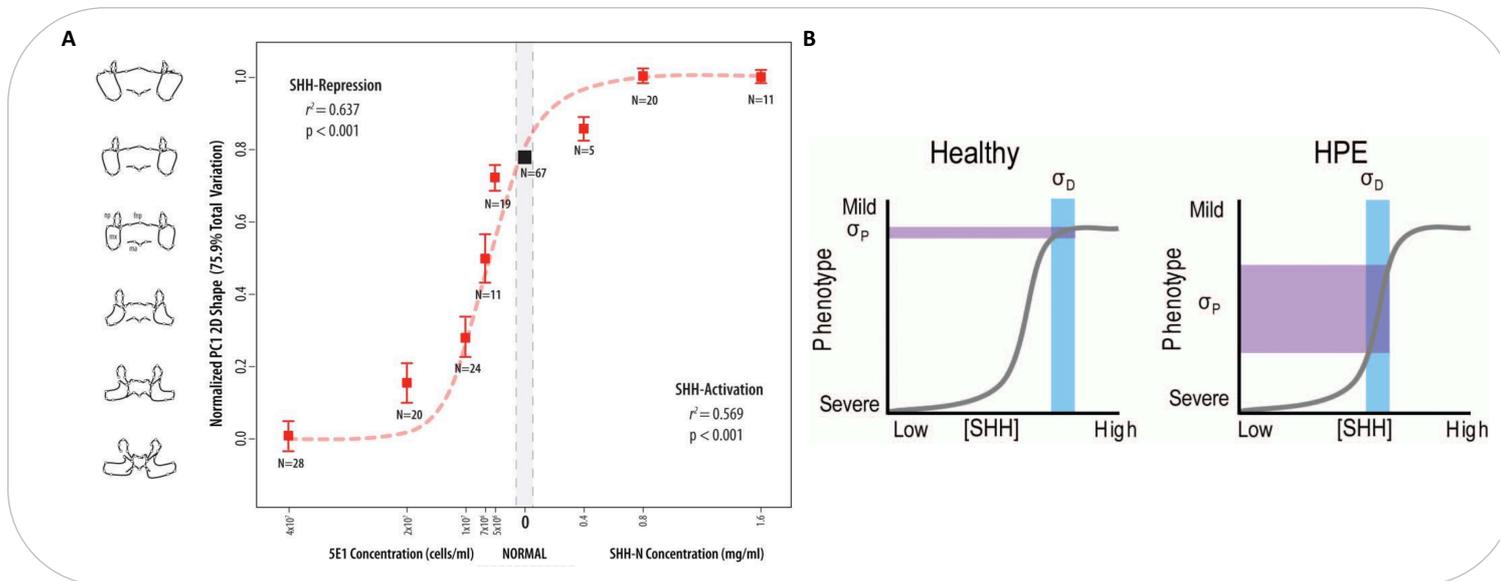


**Figure 9. Dorsoventral patterning of forebrain and HPE pathogenesis.**

#### *Signaling interactions in forebrain and facial development*

As illustrated by various HPE cases described in part II of the Introduction, impaired forebrain structures are typically associated with facial defects, indicating that defective forebrain development also impacts the morphogenesis of facial structures. In addition to the physical impact that growing brain has on facial shape, multiple studies have shown that forebrain coordinates facial development at the molecular level. In particular, SHH signaling directly affects facial structures by controlling the spatial organization of the Frontonasal Ectodermal Zone (FEZ), a signaling center that regulates facial development (Hu et al., 2003). Blocking *SHH* expression in the brain leads to altered expression of *SHH* in the FEZ and causes midline facial anomalies typically encountered in HPE (Marucio et al., 2005; Hu and Marucio, 2009). Quantitative experiments in chick model have shown that varying SHH concentration in the brain affects the head shape, the midface and the eyes, leading to various facial anomalies such as abnormal frontonasal processes, hypotelorism and jaw abnormalities (Young et al., 2010). Importantly, a significant correlation between SHH concentration and variations of midfacial shape has been illustrated, suggesting that SHH controls facial development in a dose-dependent manner (Figure 10A). The dose-dependent impact of SHH signaling on phenotypic outcome may explain the extreme

clinical variability observed in HPE patients (Marucio et al., 2011; Petryk et al., 2015) (Figure 10B).



**Figure 10. The dose-dependent effect of SHH on phenotypic outcome. (A) Impact of SHH dosage on midfacial shape in chick model.** Quantitative experiments of SHH repression (using anti-SHH antibody 5E1) or activation (SHH-N beads) illustrate a correlation between the SHH dosage (X axis) and midfacial configuration (Y axis). Bars represent standard error; the dashed line corresponds to the maximum likelihood fit (nonlinear Hill equation). Adapted from Young et al., 2010. **(B) Variations of SHH signaling may explain clinical variability of HPE.** Model proposed by Petryk et al., 2015.

Other molecular pathways of forebrain development have also been linked to facial development. Altering the expression of BMP proteins in neural crest cells results in dramatic changes of the facial skeleton leading to orofacial clefts and mandibular defects (Bonilla-Claudio et al., 2012). Conditional knockout of  $\beta$ -catenin, the key molecule of WNT signaling, results in absent upper jaw and dorsal face (Wang et al., 2011).

Altogether, these results indicate that forebrain and facial development are linked and coordinated by a complex interplay between multiple developmental pathways. Defects in these pathways result in incorrect dorsoventral patterning of the forebrain leading to the onset of HPE and associated craniofacial anomalies.

## b. Molecular pathways of HPE

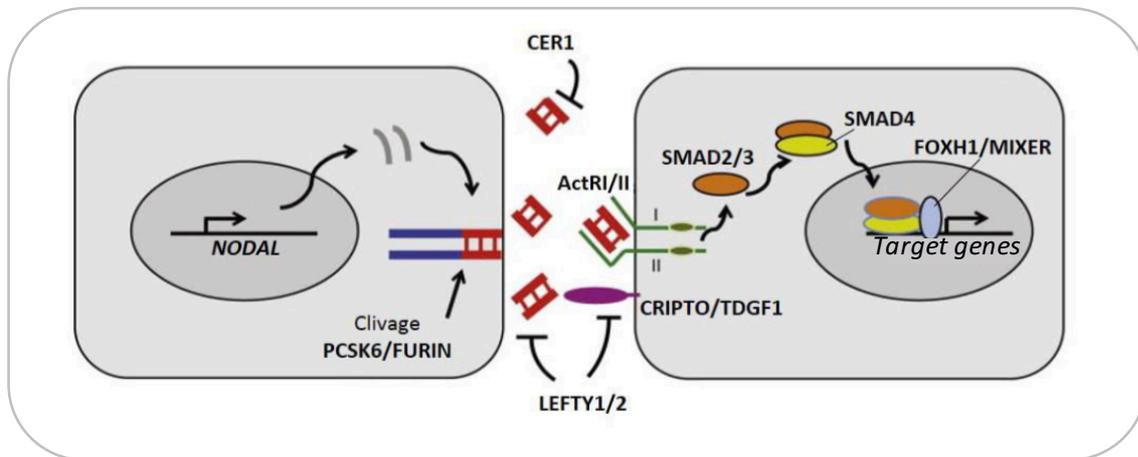
### *NODAL*

Highly conserved among vertebrates, NODAL signaling is essential for global embryonic polarity and differentiation events that take place during pre-gastrulation and gastrulation phases (Lu and Robertson, 2004).

During pre-gastrulation, NODAL signaling is crucial for the patterning of the anterior-posterior axis and the establishment of epiblast-hypoblast layers (Rossant and Tam, 2004). Specifically, *Nodal* is expressed in the early epiblast and induces the formation of the anterior hypoblast (Shen, 2007). The anterior hypoblast then produces NODAL antagonists (LEFTY1/2, CER1), thus preventing excessive NODAL activity (Perea-Gomez et al., 2002). During gastrulation, *Nodal* expression is required for the correct formation of mesodermal and endodermal layers (Vincent et al., 2003). In particular, NODAL signaling controls mesoderm differentiation, inducing axial mesoderm and exerting a dorsalizing activity (Jones et al., 1995). As demonstrated by animal studies, NODAL is also critical for establishment of the prechordal plate, the major secreting center of *SHH* (Sampath et al., 1998). Knockouts of NODAL components lead to defective patterning of the prechordal plate, reduced *SHH* expression and defective forebrain structures (Heyer et al., 1999; Hoodless et al., 2001; Nomura and Li, 1998; Song et al., 1999).

NODAL belongs to the superfamily of Transforming Growth Factor- $\beta$  (TGF- $\beta$ ) regulating proteins (Shen, 2007; Robertson et al., 2014) (Figure 11). Initially synthesized as a precursor, NODAL is cleaved into its active form by a protease PCSK6 and transduced through the transmembrane activin receptors (ACVR1B, ACVR2A/B). Binding of NODAL to activin receptors leads to the phosphorylation of SMAD proteins (SMAD2/3), which subsequently bind to SMAD4 and translocate to the nucleus. The SMAD2/3-4 complex activates transcriptional regulators such as *FOXH1*, thus enabling the transcription of target genes. NODAL signaling is regulated by extracellular co-factors including TDGF1

and CFC1, which are members of the epidermal growth factor (EGF)-cysteine rich (CFC) family.



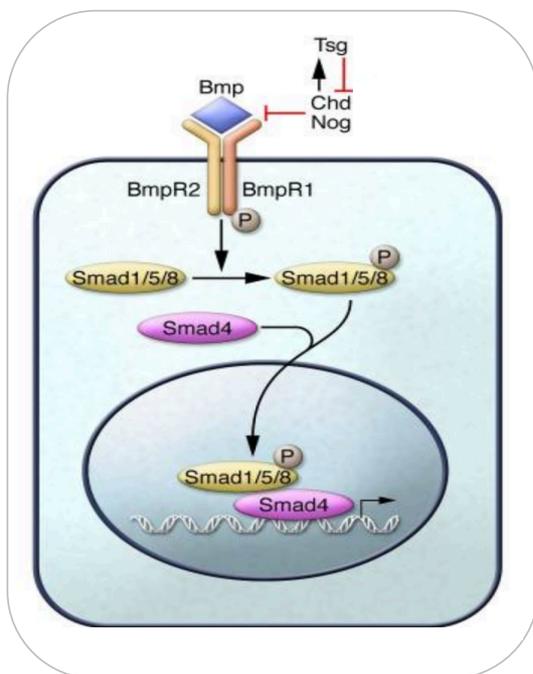
**Figure 11. NODAL signaling pathway.** Adapted from Robertson et al., 2014.

### *BMP*

Initially discovered for their role in ectopic bone formation, Bone Morphogenetic Proteins (BMPs) are now known to act as major morphogens during embryogenesis and development (Wang et al., 2014). More than 15 BMPs are known, all implicated in various developmental processes. In particular, *BMP2* and *BMP4* are essential to embryogenesis, as knockout of these genes in mouse leads to embryonic lethality (Winnier et al., 1995; Zhang and Bradley, 1996). Importantly, several BMP proteins (*BMP2*, *BMP4*, *BMP5*, *BMP7*) have been proposed to act as important dorsalizing factors during forebrain development (Furuta et al., 1997).

BMPs belong to another branch of the TGF- $\beta$  superfamily and bind to transmembrane BMP-specific receptors (*BMPR1*, *BMPR2*) (Figure 12). Similar to NODAL signaling, BMPs activate SMAD complexes upon binding, which then translocate to the nucleus and activate the expression of target genes. Chordin (*Chd*), noggin (*Nog*) and Twisted gastrulation (*Tsg*) proteins regulate the BMP signaling by preventing the receptor binding. In mouse studies, double inactivation of *Chd* and *Nog* as well as homozygous knockouts of *Tsg* cause alobar HPE-like phenotype (Bachiller et al., 2000; Zakin and Robertis, 2003). As mentioned above, inactivation of *BMPR1* receptors (*BMPR1A*,

*BMPR1B*) leads to the MIHV phenotype. Interestingly, multiple studies have shown that NODAL and BMPs pathways antagonize each other through competitive activation of the common signaling component SMAD4 (Candia et al., 1997, Furtado et al., 2008, Yamamoto et al., 2009). Additionally, BMPs can also bind activin receptors used by NODAL (ACVR1A, ACVR2A/B) (Aykul and Martinez-Hackert, 2016). The antagonistic relationship between these two signaling pathways participates in the regulation of forebrain patterning and could contribute to the complex nature of HPE pathogenesis (Yang et al., 2010).



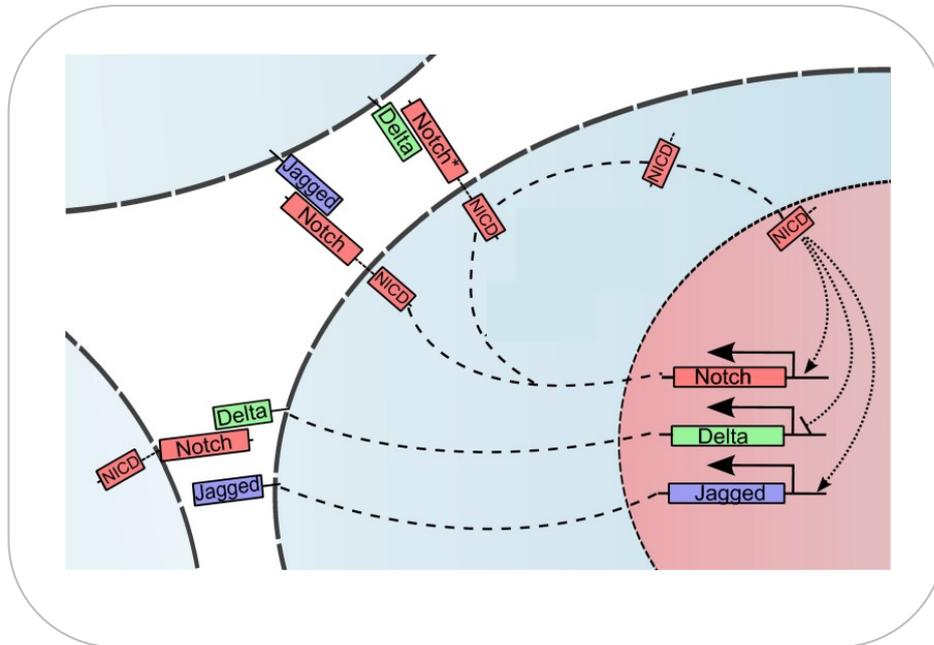
**Figure 12. BMP signaling pathway.** Adapted from Wang et al., 2014.

### *Notch*

Initially characterized in *Drosophila*, Notch signaling acts a general developmental tool of local cell interactions that is used to control a broad spectrum of cell fates and developmental processes (Artavanis Tsakonas et al., 1999).

The activity of the Notch pathway is based on the interaction between two neighbouring cells (*trans*-interaction): a transmembrane NOTCH receptor is activated by binding of ligands Delta (DLL) and/or Jagged (JAG) present on the neighboring cell (Boareto et al., 2015). Once activated, the intracellular domain of the NOTCH receptor (Notch Intracellular Domain, NICD) is translocated to the nucleus, resulting in

subsequent activation of downstream target genes. Once translocated, NICD also regulates the transcription of Notch, Jagged and Delta thus maintaining efficient intercellular communication (Figure 13). Of note, interaction between Notch receptors and ligands of the same cell (*cis*-interaction) results in the degradation of both without generation of any signal.



**Figure 13. Notch signaling pathway.** Adapted from Boareto et al., 2015.

During neurulation, Notch signaling is involved in neural tube patterning by controlling the timing of cell birth and differentiation (Stolfi et al., 2011; Cau and Blader, 2009). The temporal control of neurogenesis by Notch has been further demonstrated by studies of *Xenopus* and chick embryos (Chitnis and Kintner, 1996; Henrique et al., 1997). Recent studies demonstrated that Notch pathway regulates the SHH signal during forebrain development, indicating its relevance in HPE (Dupé et al., 2011). Specifically, Notch activity controls the localization of several members of the SHH pathway (SMO, PTCH1) within the primary cilia (Kong et al., 2015; Stasiulewicz et al., 2015). Pathogenic variants in *DLL1*, a ligand of the Notch pathway, were identified in HPE patients (Dupé et al., 2011). Moreover, a recent study conducted by our research team further revealed that deficiency of NOTCH signaling results in decreased SHH activity leading to malformations of the forebrain midline (Hamdi-Roze et al., 2020).

### *FGF signaling*

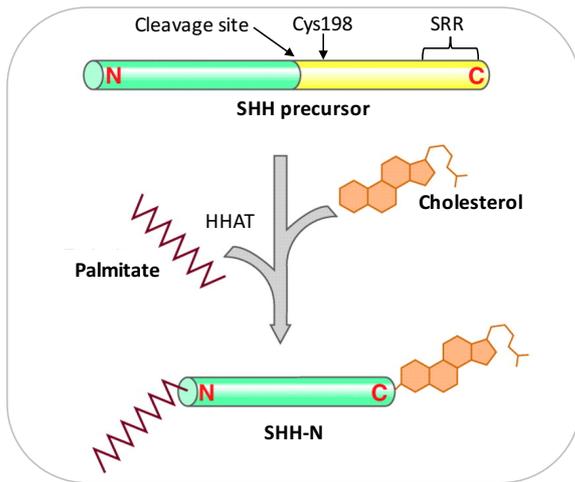
The Fibroblast Growth Factor (FGF) signaling pathway regulates fundamental cellular processes including cell proliferation, survival, migration and differentiation (Ornitz and Itoh, 2015). Prior to gastrulation, FGFs are implicated in the formation of epiblast. During gastrulation, FGF signaling induces chemotactic activity that controls morphogenetic movements necessary to the formation of endodermal, mesodermal and ectodermal layers (Kubota and Ito, 2000). In particular, *Fgf8* is crucial for this process as mice deficient for this gene lack all embryonic structures derived from the mesodermal and endodermal layers (Meyers et al., 1998). During forebrain development, FGF signaling functions downstream of SHH and participates in separation of the anterior cerebral hemispheres (Storm et al., 2006; Gutin et al., 2006). In mice, conditional deletions of *Fgf8* and *Fgfr1* result in reduced FGF signaling during forebrain development and semilobar HPE (McCabe et al., 2011; Simonis et al., 2013). The FGF pathway is comprised of secreted signaling ligands (FGFs) and a series of tyrosine kinase receptors (FGFRs). Upon ligand-receptor binding, activated FGFRs phosphorylate specific tyrosine residues and regulate the expression of target genes via one of the four downstream pathways: RAS-MAPK, PI3K-AKT, Phospholipase C $\gamma$  (PLC $\gamma$ ) and STAT. FGF signaling primarily regulates growth factors involved in cell survival and differentiation.

### *SHH signaling*

Sonic Hedgehog signaling represents the main disease pathway of HPE (Mercier et al., 2013; Dubourg et al., 2016; Kruszka et al., 2018). The Hedgehog gene (*hh*) was initially identified by a genetic screen for mutations affecting segmental patterning in *Drosophila* (Nüsslein-Volhard et al., 1980). Further studies rapidly implicated *hh* as a major developmental factor highly conserved throughout evolution (Echelard et al., 1993; Krauss et al., 1993; Chang et al., 1994). In mammals, three homologues of the Hh gene have been identified: Sonic Hedgehog (*SHH*), Desert Hedgehog (*DHH*), and Indian Hedgehog (*IHH*). *SHH*, the best characterized mammalian Hh gene, is an essential regulator of vertebrate development implicated in patterning of the neural

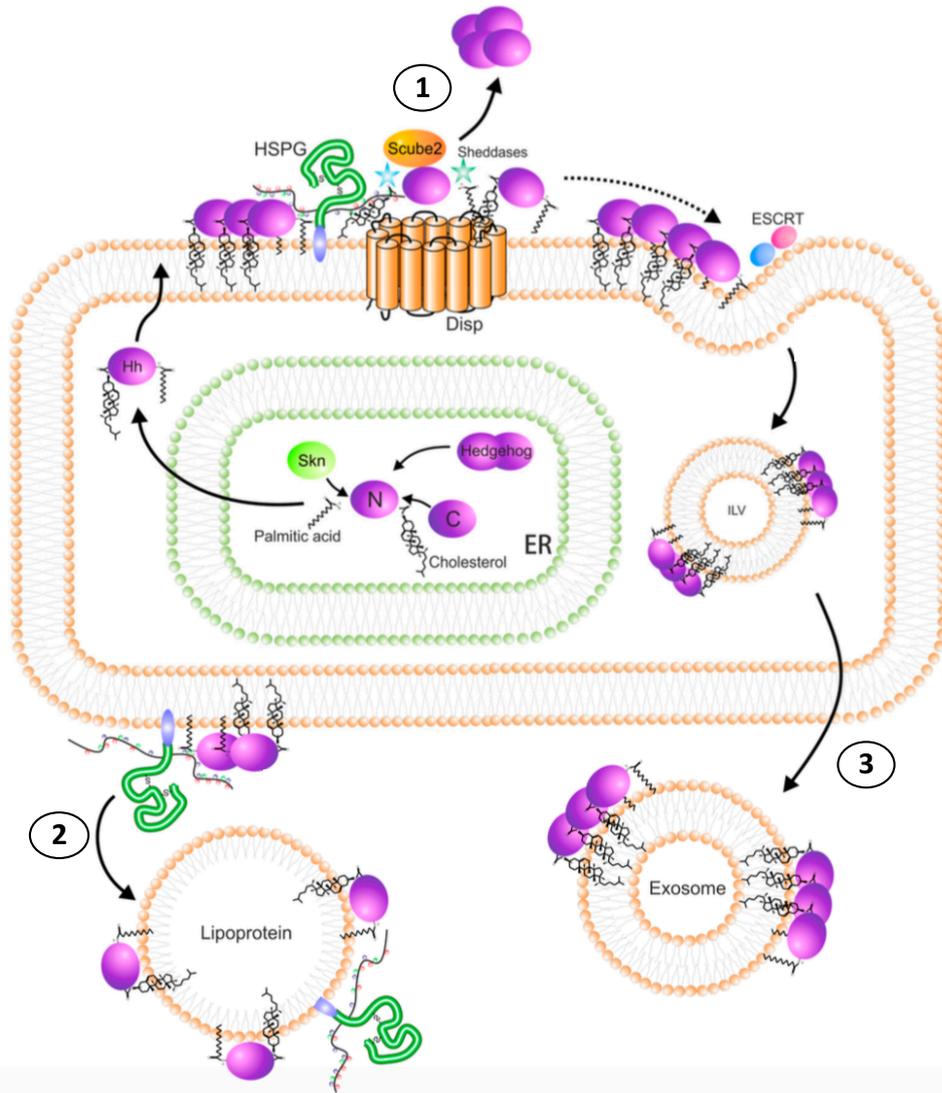
tube, forebrain and midline facial structures (Rubenstein and Beachy, 1998; Muenke and Beachy, 2000). As mentioned before, SHH signaling plays an essential role in forebrain development, as mice lacking the *Shh* gene exhibit severe HPE associated with cyclopia and absent ventral midline structures. Consistently, *SHH* is also the major gene implicated in HPE (Dubourg et al., 2016). Of note, *SHH* was named after “Sonic the Hedgehog”, a computer game of the early 90s. The game character, ‘Sonic’, appears to have two closely set eyes reminiscent of facial features encountered in HPE.

SHH functions as a morphogen with short- and long-range diffusion, capable of exerting its effect within up to 300  $\mu\text{m}$  distance from the secreting cell (Choudhry et al., 2014; Ramsbottom and Pownall, 2016). Initially synthesized as a 45-kDa precursor, SHH is then autoproteolytically cleaved to generate a 19 kDa N-terminus (SHH-N) and 26 kDa C-terminus (SHH-C) fragments (Lee et al., 1994; Dessaud et al., 2008) (Figure 14). The SHH-N fragment mediates the downstream signaling activity, while the SHH-C is necessary for the proper cleavage and maturation of the precursor (Chen et al., 2011). Specifically, SHH-C fragment contains a conserved catalytic cysteine (Cys198) and a sterol recognition (SR) motif. During maturation, the catalytic cysteine forms a thioester intermediate by attacking the polypeptide backbone while the SR motif binds a cholesterol molecule. The recruited cholesterol generates an ester linkage with the thioester and displaces the SHH-C fragment. This results in dissociation of the SHH-C fragment and attachment of the cholesterol to the SHH-N (Mann and Beachy, 2004). Subsequently, the SHH-C fragment is degraded in the endoplasmic reticulum while the SHH-N is further modified by attachment of palmitic acid, mediated by Hedgehog acyltransferase (*HHAT*, also called *Skn*). The cleavage and dual lipidation of SHH-N are essential for the correct functioning of the SHH signaling. Mutations blocking the autoproteolytic cleavage lead to loss of SHH activity, resulting in HPE (Roessler et al., 1997). Additionally, mice deficient in *HHAT* exhibit phenotypes similar to *Shh* knockout including lack of differentiated floor plate and neural tube defects (Dennis et al., 2012).



**Figure 14. Dual lipid modification of SHH protein.** Adapted from Dessaud et al., 2008

The dual lipidation of the SHH-N results in a highly hydrophobic molecule, which should complicate its release from the cell. Nevertheless, SHH has been shown to signal at a very long range, indicating that the lipidation does not restrict SHH-N to the cell but instead tightly controls its release and subsequent diffusion (Briscoe and Thérond, 2013) (Figure 15). The release of SHH-N from the cell is mainly maintained by transmembrane protein Dispatched (DISP1) and glycoprotein SCUBE2 in a cholesterol-dependent manner: DISP1 directly binds to the cholesterol part of SHH and acts in synergy with SCUBE2 to promote the release of SHH (Jakobs et al., 2016; Hall et al., 2019). The release of SHH also requires the action of Heparane Sulphate Proteoglycans (HSPG) which provide an assembly point for SHH/DISP1/SCUBE2 at the cell surface (Carrasco et al., 2005; Goetz, 2006; Ramsbottom and Pownall, 2016). Mature SHH-N peptides can also self-associate and form multimeric complexes at the cell surface. Multimeric SHH can then be released in lipoprotein particles via interaction with HSPGs; and in extracellular vesical particles (exovesicles) by endosomal sorting complexes required for transport (ESCRT) proteins (Ramsbottom and Pownall, 2016). The factors that determine the preferred type of SHH export remain unclear. Recent reports suggest that the availability of HSPG- and SCUBE2-regulated sheddases (membrane enzymes involved in cleavage of transmembrane proteins) plays a role in determining the type of SHH release (Jakobs et al., 2016).



**Figure 15. Multiple types of SHH release.** (1) SHH can be released under monomeric form via the combined action of SCUBE2, DISP1 (Disp) and HSPGs. (2) SHH can accumulate at the cell surface and form multimeric complexes, which are subsequently released as lipoprotein particles by interacting with HSPGs. (3) Unprocessed SHH can re-enter the cell and be internalized by ESCRT proteins, forming intra-luminal vesicles (ILV). ILV can subsequently fuse with the membrane and be released from the cell as exosomes. Adapted from Ramsbottom and Pownal, 2016.

Upon reaching its target cell, the SHH-N peptide will further activate the downstream signaling cascade which involves several transmembrane receptors and the primary cilium. SHH-N binds to several transmembrane proteins (Izzi et al., 2011). Patched1 (PTCH1) acts as the key receptor of SHH; CDON, BOC and GAS1 act as co-receptors enhancing the SHH/PTCH1 interaction. SHH/PTCH1 binding is necessary for the signal transduction. In absence of SHH-N ligand, PTCH1 blocks the pathway activity by inhibiting Smoothened (Smo). Binding of SHH-N to PTCH1 represses this inhibitory function, resulting in degradation of SHH-N/PTCH1 complex and activation of Smo (Figure 16).

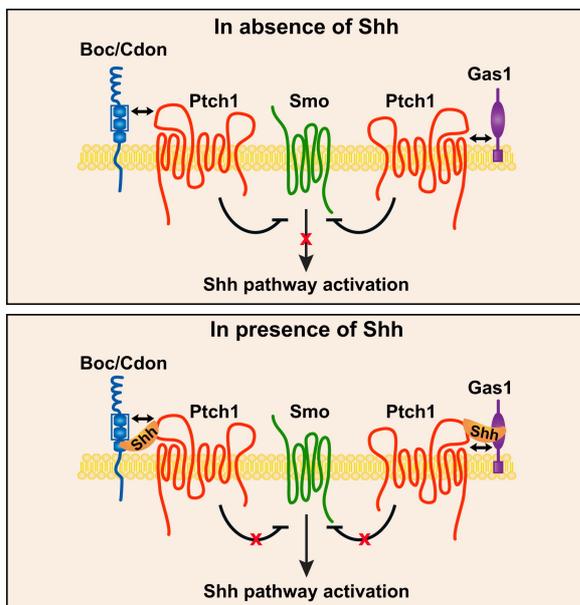
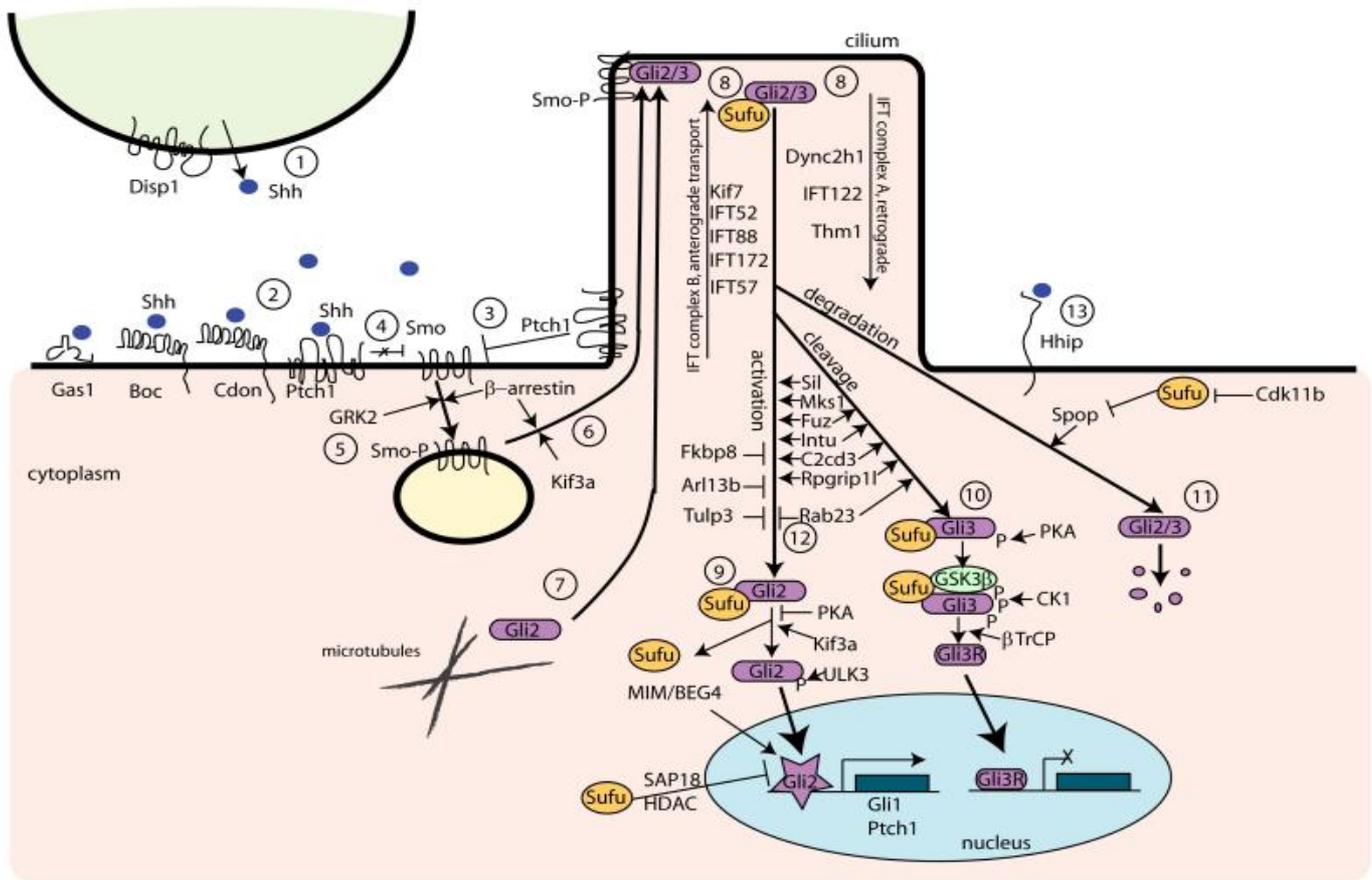


Figure 16. Reception of SHH signal. Izzi et al., 2011

Once activated, Smo transduces the SHH signal by translocating to the cilium and activating transcription factors of glioma-associated oncogene homolog family -GLI2 and GLI3 - which are the terminal effectors of the SHH signaling pathway (Rimkus et al., 2016). The precise mechanism of GLI2/3 activation remains unclear (Figure 17). In the absence of SHH signal, GLI2/3 proteins are bound to the *Suppressor of Fused* protein (SUFU) and remain in the cytoplasm. As a result, GLI2/GLI3 proteins undergo proteolytic cleavage to form transcriptional repressors (GLI2R/GLI3R). In the presence of SHH signal, the activated Smo enables the translocation of full-length GLI proteins

to the cilia thereby inhibiting their cleavage and formation of the repressor forms. Instead, GLI2/GLI3 proteins are transformed into transcriptional activators (GLI2A and GLI3A) and translocated to the nucleus to activate the target genes. Ultimately, SHH signaling activates the target genes by modulating the ratio between the repressor and activator forms of GLI2/GLI3 in the SHH-receiving cell (Jacob and Briscoe, 2003; Stamatakis et al., 2005).

SHH pathway is a complex molecular process with several regulatory nodes that can enhance or reduce the signal output. Along with other pathways described above, SHH forms a multi-layered molecular network regulating forebrain and facial development. Genetic alteration of one or several components of this network can lead to incorrect patterning of the forebrain and the onset of HPE. The complex spatio-temporal regulation and a large number of genetic factors involved in forebrain and facial development contribute to large genetic heterogeneity of this disease.



**Figure 17. Activation of GLI2/3 by SHH signaling.**

(1-4) SHH binds PTCH1 and represses its inhibitory action on SMO.

(5-7) Activated SMO is phosphorylated by GRK2 and assumes an active conformation by forming a complex with  $\beta$ -arrestin and Kif3a (Pak and Segal, 2016). The resulting Smo- $\beta$ -arrestin2-Kif3a relocates to the tip of the primary cilium and recruits SUFU-GLI2/GLI3 complex from the cytoplasm.

(8) Transport of SUFU-GLI2/GLI3 proteins into the primary cilium (anterograde transport) involves Intraflagellar Complex B (IFT-B) proteins. Once located at the tip of the primary cilium, GLI2/GLI3 are translocated back to the cytoplasm (retrograde transport) and dissociate from Sufu (Tukachinsky et al., 2010; Lin et al., 2014). Retrograde transport is assured by Intraflagellar Complex A (IFT-A) proteins.

(9) During the retrograde transport, GLI2/GLI3 proteins are transformed into transcriptional activators (GLI2A and GLI3A). This process requires phosphorylation by ULK3 kinase (Goruppi et al., 2017) and action of several ciliary proteins including Mks1, Fuz, Intu, C2cd3 and Rpgrip11 (Murdoch and Coop, 2010). After the transformation, GLI2A/GLI3A are translocated to the nucleus and activate the transcription of target genes.

(10-12) During the retrograde transport, GLI2/3 can also be degraded or cleaved into repressors (GLI2R/GLI3R). The SHH signal modulates the ratio between the repressor and activator forms of GLI2/GLI3 in the receiving cell (Stamatakis et al., 2005).

(13) Hedgehog-interacting protein (HHIP) can bind SHH ligand and repress the pathway activity (Chuang et al., 2003).

Adapted from Murdoch and Copp, 2013

### c. Genetic mechanisms of HPE

#### *Known disease genes*

##### *SHH*

In 1996, Roessler et al. analyzed 30 families presenting autosomal dominant HPE and identified 5 cases carrying heterozygous mutations in *SHH* (7q36) (Roessler et al., 1996). The identified mutations segregated with the disease and resulted in premature protein terminations or alterations of highly conserved residues. Together with previous reports indicating the implication of 7q36 region in the disease (Gurrieri et al., 1993; Muenke et al., 1994), this study identified *SHH* as the first known gene to cause HPE in human. To date, deleterious variants in *SHH* remain the most common cause of non-chromosomal cases of HPE (Dubourg et al., 2016). In a genetic analysis of 157 HPE families (396 individuals), Solomon et al. identified 141 unique *SHH* mutations (Solomon et al., 2012). A large proportion of identified variants were missense (66%), followed by nonsense (15 %) and frameshift (12 %) mutations. Several inframe deletions/insertions (~6%) and splicing mutations (1%) have also been reported.

##### *ZIC2*

*ZIC2* (*Zinc finger protein of the cerebellum 2*) is located at chromosome 13q32 and codes for a transcription factor playing several roles in regulating neurological development. Mice with complete absence of *Zic2* activity exhibit various phenotypes of the HPE spectrum (hypotelorism, cyclopia) indicating its role in craniofacial morphogenesis (Nagai et al., 2000). Similar to the case of *SHH*, *ZIC2* was first considered as HPE candidate due to recurrent deletions involving its genomic location (13q) identified in individuals presenting brain anomalies. Subsequent analyses of HPE patients identified pathogenic mutations in *ZIC2*, most of which led to the impairment of the *ZIC2* homeodomain involved in transcriptional control (Brown et al., 1998; Brown, 2001; Dubourg et al., 2004; Brown et al., 2005). For patients with normal karyotypes, mutations in *ZIC2* are found in approximately 5-8 % of cases (Solomon et al., 2009; Dubourg et al., 2016, Kruzka et al., 2018). Unlike other HPE genes, the vast majority of *ZIC2* HPE patients are sporadic cases and present with a specific phenotype

including large ears, flat nasal bridge and bitemporal narrowing (Solomon et al., 2010). The *ZIC2*-specific facial features are distinct from standard facial dysmorphisms encountered in isolated HPE, suggesting its unique role in forebrain and facial development (Dubourg et al., 2011; Mercier et al., 2011; Solomon et al., 2009).

### *SIX3*

Following the discovery of *SHH* as the first causal HPE gene, subsequent studies implicated its transcriptional regulator *SIX3* (2p21) in the disease. This transcription factor is the orthologue of the *sine oculis* (“without eyes”) gene in *Drosophila*, involved in forebrain and eye development in several organisms including mouse and zebrafish (Oliver et al., 1995; Seo et al., 1998). *SIX3* acts as a direct regulator of *SHH* expression in the ventral forebrain by binding to the *SHH* promoter region via its homeobox domain (Jeong et al., 2008). *SIX3* was first considered as HPE candidate in 1996 due to cytogenetic anomalies of 2p21 identified in HPE patients (Schell et al., 1996). A subsequent study identified 4 different mutations in the homeodomain of *SIX3* in HPE patients, thus implicating this gene in HPE (Wallis et al., 1999). *SIX3* is considered as the third most common gene implicated in isolated cases of HPE (~3 %, Dubourg et al., 2016).

### *TGIF1*

For a long time, *TGIF1* (*TGFβ-induced factor homeobox 1*) was thought to be the fourth major gene of HPE in terms of mutational frequency. Located at 18p11.3, *TGIF1* encodes a homeodomain protein with transcriptional repression activity regulating NODAL and retinoic acid signaling pathways (Powers et al., 2010). In 2000, a study by Gripp et al. implicated this gene in HPE by identifying four heterozygous *TGIF* mutations in HPE patients (Gripp et al. 2000). A case of recessive HPE caused by two compound heterozygous variants in *TGIF1* was also reported (El-Jaick et al., 2007). Mutations in this gene are thought to be responsible for 1-2 % of isolated cases of HPE, although recent reports suggest a much lower mutational frequency (Dubourg et al., 2016).

Since the discovery of *SHH*, *ZIC2*, *SIX3* and *TGIF* as causal genes in HPE, subsequent studies of the related signaling pathways have contributed to the identification of additional genetic factors implicated in the disease pathogenesis.

#### *Genes of the SHH pathway*

Most of known HPE genes belong to the SHH signaling pathway, described in part B. Mutations in *PTCH1*, the main receptor of SHH, have been identified in 10 patients presenting HPE and HPE-like phenotypes (Ming et al., 2002; Ribeiro et al., 2006). *PTCH1* acts as a repressor of SHH activity, as its partial inactivation was shown to restore normal SHH signaling and rescue the HPE phenotype in mice (Hong and Krauss, 2013). Consistently, *PTCH1* mutations found in HPE patients are thought to result in a gain-of-function effect, enhancing the *PTCH1* activity of repressing the SHH signaling (Ming et al., 2002). Duplications of *PTCH1* were also identified in several HPE patients (Derwinska et al., 2009). *GLI2*, the main transcriptional mediator of SHH signaling, can be considered as one of the major HPE genes in terms of mutational frequency (3.2%, Dubourg et al., 2016), although the reported variants cause a distinct phenotype that includes pituitary insufficiency and subtle facial features (Bear et al., 2014). Rare pathogenic mutations were also reported in SHH co-receptors enhancing the SHH-*PTCH1* interaction - *GAS1*, *CDON*, *BOC* (Pineda-Alvarez et al., 2012; Bae et al., 2011; Hong et al., 2017). *DISP1*, implicated in SHH release, was found mutated in patients presenting HPE microforms (Roessler et al., 2009). Several mutations of *SUFU*, long time HPE candidate implicated in the transduction of the SHH signal, were reported in screenings of large HPE cohorts (Dubourg et al., 2016).

#### *NODAL pathway*

In several cases, mutations in genes of the NODAL signaling pathway have been implicated in the disease. The implication of NODAL genes remains rare in liveborn HPE children, indicating that severe NODAL defects are likely to result in early embryonic lethality and not HPE (Roessler et al., 2008). Two genes of the NODAL pathway have been implicated in human HPE so far. *FOXH1*, a transcription factor implicated in the

formation of the prechordal plate and the notochord (Hoodless et al., 2001; Roessler et al., 2008); and *TDGF1 (CRIPTO)*, an essential factor of the midline development (de la Cruz et al., 2002).

#### *FGF pathway*

More recently, the FGF signalling pathway has been linked to HPE via pathogenic variants in *FGF8* and *FGFR1* (Arauz et al., 2010; Dubourg et al., 2016). In a recent analysis of 257 HPE patients, *FGF8* and *FGFR1* were found to be, respectively, the 5<sup>th</sup> and the 6<sup>th</sup> most commonly mutated genes, suggesting a major involvement of the FGF pathway in HPE pathogenesis (Dubourg et al., 2016). Inactivation of *Fgf8* in mice leads to abnormalities of the ventral telencephalon and pituitary dysfunction (McCabe et al., 2011). A study of a double mouse mutant for *Fgfr1/Fgfr2* receptors illustrated that FGF signalling acts downstream of *Shh* and *Gli3* to specify telencephalic cells (Gutin et al., 2006).

#### *New HPE genes*

Deciphering the genetic bases of HPE is an ongoing task with novel disease genes reported every year (Table 2). Recent findings report the implication of *STIL* and *CNOT1* genes in HPE. Previously associated with autosomal recessive microcephaly, *STIL* encodes a ciliary protein implicated in formation of the primary cilia, essential for transduction of SHH signal. Mutations in *STIL* were implicated in autosomal recessive cases of HPE with two consanguineous families reported so far (Mouden et al., 2015; Kakar et al., 2015). *CNOT1*, a member of CCR4-NOT complex involved in posttranscriptional regulation, was implicated in the disease by identifying an identical *de novo* missense mutation in two unrelated individuals affected by HPE (Kruszka et al., 2019). Recently identified disease genes also include *KMT2D*, *PPP1R12A*, *RAD21*, *SMC1A*, *SMC3* and *STAG2* but further investigations of large HPE cohorts are needed to further confirm these findings (Tekendo-Ngongang et al., 2019; Hughes et al., 2020; Kruszka et al., 2019). Despite numerous advances, all known HPE genes account for only a fraction of the genetic etiology in isolated HPE, with approximately 70 % of

patients remaining without an identified molecular cause of the disease (Dubourg et al., 2016). The low diagnostic yield of isolated HPE indicates the existence of unelucidated genetic factors underlying the disease pathogenesis.

Gene	Frequency of mutations	References
<i>SHH</i>	5.4%-5.9%	Roessler et al., 1996
		Nanni et al., 1999
<i>ZIC2</i>	4.8%-5.2%	Brown et al., 1998
		Brown et al., 2001
<i>SIX3</i>	~3%	Wallis et al., 1999
		Lacbawan et al., 2009
<i>GLI2</i>	~3.2%	Roessler et al., 2003 Dubourg et al., 2016
<i>TGIF1</i>	<1%	Gripp et al., 2000
		El-Jaick et al., 2007
<i>FGF8</i>	<2.2%	Arauz et al., 2010
		McCabe et al., 2011
<i>FGFR1</i>	~1.2%	Simonis et al., 2013
		Dubourg et al., 2016
<i>DISP1</i>	<1.2%	Roessler et al., 2009
		Mouden et al., 2016
<i>CNOT1</i>	<1.5%	Kruzka et al., 2009
<i>PTCH1</i>	Rare	Ming et al., 2002
		Ribeiro et al., 2006
<i>GAS1</i>	Rare	Pineda-Alvarez et al., 2012
<i>CDON</i>	Rare	Bae et al., 2011
<i>BOC</i>	Rare	Hong et al., 2017
<i>FOXH1</i>	Rare	Hoodless et al., 2001
		Roessler et al., 2008
<i>TDGF1</i>	Rare	de la Cruz et al., 2002
<i>STIL</i>	Rare	Mouden et al., 2015
		Kakar et al., 2015
<i>KMT2D</i>	Rare	Tekendo-Ngongang et al., 2019
<i>PPP1R12A</i>	Rare	Hughes et al., 2020
<i>RAD21</i>	Rare	Kruszka et al., 2019
<i>SMC1A</i>	Rare	Kruszka et al., 2019
<i>SMC3</i>	Rare	Kruszka et al., 2019
<i>STAG2</i>	Rare	Kruszka et al., 2019

**Table 2. Known HPE genes.** Estimated mutational frequencies have been retrieved from various sources detailed in Tekendo-Ngongang et al., 2000 (Updated Mar, 2020).

### *Inheritance*

In 1998, an epidemiologic study of 258 cases of isolated HPE concluded that the most compatible mode of disease transmission was autosomal dominant (Odent et al., 1998). The disease penetrance was estimated at 82 % for major HPE subtypes and 88 % when including minor forms. The autosomal dominant inheritance was further supported by a high rate of observed sporadic cases (68 %) caused by *de novo* mutations. However, subsequent studies challenged this notion. In 2011, a genetic screening of 645 HPE probands revealed that mutations in major HPE genes are inherited from asymptomatic or mildly affected parents in most cases (70% for *SHH* and *SIX3*), illustrating an extremely high incomplete penetrance and variable expressivity of pathogenic mutations (Mercier et al., 2011).

Several cases of autosomal recessive HPE have also been reported. The first recessive case was described in 2007 involving two compound heterozygous variants in *TGIF1*, but was interpreted as a rare coincidence since one of the two identified variants was found to be functionally normal in its activity (El-Jaick et al., 2007). In 2011, a pathogenic homozygous mutation in *FGF8* was reported in a consanguineous patient presenting semilobar HPE (McCabe et al., 2011). Together with reports of recessive HPE caused by mutations in *STIL* and *DISP1* (Mouden et al., 2015, 2016), these findings indicate that recessive inheritance may account for a certain part of isolated HPE cases.

Digenic inheritance of HPE, involving two mutations in two distinct HPE genes, was also observed (Table 3). In 1999, Nanni et al., reported 3 families presenting a mutation in *SHH* associated with variants in another disease gene (*ZIC2*, *TGIF1*), suggesting that the HPE phenotype may result from interactions of multiple gene products (Nanni et al., 1999). Subsequent studies of major HPE genes described isolated cases with mutations in two HPE genes (Ming & Muenke, 2002) and several cases of chromosomal rearrangements associated with a mutation in a known HPE gene (Mercier et al., 2011).

Gene	Mutation	References
<i>SHH</i>	p.Gly290Asp	Nanni et al, 1999
<i>ZIC2</i>	c.1377-1406dup30	
<i>SHH</i>	p.Pro424Ala	Nanni et al, 1999
<i>TGIF1</i>	del18p11	
<i>SHH</i>	Del378-380	Nanni et al, 1999
<i>TGIF1</i>	p.Thr151Ala	
<i>GLI2</i>	p.Arg151Gly	Rahimov et al, 2006
<i>PTCH1</i>	p.Thr328Ala	
<i>SIX3</i>	p.Ala93Asp	Ming et al, 2002
<i>PTCH1</i>	p.Ala393Thr	
<i>SIX3</i>	p.Ala284Pro	Lacbawan et al, 2009
<i>ZIC2</i>	p.Trp304Arg	
<i>GAS1</i>	p.Asp270Tyr	Ribeiro et al, 2010
<i>SHH</i>	p.Leu218Pro	

**Table 3. Digenic cases of HPE**

Studies of mouse HPE mutants further underline the complex genetic mechanisms involved in the disease pathogenesis. While homozygous mice mutants for *CDON* (*Cdo*<sup>-/-</sup>) exhibit a mild HPE phenotype, a heterozygous inactivation of *SHH* in this mutant background (*Cdo*<sup>-/-</sup>; *SHH*<sup>+/-</sup>) dramatically exacerbates the phenotype leading to severe HPE-like defects including proboscis, suggesting a genetic interaction between *CDON* and *SHH* to promote disease pathogenesis (Tenzen *et al.*, 2006). In NODAL signaling pathway, double inactivation of *Nodal* and *Hnf3b* results in severe HPE phenotype (cyclopia), which is absent in case of *Nodal/Hnf3b* single gene inactivation (Varlet *et al.*, 1997). The same pattern is observed in case of double inactivation of *Chd* and *Nog*, resulting in proboscis, cyclopia and agnathia (Bachiller *et al.*, 2000). Double inactivation of *Fgfr1;Fgfr2* and *Bmp1ra;Bmp1rb* of, respectively, FGF and BMP pathways, lead to HPE-like phenotypes (Gutin *et al.*, 2006; Fernandes *et al.*, 2007). Mouse studies also underline the importance of genetic background in the development of HPE phenotype. *Cdo*<sup>-/-</sup> mice develop semilobar HPE on a pure C57BL/6 genetic background but exhibit only a microform phenotype on a mixed genetic background (Helms *et al.*, 2005; Zhang *et al.*, 2006). Modifying factors due to the genetic background seem insufficient to induce an HPE phenotype but can impact on the regulation of the involved signalling pathways and represent important risk factors.

## d. Non-mendelian HPE - Article 1.

---

### Recent advances in understanding the inheritance of holoprosencephaly

Christèle Dubourg, Artem Kim, Erwan Watrin, Marie de Tayrac, Sylvie Odent, Véronique David<sup>1</sup>,  
Valérie Dupé

1 - corresponding author: [veronique.david@univ-rennes1.fr](mailto:veronique.david@univ-rennes1.fr)

American Journal of Medical Genetics, May 2018

---

The increasing number of causative genes with varying inheritance, the incomplete penetrance and variable expressivity of disease-causing mutations, the importance of genetic background as well as complex molecular basis underlying forebrain and facial development have led to the idea of HPE being a complex genetic disorder with non-Mendelian inheritance resulting from accumulation of several variants in functionally-connected genes. Under this model, a single pathogenic mutation can be necessary but not sufficient for the disease pathogenesis. Additional variants can either modulate the clinical phenotype induced by the pathogenic mutation (modifier effect) or can be required for the disease onset (oligogenic inheritance). Such additional variants can be hypomorphic (*i.e.*, of small effect) and be present in asymptomatic/less affected relatives of the HPE probands. Incomplete penetrance and variable expressivity of pathogenic mutations observed in HPE can be, therefore, explained by the presence of additional variants required for the disease manifestation. As such mutations are likely to be overlooked in genetic studies, the non-Mendelian inheritance model accounts for the significant part of cases for which no genetic etiology could be established by conventional approaches. Given the potential modifier effect, the non-Mendelian inheritance pattern can also explain the wide variability of clinical features observed in HPE.



## RESEARCH REVIEW

# Recent advances in understanding inheritance of holoprosencephaly

Christèle Dubourg<sup>1,2</sup> | Artem Kim<sup>1</sup> | Erwan Watrin<sup>1</sup> | Marie de Tayrac<sup>1,2</sup> |  
Sylvie Odent<sup>1,3</sup> | Véronique David<sup>1,2</sup>  | Valérie Dupé<sup>1</sup>

<sup>1</sup>Univ Rennes, CNRS, IGDR (Institut de génétique et développement de Rennes) - UMR 6290, F - 35000, Rennes, France

<sup>2</sup>Service de Génétique Moléculaire et Génomique, CHU, Rennes, France

<sup>3</sup>Service de Génétique Clinique, CHU, Rennes, France

**Correspondence**

Véronique David, Univ Rennes, CNRS, IGDR (Institut de génétique et développement de Rennes), UMR 6290, Rennes F-35000, France. Email: veronique.david@univ-rennes1.fr

Holoprosencephaly (HPE) is a complex genetic disorder of the developing forebrain characterized by high phenotypic and genetic heterogeneity. HPE was initially defined as an autosomal dominant disease, but recent research has shown that its mode of transmission is more complex. The past decade has witnessed rapid development of novel genetic technologies and significant progresses in clinical studies of HPE. In this review, we recapitulate genetic epidemiological studies of the largest European HPE cohort and summarize the novel genetic discoveries of HPE based on recently developed diagnostic methods. Our main purpose is to present different inheritance patterns that exist for HPE with a particular emphasis on oligogenic inheritance and its implications in genetic counseling.

**KEYWORDS**

complex disorder, holoprosencephaly, oligogenic inheritance, Sonic hedgehog

## 1 | INTRODUCTION

Holoprosencephaly (HPE) is a severe developmental disorder classically defined as incomplete cleavage of the forebrain that originates from failed midline delineation during early development. There are several degrees of severity defined by the extent of the brain malformations. For the most severe cases, malformations are divided into alobar, semilobar, or lobar forms. These brain abnormalities are associated with facial anomalies that are also of varying severity ranging from cyclopia to milder signs such as ocular hypotelorism. The full spectrum of HPE also includes microforms characterized by facial midline defects (e.g., single median incisor) without brain malformations typical of HPE (Cohen, 2006; Dubourg et al., 2007; Hahn, Barnes, Clegg, & Stashinko, 2010; Muenke & Beachy, 2000).

HPE occurs in most ethnic groups worldwide. Although implication of maternal diabetes in HPE has been reported (Barr et al., 1983), the well-established origin of HPE remains almost exclusively genetic and consists of chromosomal abnormalities and nucleotide-based variants (Dubourg et al., 2007, 2016; Solomon et al., 2012). So far, 17 genes have been implicated in HPE, all of which encode proteins belonging to brain development pathways. *Sonic Hedgehog* (*SHH*) was the first discovered HPE gene (Roessler et al., 1996) and its alterations remain the most common cause of nonchromosomal HPE (Dubourg et al., 2016). *SHH* has been extensively studied and its functions during early brain development are now well described. A morphogenetic gradient of

*SHH* is established from the ventral midline of the diencephalon to induce appropriate cleavage of both forebrain and eye field. Remarkably, all HPE genes described so far are involved in the regulation of *SHH* activity (Sun et al., 2014; Xavier et al., 2016).

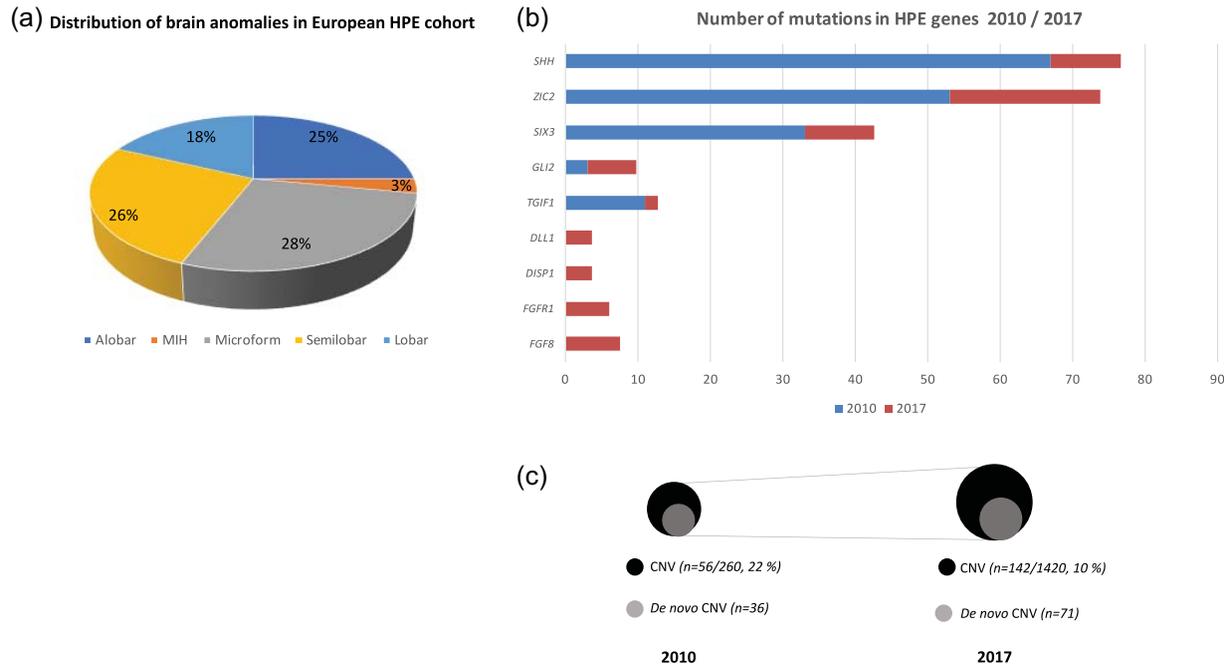
Initially described as an autosomal dominant trait with incomplete penetrance and variable expressivity, the mode of inheritance of HPE has been progressively redefined. The apparent autosomal dominant transmission with incomplete penetrance observed in a few HPE families may well be due to the cumulative effects of rare variants in two genes or more. Undeniably, the prevalence of oligogenicity has increased for several developmentalcc pathologies since next generation sequencing (NGS) technologies became accessible (Bamshad et al., 2011). Despite technical advances, defining the causative gene for HPE remains a difficult task, and even when one underlying variant is known, prenatal prediction remains uncertain.

Here, we describe clinical features and inheritance aspects of this disease with examples from our experience.

## 2 | CLINICAL AND GENETICS FEATURES OF THE COHORT

### 2.1 | The European HPE cohort

In 1996, a European HPE network was established in Rennes, France. Patients are recruited by clinicians from the different French centers of



**FIGURE 1** Clinical and molecular details of the European HPE cohort. (a) Brain anomalies distribution in European HPE cohort (1,420 probands). (b) Mutational spectrum of HPE. In blue, number of variants found in 2010. About 164/642 (25.4%) patients were found to harbor variants in *SHH*, *ZIC2*, *SIX3*, or *TGIF1*. *GLI2* variants were identified in 3/208 patients (Mercier et al., 2011). In red, number of variants found in 2017. Molecular screening of major HPE genes in 1,420 patients revealed 207 variants in *SHH*, *ZIC2*, *SIX3*, and *TGIF1*. Complementary screenings on a series of 302 patients revealed 32 variants in *DLL1*, *FGF8*, *FGFR1*, *DISP1*, and *GLI2*. (c) Contribution of CGH-array to molecular diagnosis. In 2010, CNVs were detected in 22% of the 260 patients analyzed by CGH-array, including 36 occurring de novo. In 2017, the screening of our entire cohort (1,420 patients) reported CNVs in 142 patients (10%), including 71 occurring de novo

reference for rare developmental diseases as well as from several European clinical centers in the United Kingdom, Belgium, Italy, Spain, and Portugal. Half of collected samples were fetuses, which enriched our cohort for severe HPE phenotypes (Mercier et al., 2011). Over the years, we have collected over 2,700 blood deoxyribonucleic acid (DNA) samples or frozen fetal tissues, including patients and relatives, and gathered clinical data and DNA for 1,420 HPE probands. Our cohort contains both apparently sporadic and familial cases, excluding those associated with other malformation syndromes or with known chromosomal abnormalities that could be revealed by standard cytogenetic analysis (e.g., trisomies 18 and 13).

HPE presents a wide continuous spectrum of clinical malformations ranging from severe to milder forms. Our cohort is representative of the full clinical spectrum of HPE phenotypes (Mercier et al., 2011). Within this cohort, the most severe brain malformations are categorized into alobar, semilobar, or lobar form. Middle interhemispheric fissure (syntelencephaly)—incomplete separation of the posterior frontal and parietal regions—also belongs to the HPE spectrum (Figure 1). These brain abnormalities are significantly correlated with a variety of distinct facial anomalies ranging from cyclopia, the most severe form, to milder signs such as ocular hypotelorism. Severe facial phenotypes, such as cyclopia, ethmocephaly (proboscis), and cebocephaly are more highly associated with alobar HPE. Similarly, premaxillary agenesis, cleft lip or palate and milder ocular abnormalities (coloboma, retinal

dysplasia) are mostly associated with semilobar HPE while the mild midface malformations, such as pyriform sinus stenosis and choanal stenosis are mostly found in lobar HPE. Patients who present the mildest facial abnormalities, such as hypotelorism, cleft lip, or single median incisor, generally do not present easily detectable brain malformations (Mercier et al., 2011). These HPE microforms can nonetheless be associated with microcephaly and intellectual disability and their molecular diagnosis is therefore important for proper patient care (Bruel et al., 2017; Solomon et al., 2010). Notably, most of these microforms have been diagnosed only because they were relatives of patients with severe HPE. Some families manifest a wide range of phenotypes, from typical alobar HPE with perinatal lethality to microforms such as microcephaly, hypotelorism, or both (Mercier et al., 2011). These different observations and clinical correlations made on a European cohort can be extended to all HPE cohorts, as they are very similar to those of North American patients (Lacbawan et al., 2009; Solomon et al., 2010).

As the frequency of microform HPE is underestimated, we are currently expanding our diagnostic approach to mildly affected relatives of a classical HPE patient. When a typical HPE patient is diagnosed in a family, we routinely perform a careful examination of all family members including neuroimaging techniques (MRI) and determination of clinical features that are not traditionally considered as a part of the HPE spectrum. Our goal is to expand our cohort to a larger number of HPE microforms to ensure that we cover the entire HPE spectrum.

TABLE 1 Summary of chromosomal abnormalities in HPE cases

Patient	Cytoband	Type of CNV	Start-end (GRCh37)	CNV size	Inheritance
1	1q43q44 3p25.2p22.1	Del Dup	242094954_249212668 0_39848444	7 Mb 40 Mb	Inherited from parental balanced t(1;3)
2	2p15	Del	61668439_61777447	109 kb	De novo
3	2p11.2 16p13.11 <sup>a</sup>	Dup Dup	85824180_86469217 15492317_16276115	645 kb 784 kb	De novo Inherited from father
4	3p22;20q11.2	Balanced translocation	-	-	Inherited from mother
5	3q25.32	Del	157105931_157154842	50 kb	Inherited from mother
6	4q12	Del	55193357_58196685	3,00 Mb	ND
7	5q35.3	Del	177296851_178323802	1 Mb	Inherited from mother
8	5q35.3	Dup	178038828_179766520	1.73 Mb	Inherited
9	7p22.1	Del	5399371_6871084	1.4 Mb	De novo
10	7p22.1	Dup	5057686_5166175	108 kb	Inherited from father
11	8q23.3q24.11	Del	117641330_118051191	410 kb	ND
12	Whole chromosome 8	Dup in mosaic	0_146294098	146 Mb	De novo
13	10p15.3p14	Del	136361_9946915	9.8 Mb	De novo
14	11q13.4	Dup	74931870_75109882	178 kb	Inherited from mother
15	12q21.32	Del	86984993_87656628	672 kb	ND
16	14q23.1	Del	60950490_61006021	55 kb	Inherited from mother
17	14q23.1	Dup	Unavailable data	900kb	Inherited from mother
18	15q11.2 <sup>a</sup>	Del	22765628_23208901	443 kb	Inherited from mother
19	16p13.11 <sup>a</sup> 16p11.2 <sup>a</sup>	Del Dup	15492317_16267306 29652999_30197341	1.3 Mb 544 kb	ND Inherited from mother
20	18q22.1	Dup	64152648_64324336	172 kb	Inherited from mother
21	19q13.42q13.43	Dup	56228025_57744093	1.5 Mb	De novo
22	22q11.21 <sup>a</sup>	Del	20719112_21464119	745 kb	ND
23	22q11.21 <sup>a</sup>	Del	21081060_21505558	424 kb	De novo
24	22q11.21 <sup>a</sup>	Del	18894835_21464119	2.6 Mb	De novo
25	22q11.21 <sup>a</sup>	Del	18894835_21464119	2.6 Mb	De novo
26	22q11.21 <sup>a</sup>	Del	18651614_21801661	3.2 Mb	ND
27	Xq25q28	Del	123176394_152515593	29.34 Mb	De novo

Del = deletion; Dup = duplication; ND = not determined.

## 2.2 | The evolution of genetic strategies for HPE diagnosis

### 2.2.1 | Sanger sequencing and detection of microdeletions in the major HPE genes

From the discovery of the first genes responsible for HPE until recently, genetic analysis of HPE patients has mainly relied on a Sanger sequencing approach (Mercier et al., 2011). During that period (1997–2010), screening our cohort for nucleotide-based variants in the four HPE genes—*SHH*, *Zinc Finger Protein 2 (ZIC2)*, *Six Homeobox 3 (SIX3)*, and the Homeobox protein *TGIF1*—provided a global variant detection rate of 20% (8.2% for *SHH*, 7.4% for *ZIC2*, 3.9% for *SIX3*, and 1.1% for *TGIF1*)

(Mercier et al., 2011). All variants in these genes were detected heterozygously, and were shown to be loss-of-function variants (Roessler, El-Jaick, et al., 2009; Roessler, Lacbawan, et al., 2009). These genes have constituted the four major genes of HPE. Meanwhile, novel genes have been implicated in sporadic cases of HPE (Roessler, El-Jaick, et al., 2009). The implication of each gene represents less than 1% of HPE patients and they are therefore referred to as minor genes.

In 2003, variants in *GLI2*, one crucial effector of *SHH* signaling pathway (Ruiz i Altaba, Palma, & Dahmane, 2002), had been described in HPE patients (Roessler et al., 2003), and screening of our cohort (Figure 1) revealed *GLI2* variants in 3.2% of the 302 patients tested (Mercier et al., 2011), thus placing *GLI2* as a major gene.

Identification of variants combined with detailed clinical assessment of HPE patients has allowed establishing some genotype–phenotype correlations (Mercier et al., 2011; Solomon et al., 2010). The most severe types of HPE (alobar and semilobar) tend to be associated with *ZIC2* and *SIX3* alterations while *SHH* tends to be more frequently associated with microforms. Remarkably, in *SHH* and *SIX3* cases, the facial dysmorphism is associated with brain anomalies while the probands of the *ZIC2* group tend to have a combination of severe HPE with few of the facial features. *GLI2* variants were preferentially found in patients presenting HPE microforms together with secondary specific features such as pituitary anomalies (Bear et al., 2014). These genotype–phenotype correlations have contributed to facilitate molecular analysis and genetic counseling for HPE.

Initially, Sanger sequencing has successfully allowed establishing molecular diagnoses for about 20% of HPE patients present in our cohort (Mercier et al., 2011). In order to explain, at least part of the remaining 80% of unsolved cases, we have searched for microdeletions in the major HPE genes (*SHH*, *ZIC2*, *SIX3*, and *TGIF1*) first by quantitative multiplex PCR of short fluorescent fragments and then by multiplex ligation-dependent probe amplification. Deletions in HPE genes were thereby shown to be a common cause of HPE in up to 8.5% of fetuses and in 5% in our whole cohort (Bendavid et al., 2006).

In 2006, the combination of these different approaches allowed us to diagnose molecularly 25% of our patients, very similar to the success rate of our American colleagues (Roessler et al., 2009; Solomon et al., 2010, 2012). As 75% of cases remained unsolved, we and others in the field have considered alternative genetic causes of HPE.

## 2.2.2 | Chromosomal abnormalities and copy number variants in HPE

Since 2006, a pangenomic technique named comparative genomic hybridization (CGH) array has been used to screen the entire genome for copy number variations (CNVs). The first study, we carried out during 2006–2009 revealed an impressively high rate of chromosomal rearrangements in HPE patients (22%), of which 14% occurred de novo and 8% were inherited (Bendavid et al., 2009, 2010). Furthermore, the observation of these CNVs can also lead to the detection of parental-balanced translocations and can subsequently improve prenatal diagnosis in such families.

In addition, these CNVs involved novel potential HPE loci. Despite relatively low recurrence rates of CNVs, overlapping 6qter region deletions among four unrelated patients allowed identification of a ligand of the NOTCH signaling pathway, Delta-like 1 (*DLL1*) as candidate gene (Dupé et al., 2011). Subsequent detection of a nucleotide-based variant in a distinct patient provided further evidence for *DLL1* as HPE gene, and, together with expression and functional studies in vertebrates (Ratié et al., 2013; Ware, Hamdi-Rozé, & Dupé, 2014), allowed defining Notch as a novel signaling pathway involved in HPE.

In addition, microrearrangements found in unique cases, pointed to candidate genes such as *SIX6* and *OTX2*, which are both implicated in early brain development (Jean, Bernier, & Gruss, 1999; Jin, Harpal, Ang, & Rossant, 2001). Another study using a similar approach has also highlighted *OTX2* as a candidate gene (Rosenfeld et al., 2010).

Since this work, CGH-array has been and still is part of our systematic molecular screening of HPE patients. The novel detected CNVs are summarized in Table 1. The proportion of CNVs observed in this second period (2010–2017) is reduced as compared to that of our previous study (Bendavid et al., 2009). The rate of disease-relevant CNVs is now shown to be 10%, half of which being de novo. Several of these rearrangements are recurrently observed in cases of intellectual disability, such as 2.6 Mb-microdeletion of 22q11.21 (proximal deletion) corresponding to DiGeorge syndrome (Burnside, 2015), 16p11.2 microduplication that confers susceptibility to autism (Fernandez et al., 2010), 16p13.11 encompassing the *NDE1* gene involved in brain neurogenesis and rhombencephalosynapsis (Bakircioglu et al., 2011; Démurger et al., 2013) and 15q11.2 microdeletion emerging as one of the most common cytogenetic abnormalities in intellectual disability and autism spectrum disorder (Butler, 2017). These deletions and duplications are thus at the origin of other neurodevelopmental disorders but are not sufficient to fully explain HPE. Nevertheless, CNV detection has increased the diagnostic yield from 25% to 35% in our cohort.

## 2.2.3 | NGS methods and their use for HPE diagnosis

The discovery of *SHH* in 1996 was followed by that of other genes—*ZIC2*, *SIX3*, *TGIF1*, and *GLI2*. Since then, subsequent studies of the pathways implicating these major genes have contributed to the identification of additional HPE genes (Table 2). Variants in genes involved in the SHH signaling pathway—*PTCH1*, *DISP1*, *CDON*, *GAS1*, *BOC*, and *SUFU*—were described in some HPE patients (Bae et al.,

TABLE 2 List of HPE genes and corresponding percentages of variants found in our HPE cohort

Chromosome	Gene	NM	%
7	<i>SHH</i>	000193.2	5.4
13	<i>ZIC2</i>	007129.3	5.2
2	<i>GLI2</i>	005270.4	3.2
2	<i>SIX3</i>	005413.3	3.0
10	<i>FGF8</i>	033163.3	2.5
8	<i>FGFR1</i>	023110.2	2.0
1	<i>DISP1</i>	032890.3	1.2
6	<i>DLL1</i>	005618.3	1.2
18	<i>TGIF1</i>	170695.2	0.9
10	<i>SUFU</i>	016169.3	0.4
1	<i>STIL</i>	001048166.1	1 case/375
9	<i>GAS1</i>	002048.2	0
3	<i>TDGF1</i>	003212.3	0
11	<i>CDON</i>	016952.4	0
8	<i>FOXH1</i>	003923.2	0
10	<i>NODAL</i>	018055.4	0
3	<i>BOC</i>	001301861.1	Not tested

<sup>a</sup>Copy number variation (CNV) usually observed in intellectual deficiency.

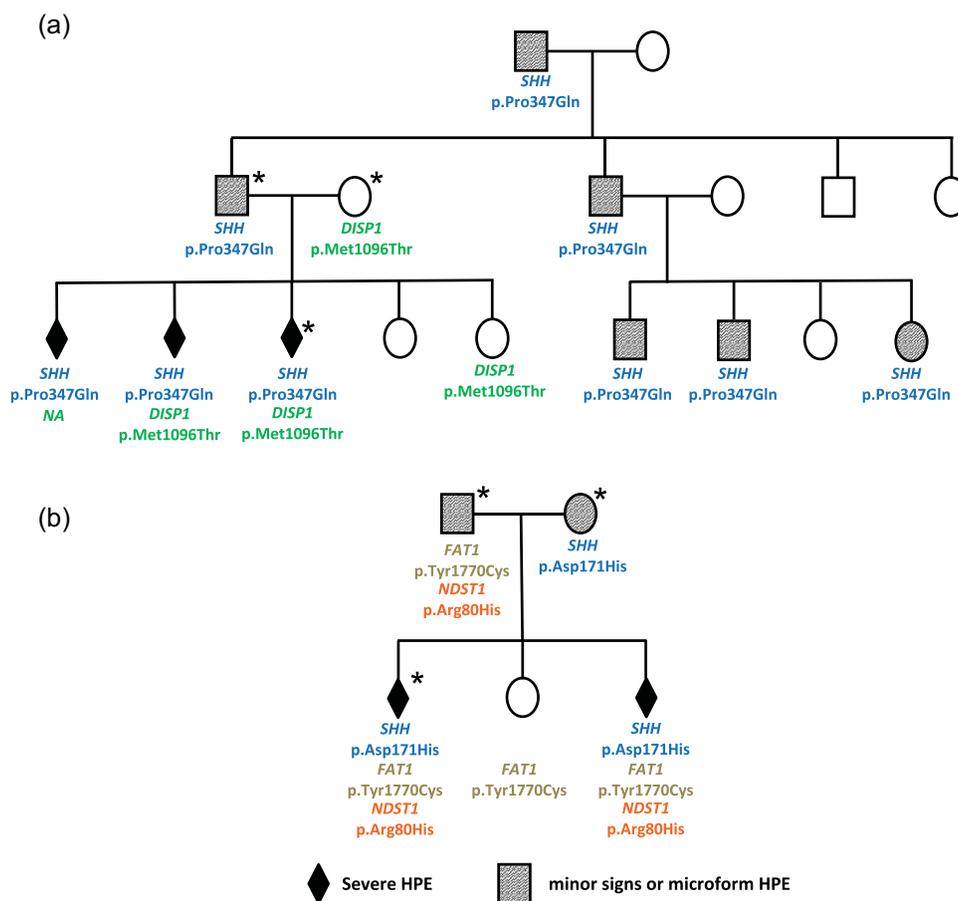
2011; Dubourg et al., 2016; M. Hong et al., 2017; Ming et al., 2002; Pineda-Alvarez et al., 2012; Roessler, Ma, et al., 2009). A few variants have been described in *NODAL*, *TDGF1*, *FOXH1*, which encode proteins belonging to the Nodal/TGF-beta pathway (de la Cruz et al., 2002; Roessler, Pei, et al., 2009; Roessler et al., 2008). The fibroblast growth factor pathway has also been implicated in HPE through variants in *FGF8* and its receptor *FGFR1* (Arauz et al., 2010; Simonis et al., 2013). More recently, variants in *STIL*, a gene implicated in the formation of the primary cilia, were also described in HPE families (Kakar et al., 2015; Mouden et al., 2015). These are minor genes as they are reported in less than 1% of HPE cases. Importantly, all these genes have in common the ability to affect SHH activity (Sun et al., 2014; Xavier et al., 2016).

Taking advantage of NGS in 2010, we established the first gene-panel sequencing method that targeted all known HPE genes (Dubourg et al., 2016). At the time we are writing this manuscript, more than 300 HPE patients have been tested. This study revealed that *SHH*, *ZIC2*, *SIX3*, and *GLI2* retain their position of major genes and *TGIF1* is relegated to the minor gene group. Furthermore, the identification of numerous variants in *FGF8* and *FGFR1* strengthens the involvement of FGF signaling in HPE (Figure 1b). Recent functional analysis in

Zebrafish has confirmed the contribution of *FGF8* variants in HPE (Hong, Hu, Roessler, Hu, & Muenke, 2018; Hong et al., 2016) in accordance with the known function of FGF signaling during the specification of the dorso-ventral axis of the forebrain (Storm et al., 2006).

From the results we obtained over the years from our 1,420 probands, the 10 first-ranked genes involved are *SHH* (5.4%), *ZIC2* (5.2%), *GLI2* (3.2%), *SIX3* (3.0%), *FGF8* (2.5%), *FGFR1* (2.0%), *DISP1* (1.2%), *DLL1* (1.2%), and *TGIF1* (0.9%) (Figure 2; Table 2). Some rare deleterious variants have been found in *SUFU*, a regulator of the Sonic-hedgehog-signaling pathway (Dubourg et al., 2016). By contrast, we did not detect any pathogenic variants in the following minor genes: *PTCH1*, *NODAL*, *GAS1*, *TDGF1*, *CDON*, and *FOXH1*. To date, only a small number of variants have been reported for these genes (Bae et al., 2011; de la Cruz et al., 2002; Ming et al., 2002; Roessler, Pei, et al., 2009; Roessler et al., 2008). Therefore, more data need to be collected to unambiguously assign these genes to HPE etiology.

As NGS technology evolved and became more accessible, whole exome sequencing (WES) is now routinely used to investigate novel HPE patients (see paragraph below). We expect this new approach will allow us not only to increase the diagnostic yield but also to identify novel HPE genes. Along the same lines, we are now establishing whole



**FIGURE 2** Examples of complex inheritance in HPE. (a) Family presenting a digenic mode of transmission associating variants in *SHH* and *DISP1* (Mouden et al., 2016). Minor signs refer to microcephaly. (b) Family presenting an oligogenic pattern with combined inherited variants in *SHH*, *FAT1*, and *NDST1*. Minor signs refer to hypotelorism (father) and epicanthus (mother). Individuals marked with asterisk were analyzed by whole exome sequencing. NA: not available for *DISP1* sequencing

genome sequencing approach in order to identify other alterations located in the noncoding part of the genome (~98%) that has remained largely unexplored until now in the context of HPE.

An important observation has emerged from our experience as clinical reference center for HPE. When comparing the percentage of positive diagnoses between 2010 and 2017 it appears surprisingly stable (about 35%), despite the fact that sequencing technology has improved and that the number of HPE genes has increased. A closer examination of our data indicates that the fraction of patients with a CNV or with a variant in one of the major genes (*SHH*, *ZIC2*, and *SIX3*) has been reduced by half as compared to that of 2010. Paradoxically, this reduction of cases is due to the increase of knowledge in diagnosis of HPE. Indeed, more and more clinical genetic centers are now able to perform their own molecular diagnoses for HPE. When a deleterious variant is found in one of the major HPE genes or when there is a CNV, the patient is not systematically referred to the reference center (i.e., Rennes, France) anymore. In contrast, when a molecular diagnosis could not be established, the patient data are sent to us for further investigation. As a consequence, the proportion of cases with an alteration in the major HPE genes has decreased in our cohort, which explains why our success rate of molecular diagnosis has remained stable over the years.

### 3 | MODES OF INHERITANCE IN HPE

In 1996, we initiated an epidemiologic study on 258 HPE cases and concluded that for nonsyndromic and nonchromosomal HPE, the most compatible mode of transmission was autosomal dominant with incomplete penetrance and a variable expressivity (Odent, Le Marec, Munnich, Le Merrer, & Bonaïti-Pelli  , 1998).

In HPE, the analysis of the first four major genes revealed that in most of the cases (70% for *SHH* and *SIX3*) the variants were inherited from a parent who was asymptomatic or only mildly affected (microform HPE). By classic textbook definitions, autosomal dominant inheritance is defined as the transmission of disease from an affected parent to an affected offspring. In this model, half of this parent's offspring is expected to be affected. Our observations show that this model (and

ratios of affected offspring) does not appear to apply to HPE. Furthermore, although as many as 17 genes have been linked to HPE, variants in these genes collectively explain only 25% of all HPE cases. It suggests that HPE is a complex disease with an increasing number of causative genes for which inheritance can vary depending of the affected gene as well as other related factors. This observation stresses the need for clarifying potential modes of its inheritance. In this section, we present different cases from our cohort and discuss their corresponding inheritance pattern.

#### 3.1 | De novo versus inherited variants in HPE

Systematic sequencing of the major HPE genes has shown a high proportion of de novo variants in *ZIC2* (70%), *SHH* (30%), and *SIX3* (30%) (Figure 3; Mercier et al., 2011). Thus, de novo variants are implicated in numerous sporadic HPE cases. In accordance with the essential role of these genes in early brain development, these de novo variants are loss-of-function and tend to be more deleterious than inherited ones (Roessler et al., 2009).

In our cohort, 50% of *FGFR1* variants appeared de novo. Adding to complexity, *FGFR1* is prone to mosaicism as we could show in one family in which a patient's father presents an attenuated HPE phenotype (Dubourg et al., 2016). Interestingly, cases of mosaicism involving *FGFR1* were described in other diseases like Hartsfield syndrome and encephalocraniocutaneous lipomatosis (Bennett et al., 2016; Dhamija et al., 2014). It should be noted that mosaic variants could be overlooked depending on the type of tissue tested and on the detection method (Braunholz et al., 2015), which could artificially increase the proportion of de novo variants.

#### 3.2 | Rare examples of autosomal recessive inheritance in HPE patients

In 2007, first case of recessive inheritance was described, involving two compound heterozygous variants in *TGIF1* (El-Jaick et al., 2007). Then, homozygous variant in *FGF8* was identified in one consanguineous HPE family (McCabe et al., 2011). More recently, another case of recessive inheritance involving *FGF8* was described (Hong et al., 2018). This *FGF8*

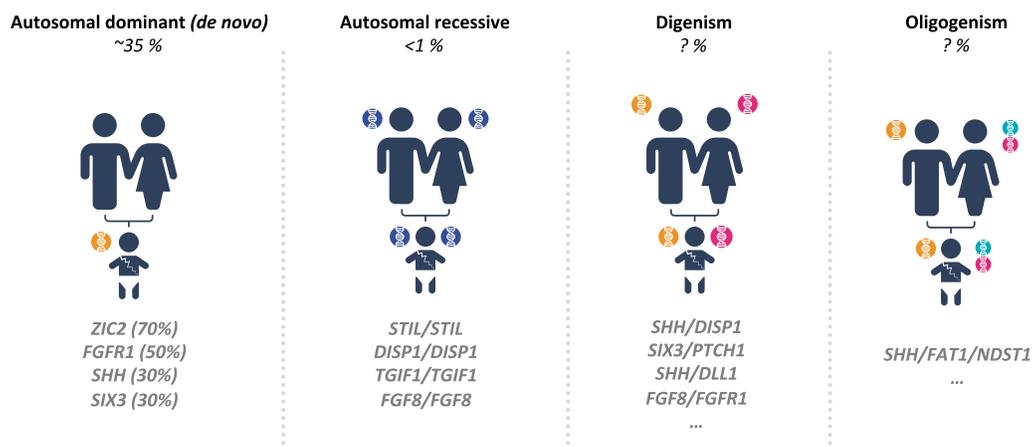


FIGURE 3 Different inheritance patterns for HPE are presented together with illustrative cases from our unpublished data

variant was functionally validated and shown to be hypomorph, which is consistent with the indispensable role of *FGF8* during early development (Sun, Meyers, Lewandoski, & Martin, 1999). A severe loss of function is probably not compatible with embryonic development.

Still, consanguineous families should predispose descendants to autosomal recessive gene combinations. Homozygosity mapping performed on eight consanguineous HPE families from our cohort did not initially reveal any homozygous variants. Only by complementing mapping with WES we were able to detect in one of these families a homozygous hypomorphic variant for the gene *STIL* (Mouden et al., 2015). A homozygous nonsense variant of *STIL* was independently reported by others (Kakar et al., 2015). *STIL* is localized to centrioles where it participates in SHH signaling through its function in primary cilia biology and is known to be involved in microcephaly (David et al., 2014). Notably, in these two families, HPE is transmitted as a recessive trait associated with severe microcephaly. As no additional variant of *STIL* was described on more than 100 HPE patients tested (Karkera et al., 2002; Mouden et al., 2015), *STIL* gene therefore belongs to the minor HPE genes.

We also described a female HPE patient displaying two different mutated *DISP1* alleles both inherited from her two healthy parents (Mouden et al., 2016). These missense variants in the exon 10 were predicted to be deleterious. *DISP1* is a protein that mediates the secretion of SHH, that is required for long-range cell-to-cell signaling (Tian, Jeong, Harfe, Tabin, & McMahon, 2005). This patient presents a localized fusion of forebrain hemispheres (mild form of HPE). We believe that hypomorphic effect of *DISP1* missense variant impacts SHH secretion to such an extent that global SHH signaling is decreased to pathological level.

All cases of autosomal recessive inheritance reported so far involve unaffected parents and concern minor HPE genes (Figure 3). Despite systematic sequencing of the major genes (*SHH*, *ZIC2*, and *SIX3*) in our 1,420 probands and others (Roessler, Vélez, Zhou, & Muenke, 2012), no recessive case has been reported for these genes. It suggests that homozygous variants in major genes are not compatible with embryonic development, which is fully consistent with their crucial roles during early developmental stages (Geng & Oliver, 2009; Schachter & Krauss, 2008) and may explain why recessive inheritance is rare in HPE.

### 3.3 | Oligogenic inheritance

As presented above, variants in HPE genes are frequently inherited from a parent without a typical HPE phenotype (Mercier et al., 2011). Within a given family, we observed a large phenotypic heterogeneity between the different variant carriers, illustrating the incomplete expressivity of these variants (Kruszka, Hart, Hadley, Muenke, & Habal, 2015; Mercier et al., 2011; Solomon et al., 2010; Stokes et al., 2018). These observations support the hypothesis that in these families variant in one HPE-related gene is necessary but not sufficient for the disease to occur, which implies more variants are required for complete phenotypic spectrum. This oligogenic mode of inheritance has already been proposed by our colleagues, who referred to it as “autosomal dominant with modifier effects” (S. Hong et al., 2016; M. Hong et al., 2017). In this oligogenic model, penetrance and expressivity of existing heterozygous variant is modulated by variants in other genes

associated with HPE. Such a synergistic effect between distinct deleterious genetic events is now well-documented in several other hereditary developmental diseases (e.g., Alport, Bardet-Biedl, and Kallman syndromes) (M'hamdi et al., 2014; Maione et al., 2018; Mencarelli et al., 2015). In a similar manner, HPE could result from cumulated effects of distinct variants. The use of animal models has reinforced this possibility Mercier et al., 2013. Numerous examples of double heterozygous mutant mice displaying HPE-like phenotypes provided evidence for oligogenism by implicating genes controlling either the same or distinct signaling pathways (Krauss, 2007).

For years, when a patient presented a deleterious variant in HPE gene, the analysis was interrupted because a likely genetic cause for the disease had been felt to be found. We believe it explains why rare HPE cases compatible with oligogenism have been reported so far (Ming & Muenke, 2002).

High-throughput sequencing and genomic technologies provided a unique opportunity to address this oligogenic inheritance in HPE. Thanks to WES, we have recently started to address the presence of additional events in HPE patients with a known variant in a HPE gene. Some of our most recent unpublished results on the subject are presented below.

#### 3.3.1 | Examples of digenic inheritance in HPE

Digenic inheritance is the simplest form of inheritance for genetically complex diseases. Systematic sequencing of major HPE genes have described isolated cases with variants in two genes (e.g., *SHH/ZIC2*; *SHH/TGIF*) (Ming & Muenke, 2002). We also reported several cases of chromosomal rearrangement (CNV) associated with variant in HPE-related gene (Mercier et al., 2011).

In one family, presented in Figure 2a, deleterious variants in both *SHH* and *DISP1* co-segregated with the disease, while relatives carrying variant only in either *SHH* or *DISP1* presented a mild or no disease phenotype (Mouden et al., 2016). This first digenic case in a family with several HPE patients prompted us to consider and further investigate digenic inheritance in families with HPE. This hypothesis was significantly reinforced by our experience with HPE genes routinely analyzed by targeted NGS. This study revealed that out of 257 HPE probands, 16% of the variants used for diagnosis were found in association with a second variant (e.g., *FGF8/FGFR1*, *FGF8/DLL1*, *DLL1/SHH*, *DISP1/SUFU*) (Dubourg et al., 2016). More recently, two more cases of digenism were reported (*ZIC2/BOC* and *TGIF/BOC*) in HPE patients (Hong et al., 2017). Considering the systematic use of NGS on HPE patients and the increasing number of HPE genes, we expect that digenic cases will continue to accumulate further.

#### 3.3.2 | Cases of oligogenic inheritance in HPE

The increasing number of digenic cases provides novel insights into genetic etiology of HPE such as oligogenic inheritance. Oligogenic pattern has never been described in HPE, but is often suggested in other complex diseases such as ciliopathies (Reiter & Leroux, 2017), retinitis pigmentosa (Ali, Rahman, Cao, & Yuan, 2017), autism spectrum disorder (Yin & Schaaf, 2017), amyotrophic lateral sclerosis (Nguyen, Van Broeckhoven, & van der Zee, 2018), or porphyria (Lenglet et al., 2018).

These disorders share with HPE high genetic and phenotypic heterogeneity as well as incomplete penetrance. We therefore considered oligogenic inheritance as a likely cause of HPE that may account for at least a substantial part of enigmatic cases.

With constantly evolving knowledge of disease genes and emergence of new analysis methods, it has become important to us to reanalyze systematically previous unsolved cases. In that aim, we reevaluated unsolved HPE families by taking into account the possibility of oligogenic inheritance. The study included families with no identified causal variant as well as families with variants in *SHH*, *ZIC2*, *SIX3*, or *TGIF1* inherited from clinically unaffected/mildly affected parent. Parts of this unpublished work are presented and discussed hereafter.

We combined trio-based WES with deep clinical phenotyping of the patients. Variant analysis was further improved by a gene prioritization approach based on clinical ontologies and co-expression networks of known disease-related signaling pathways.

In one family, presented in Figure 2b, we identified a variant in *SHH*, which was inherited from the mother presenting with hypotelorism. Considering oligogenic inheritance, we addressed whether paternal variants could contribute to HPE phenotype in combination with the *SHH* variant inherited from the mother. Indeed, our analysis revealed paternally inherited variants in two candidate HPE genes—*FAT1* and *NDST1*. Both variants were rare (minor allele frequency below 1%) and were predicted deleterious by the majority of bioinformatics algorithms (Combined Annotation Dependent Depletion (CADD) score >20). *FAT1* is a protocadherin and its knockdown in mouse causes severe midline defects including HPE (Ciani, Patel, Allen, & French-Constant, 2003) and *NDST1* is an N-deacetylase sulfotransferase and the corresponding mice mutants exhibit reduced SHH signaling and HPE-like phenotype (Grobe, 2005). Additionally, clinical phenotyping and analysis of cross-species similarities provided further evidence of causality for these genes by revealing a strong overlap of clinical features between the patient and the *FAT1* (proboscis) and *NDST1* (eye defects) mutant mice. Finally, the segregation analysis showed that the combination of variants in *SHH/FAT1/NDST1* was exclusively found in the two affected individuals of this family (Figure 2b). This example nicely illustrates oligogenic inheritance in HPE where the disease results from accumulation of multiple variants in genes associated to HPE phenotypes and/or implicated in SHH signaling (Figure 3). We therefore pursue a systematic reevaluation of all unsolved HPE cases. We believe that numerous other genes will be characterized in patients with oligogenism transmission. An exciting future challenge will be to test experimentally the combined effect of these different variants on early brain development.

#### 4 | OUR CLINICAL APPROACH

Over the last two decades, we have followed for prenatal diagnosis 26 HPE families affected by a first case of severe HPE carrying a variant in one of the major HPE gene. In 18 instances, we were able to reassure the parents after establishing the absence of the gene alteration in the fetus. Fetal MRI scan was normal later in pregnancy, and no child had

medical problems after birth. A genetic alteration (in *SHH*, *SIX3*, or *TGIF1*) was found in the eight other cases: five children were born, either without brain malformation and asymptomatic, or presenting a less severe form than the proband as predicted by the fetal brain surveillance. Three pregnancies were interrupted after MRI scans showed HPE features.

Nowadays, even when a causative gene for HPE has been found in a patient, the molecular diagnosis is probably not fully established. In order to properly address the molecular diagnosis, it will be necessary to compare the detailed phenotypes of the different family members with the segregation of relevant rare variants. For practitioners involved in counseling, an important consideration is how to communicate results of genetic analysis when potentially deleterious variants are identified but not yet functionally validated. Ideally, determining the contribution of each variant to the phenotype would be a condition for reliable genetic counseling in HPE families with oligogenic transmission.

#### 5 | CONCLUSION

In this review, we aimed to present the different patterns of inheritance of HPE in the light of our experience (Figure 3). In some cases, the disease is due to de novo variants; in rare cases the disease exhibits classical Mendelian inheritance with autosomal recessive transmission. In most cases, it emerges that the penetrance and the phenotypic variability have digenic or oligogenic origin. This complex genetic architecture will be better understood by analysis of hundreds of genes with NGS techniques on unsolved HPE cases. Our future challenge will be to differentiate rare variants that have significant impact on the observed phenotype from those with no effect.

#### CONFLICT OF INTEREST

None.

#### ORCID

Véronique David  <http://orcid.org/0000-0003-1489-1345>

#### REFERENCES

- Ali, M. U., Rahman, M. S. U., Cao, J., & Yuan, P. X. (2017). Genetic characterization and disease mechanism of retinitis pigmentosa; current scenario. *3 Biotech*, 7(4), 251. <https://doi.org/10.1007/s13205-017-0878-3>
- Arauz, R. F., Solomon, B. D., Pineda-Alvarez, D. E., Gropman, A. L., Parsons, J. A., Roessler, E., & Muenke, M. (2010). A hypomorphic allele in the *FGF8* gene contributes to Holoprosencephaly and is allelic to gonadotropin-releasing hormone deficiency in humans. *Molecular Syndromology*, 1(2), 59–66. <https://doi.org/10.1159/000302285>
- Bae, G.-U., Domené, S., Roessler, E., Schachter, K., Kang, J.-S., Muenke, M., & Krauss, R. S. (2011). Mutations in *CDON*, encoding a hedgehog receptor, result in holoprosencephaly and defective interactions with other hedgehog receptors. *American Journal of Human Genetics*, 89(2), 231–240. <https://doi.org/10.1016/j.ajhg.2011.07.001>
- Bakircioglu, M., Carvalho, O. P., Khurshid, M., Cox, J. J., Tuysuz, B., Barak, T., ... Woods, C. G. (2011). The essential role of centrosomal NDE1

- in human cerebral cortex neurogenesis. *American Journal of Human Genetics*, 88(5), 523–535. <https://doi.org/10.1016/j.ajhg.2011.03.019>
- Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A., & Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics*, 12(11), 745–755. <https://doi.org/10.1038/nrg3031>
- Barr, M., Hanson, J. W., Currey, K., Sharp, S., Toriello, H., Schmickel, R. D., & Wilson, G. N. (1983). Holoprosencephaly in infants of diabetic mothers. *The Journal of Pediatrics*, 102(4), 565–568.
- Bear, K. A., Solomon, B. D., Antonini, S., Arnhold, I. J., França, M. M., Gerkes, E. H., ... Muenke, M. (2014). Pathogenic mutations in *GLI2* cause a specific phenotype that is distinct from holoprosencephaly. *Journal of Medical Genetics*, 51(6), 413–418. <https://doi.org/10.1136/jmedgenet-2013-102249>
- Bendavid, C., Dubourg, C., Gicquel, I., Pasquier, L., Saugier-Verber, P., Durou, M. R., ... David, V. (2006). Molecular evaluation of fetuses with holoprosencephaly shows high incidence of microdeletions in the HPE genes. *Human Genetics*, 119(1–2), 1–8. <https://doi.org/10.1007/s00439-005-0097-6>
- Bendavid, C., Dupé, V., Rochard, L., Gicquel, I., Dubourg, C., & David, V. (2010). Holoprosencephaly: An update on cytogenetic abnormalities. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, 154C(1), 86–92. <https://doi.org/10.1002/ajmg.c.30250>
- Bendavid, C., Rochard, L., Dubourg, C., Seguin, J., Gicquel, I., Pasquier, L., ... David, V. (2009). Array-CGH analysis indicates a high prevalence of genomic rearrangements in holoprosencephaly: An updated map of candidate loci. *Human Mutation*, 30(8), 1175–1182. <https://doi.org/10.1002/humu.21016>
- Bennett, J. T., Tan, T. Y., Alcantara, D., Tétrault, M., Timms, A. E., Jensen, D., ... McDonnell, L. M. (2016). Mosaic activating mutations in *FGFR1* cause encephalocraniocutaneous lipomatosis. *American Journal of Human Genetics*, 98(3), 579–587. <https://doi.org/10.1016/j.ajhg.2016.02.006>
- Braunholz, D., Obieglo, C., Parenti, I., Pozojevic, J., Eckhold, J., Reiz, B., ... Kaiser, F. J. (2015). Hidden mutations in Cornelia de Lange syndrome limitations of sanger sequencing in molecular diagnostics. *Human Mutation*, 36(1), 26–29. <https://doi.org/10.1002/humu.22685>
- Bruel, A. L., Franco, B., Duffourd, Y., Thevenon, J., Jego, L., Lopez, E., ... Thauvin-Robinet, C. HRISTEL. (2017). Fifteen years of research on oral-facial-digital syndromes: From 1 to 16 causal genes. *Journal of Medical Genetics*, 54(6), 371–380. <https://doi.org/10.1136/jmedgenet-2016-104436>
- Burnside, R. D. (2015). 22q11.21 deletion syndromes: A review of proximal, central, and distal deletions and their associated features. *Cytogenetic and Genome Research*, 146(2), 89–99. <https://doi.org/10.1159/000438708>
- Butler, M. G. (2017). Clinical and genetic aspects of the 15q11.2 BP1-BP2 microdeletion disorder. *Journal of Intellectual Disability Research*, 61(6), 568–579. <https://doi.org/10.1111/jir.12382>
- Ciani, L., Patel, A., Allen, N. D., & French-Constant, C. (2003). Mice lacking the giant protocadherin *mFAT1* exhibit renal slit junction abnormalities and a partially penetrant cyclopia and anophthalmia phenotype. *Molecular and Cellular Biology*, 23(10), 3575–3582.
- Cohen, M. M. (2006). Holoprosencephaly: Clinical, anatomic, and molecular dimensions. *Birth Defects Research Part A: Clinical and Molecular Teratology*, 76(9), 658–673. <https://doi.org/10.1002/bdra.20295>
- David, A., Liu, F., Tibelius, A., Vulprecht, J., Wald, D., Rothermel, U., ... Krämer, A. (2014). Lack of centrioles and primary cilia in *STIL*(-/-) mouse embryos. *Cell Cycle*, 13(18), 2859–2868. <https://doi.org/10.4161/15384101.2014.946830>
- de la Cruz, J. M., Bamford, R. N., Burdine, R. D., Roessler, E., Barkovich, J. A., Donnai, D., ... Muenke, M. (2002). A loss-of-function mutation in the CFC domain of *TDGF1* is associated with human forebrain defects. *Human Genetics*, 110(5), 422–428. <https://doi.org/10.1007/s00439-002-0709-3>
- Démurger, F., Pasquier, L., Dubourg, C., Dupé, V., Gicquel, I., Evain, C., ... David, V. (2013). Array-CGH analysis suggests genetic heterogeneity in rhombencephalosynapsis. *Molecular Syndromology*, 4(6), 267–272. <https://doi.org/10.1159/000353878>
- Dhamija, R., Kirmani, S., Wang, X., Ferber, M. J., Wieben, E. D., Lazaridis, K. N., & Babovic-Vuksanovic, D. (2014). Novel de novo heterozygous *FGFR1* mutation in two siblings with Hartsfield syndrome: A case of gonadal mosaicism. *American Journal of Medical Genetics Part A*, 164(9), 2356–2359. <https://doi.org/10.1002/ajmg.a.36621>
- Dubourg, C., Bendavid, C., Pasquier, L., Henry, C., Odent, S., & David, V. (2007). Holoprosencephaly. *Orphanet Journal of Rare Diseases*, 2(1), 8. <https://doi.org/10.1186/1750-1172-2-8>
- Dubourg, C., Carré, W., Hamdi-Rozé, H., Mouden, C., Roume, J., Abdelmajid, B., ... David, V. (2016). Mutational spectrum in holoprosencephaly shows that FGF is a new major signaling pathway. *Human Mutation*, 37(12), 1329–1339. <https://doi.org/10.1002/humu.23038>
- Dupé, V., Rochard, L., Mercier, S., Le Petillon, Y., Gicquel, I., Bendavid, C., ... David, V. (2011). NOTCH, a new signaling pathway implicated in holoprosencephaly. *Human Molecular Genetics*, 20(6), 1122–1131. <https://doi.org/10.1093/hmg/ddq556>
- El-Jaick, K. B., Powers, S. E., Bartholin, L., Myers, K. R., Hahn, J., Orioli, I. M., ... Muenke, M. (2007). Functional analysis of mutations in *TGIF* associated with holoprosencephaly. *Molecular Genetics and Metabolism*, 90(1), 97–111. <https://doi.org/10.1016/j.ymgme.2006.07.011>
- Fernandez, B. A., Roberts, W., Chung, B., Weksberg, R., Meyn, S., Szatmari, P., ... Scherer, S. W. (2010). Phenotypic spectrum associated with de novo and inherited deletions and duplications at 16p11.2 in individuals ascertained for diagnosis of autism spectrum disorder. *Journal of Medical Genetics*, 47(3), 195–203. <https://doi.org/10.1136/jmg.2009.069369>
- Geng, X., & Oliver, G. (2009). Pathogenesis of holoprosencephaly. *Journal of Clinical Investigation*, 119(6), 1403–1413. <https://doi.org/10.1172/JCI38937>
- Grobe, K. (2005). Cerebral hypoplasia and craniofacial defects in mice lacking heparan sulfate *Ndst1* gene function. *Development*, 132(16), 3777–3786. <https://doi.org/10.1242/dev.01935>
- Hahn, J. S., Barnes, P. D., Clegg, N. J., & Stashinko, E. E. (2010). Septopreoptic holoprosencephaly: A mild subtype associated with midline craniofacial anomalies. *American Journal of Neuroradiology*, 31(9), 1596–1601. <https://doi.org/10.3174/ajnr.A2123>
- Hong, M., Srivastava, K., Kim, S., Allen, B. L., Leahy, D. J., Hu, P., ... Muenke, M. (2017). *BOC* is a modifier gene in holoprosencephaly. *Human Mutation*, 38(11), 1464–1470. <https://doi.org/10.1002/humu.23286>
- Hong, S., Hu, P., Marino, J., Hufnagel, S. B., Hopkin, R. J., Toromanović, A., ... Muenke, M. (2016). Dominant-negative kinase domain mutations in *FGFR1* can explain the clinical severity of Hartsfield syndrome. *Human Molecular Genetics*, 25(10), 1912–1922. <https://doi.org/10.1093/hmg/ddw064>
- Hong, S., Hu, P., Roessler, E., Hu, T., & Muenke, M. (2018). Loss-of-function mutations in *FGF8* can be independent risk factors for holoprosencephaly. *Human Molecular Genetics*, <https://doi.org/10.1093/hmg/ddy106> [Epub ahead of print]
- Jean, D., Bernier, G., & Gruss, P. (1999). *Six6* (*Optx2*) is a novel murine *Six3*-related homeobox gene that demarcates the presumptive pituitary/hypothalamic axis and the ventral optic stalk. *Mechanisms of Development*, 84(1–2), 31–40.
- Jin, O., Harpal, K., Ang, S. L., & Rossant, J. (2001). *Otx2* and *HNF3beta* genetically interact in anterior patterning. *The International Journal of Developmental Biology*, 45(1), 357–365.

- Kakar, N., Ahmad, J., Morris-Rosendahl, D. J., Altmüller, J., Friedrich, K., Barbi, G., ... Borck, G. (2015). STIL mutation causes autosomal recessive microcephalic lobar holoprosencephaly. *Human Genetics*, 134(1), 45–51. <https://doi.org/10.1007/s00439-014-1487-4>
- Karkera, J. D., Izraeli, S., Roessler, E., Dutra, A., Kirsch, I., & Muenke, M. (2002). The genomic structure, chromosomal localization, and analysis of SIL as a candidate gene for holoprosencephaly. *Cytogenetic and Genome Research*, 97(1–2), 62–67. <https://doi.org/10.1159/000064057>
- Krauss, R. S. (2007). Holoprosencephaly: New models, new insights. *Expert Reviews in Molecular Medicine*, 9(26), 1–17. <https://doi.org/10.1017/S1462399407000440>
- Kruszka, P., Hart, R. A., Hadley, D. W., Muenke, M., & Habal, M. B. (2015). Expanding the phenotypic expression of sonic hedgehog mutations beyond holoprosencephaly. *The Journal of Craniofacial Surgery*, 26(1), 3–5. <https://doi.org/10.1097/SCS.0000000000001377>
- Lacbawan, F., Solomon, B. D., Roessler, E., El-Jaick, K., Domene, S., Velez, J. I., ... Muenke, M. (2009). Clinical spectrum of SIX3-associated mutations in holoprosencephaly: Correlation between genotype, phenotype and function. *Journal of Medical Genetics*, 46(6), 389–398. <https://doi.org/10.1136/jmg.2008.063818>
- Lenglet, H., Schmitt, C., Grange, T., Manceau, H., Karboul, N., Bouchet-Crivat, F., ... Gouya, L. (2018). From a dominant to an oligogenic model of inheritance with environmental modifiers in acute intermittent porphyria. *Human Molecular Genetics*, 27(7), 1164–1173. <https://doi.org/10.1093/hmg/ddy030>
- Maione, L., Dwyer, A. A., Francou, B., Guiochon-Mantel, A., Binart, N., Bouligand, J., & Young, J. (2018). GENETICS IN ENDOCRINOLOGY: Genetic counseling for congenital hypogonadotropic hypogonadism and Kallmann syndrome: New challenges in the era of oligogenism and next-generation sequencing. *European Journal of Endocrinology*, 178(3), R55–R80. <https://doi.org/10.1530/EJE-17-0749>
- McCabe, M. J., Gaston-Massuet, C., Tziaferi, V., Gregory, L. C., Alatzoglou, K. S., Signore, M., ... Dattani, M. T. (2011). Novel FGF8 mutations associated with recessive holoprosencephaly, craniofacial defects, and hypothalamo-pituitary dysfunction. *The Journal of Clinical Endocrinology and Metabolism*, 96(10), E1709–E1718. <https://doi.org/10.1210/jc.2011-0454>
- Mencarelli, M. A., Heidet, L., Storey, H., van Geel, M., Knebelmann, B., Fallerini, C., ... Renieri, A. (2015). Evidence of digenic inheritance in Alport syndrome. *Journal of Medical Genetics*, 52(3), 163–174. <https://doi.org/10.1136/jmedgenet-2014-102822>
- Mercier, S., Dubourg, C., Garcelon, N., Campillo-Gimenez, B., Gicquel, I., Belleguic, M., ... Odent, S. (2011). New findings for phenotype-genotype correlations in a large European series of holoprosencephaly cases. *Journal of Medical Genetics*, 48(11), 752–760. <https://doi.org/10.1136/jmedgenet-2011-100339>
- Mercier, S., David, V., Ratié, L., Gicquel, I., Odent, S., & Dupé, V. (2013). NODAL and SHH dose-dependent double inhibition promotes an HPE-like phenotype in chick embryos. *Disease Models and Mechanisms*, 6(2), 537–543. <https://doi.org/10.1242/dmm.010132>
- M'hamedi, O., Redin, C., Stoetzel, C., Ouertani, I., Chaabouni, M., Maazoul, F., ... Chaabouni, H. (2014). Clinical and genetic characterization of Bardet-Biedl syndrome in Tunisia: Defining a strategy for molecular diagnosis. *Clinical Genetics*, 85(2), 172–177. <https://doi.org/10.1111/cge.12129>
- Ming, J. E., Kaupas, M. E., Roessler, E., Brunner, H. G., Golabi, M., Tekin, M., ... Muenke, M. (2002). Mutations in PATCHED-1, the receptor for SONIC HEDGEHOG, are associated with holoprosencephaly. *Human Genetics*, 110(4), 297–301. <https://doi.org/10.1007/s00439-002-0695-5>
- Ming, J. E., & Muenke, M. (2002). Multiple hits during early embryonic development: Digenic diseases and holoprosencephaly. *American Journal of Human Genetics*, 71(5), 1017–1032. <https://doi.org/10.1086/344412>
- Mouden, C., Dubourg, C., Carré, W., Rose, S., Quelin, C., Akloul, L., ... David, V. (2016). Complex mode of inheritance in holoprosencephaly revealed by whole exome sequencing. *Clinical Genetics*, 89(6), 659–668. <https://doi.org/10.1111/cge.12722>
- Mouden, C., de Tayrac, M., Dubourg, C., Rose, S., Carré, W., Hamdi-Rozé, H., ... David, V. (2015). Homozygous STIL mutation causes holoprosencephaly and microcephaly in two siblings. *Plos One*, 10(2), e0117418. <https://doi.org/10.1371/journal.pone.0117418>
- Muenke, M., & Beachy, P. A. (2000). Genetics of ventral forebrain development and holoprosencephaly. *Current Opinion in Genetics & Development*, 10(3), 262–269.
- Nguyen, H. P., Van Broeckhoven, C., & van der Zee, J. (2018). ALS genes in the genomic era and their implications for FTD. *Trends in Genetics*. <https://doi.org/10.1016/j.tig.2018.03.001> [Epub ahead of print]
- Odent, S., Le Marec, B., Munnich, A., Le Merrer, M., & Bonaïti-Pellié, C. (1998). Segregation analysis in nonsyndromic holoprosencephaly. *American Journal of Medical Genetics*, 77(2), 139–143.
- Pineda-Alvarez, D. E., Roessler, E., Hu, P., Srivastava, K., Solomon, B. D., Siple, C. E., ... Muenke, M. (2012). Missense substitutions in the GAS1 protein present in holoprosencephaly patients reduce the affinity for its ligand, SHH. *Human Genetics*, 131(2), 301–310. <https://doi.org/10.1007/s00439-011-1078-6>
- Ratié, L., Ware, M., Barloy-Hubler, F., Romé, H., Gicquel, I., Dubourg, C., ... Dupé, V. (2013). Novel genes upregulated when NOTCH signaling is disrupted during hypothalamic development. *Neural Development*, 8(1), 25. <https://doi.org/10.1186/1749-8104-8-25>
- Reiter, J. F., & Leroux, M. R. (2017). Genes and molecular pathways underpinning ciliopathies. *Nature Reviews Molecular Cell Biology*, 18(9), 533–547. <https://doi.org/10.1038/nrm.2017.60>
- Roessler, E., Belloni, E., Gaudenz, K., Jay, P., Berta, P., Scherer, S. W., ... Muenke, M. (1996). Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. *Nature Genetics*, 14(3), 357–360. <https://doi.org/10.1038/ng1196-357>
- Roessler, E., Du, Y.-Z., Mullor, J. L., Casas, E., Allen, W. P., Gillesen-Kaesbach, G., ... Muenke, M. (2003). Loss-of-function mutations in the human GLI2 gene are associated with pituitary anomalies and holoprosencephaly-like features. *Proceedings of the National Academy of Sciences of the United States of America*, 100(23), 13424–13429. <https://doi.org/10.1073/pnas.2235734100>
- Roessler, E., El-Jaick, K. B., Dubourg, C., Vélez, J. I., Solomon, B. D., Pineda-Álvarez, D. E., ... Muenke, M. (2009). The mutational spectrum of holoprosencephaly-associated changes within the SHH gene in humans predicts loss-of-function through either key structural alterations of the ligand or its altered synthesis. *Human Mutation*, 30(10), E921–E935. <https://doi.org/10.1002/humu.21090>
- Roessler, E., Lacbawan, F., Dubourg, C., Paulussen, A., Herbergs, J., Hehr, U., ... Muenke, M. (2009). The full spectrum of holoprosencephaly-associated mutations within the ZIC2 gene in humans predicts loss-of-function as the predominant disease mechanism. *Human Mutation*, 30(4), E541–E554. <https://doi.org/10.1002/humu.20982>
- Roessler, E., Ma, Y., Ouspenskaia, M. V., Lacbawan, F., Bendavid, C., Dubourg, C., ... Muenke, M. (2009). Truncating loss-of-function mutations of DISP1 contribute to holoprosencephaly-like microform features in humans. *Human Genetics*, 125(4), 393–400. <https://doi.org/10.1007/s00439-009-0628-7>
- Roessler, E., Ouspenskaia, M. V., Karkera, J. D., Vélez, J. I., Kantipong, A., Lacbawan, F., ... Muenke, M. (2008). Reduced NODAL signaling strength via mutation of several pathway members including FOXH1 is linked to human heart defects and holoprosencephaly. *American Journal of Human Genetics*, 83(1), 18–29. <https://doi.org/10.1016/j.ajhg.2008.05.012>

- Roessler, E., Pei, W., Ouspenskaia, M. V., Karkera, J. D., Veléz, J. I., Banerjee-Basu, S., ... Towbin, J. A. (2009). Cumulative ligand activity of NODAL mutations and modifiers are linked to human heart defects and holoprosencephaly. *Molecular Genetics and Metabolism*, *98*(1–2), 225–234. <https://doi.org/10.1016/j.ymgme.2009.05.005>
- Roessler, E., Vélez, J. I., Zhou, N., & Muenke, M. (2012). Utilizing prospective sequence analysis of SHH, ZIC2, SIX3 and TGIF in holoprosencephaly probands to describe the parameters limiting the observed frequency of mutant gene×gene interactions. *Molecular Genetics and Metabolism*, *105*(4), 658–664. <https://doi.org/10.1016/j.ymgme.2012.01.005>
- Rosenfeld, J. A., Ballif, B. C., Martin, D. M., Aylsworth, A. S., Bejjani, B. A., Torchia, B. S., & Shaffer, L. G. (2010). Clinical characterization of individuals with deletions of genes in holoprosencephaly pathways by aCGH refines the phenotypic spectrum of HPE. *Human Genetics*, *127*(4), 421–440. <https://doi.org/10.1007/s00439-009-0778-7>
- Ruiz i Altaba, A., Palma, V., & Dahmane, N. (2002). Hedgehog-Gli signaling and the growth of the brain. *Nature Reviews Neuroscience*, *3*(1), 24–33. <https://doi.org/10.1038/nrn704>
- Schachter, K. A., & Krauss, R. S. (2008). Murine models of holoprosencephaly. *Current Topics in Developmental Biology*, *84*, 139–170. [https://doi.org/10.1016/S0070-2153\(08\)00603-0](https://doi.org/10.1016/S0070-2153(08)00603-0)
- Simonis, N., Migeotte, I., Lambert, N., Perazzolo, C., de Silva, D. C., Dimitrov, B., ... Vilain, C. (2013). FGFR1 mutations cause Hartsfield syndrome, the unique association of holoprosencephaly and ectrodactyly. *Journal of Medical Genetics*, *50*(9), 585–592. <https://doi.org/10.1136/jmedgenet-2013-101603>
- Solomon, B. D., Bear, K. A., Wyllie, A., Keaton, A. A., Dubourg, C., David, V., ... Muenke, M. (2012). Genotypic and phenotypic analysis of 396 individuals with mutations in Sonic Hedgehog. *Journal of Medical Genetics*, *49*(7), 473–479. <https://doi.org/10.1136/jmedgenet-2012-101008>
- Solomon, B. D., Mercier, S., Vélez, J., Pineda-Alvarez, D. E., Wyllie, A., Zhou, N., ... Muenke, M. (2010). Analysis of genotype-phenotype correlations in human holoprosencephaly. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, *154C*(1), 133–141. <https://doi.org/10.1002/ajmg.c.30240>
- Stokes, B., Berger, S. I., Hall, B. A., Weiss, K. A., Martinez, A. F., Hadley, D. W., ... Muenke, M. (2018). SIX3 deletions and incomplete penetrance in families affected by holoprosencephaly. *Congenital Anomalies*, *58*(1), 29–32. <https://doi.org/10.1111/cga.12234>
- Storm, E. E., Garel, S., Borello, U., Hebert, J. M., Martinez, S., McConnell, S. K., ... Rubenstein, J. L. R. (2006). Dose-dependent functions of Fgf8 in regulating telencephalic patterning centers. *Development*, *133*(9), 1831–1844. <https://doi.org/10.1242/dev.02324>
- Sun, L., Carr, A. L., Li, P., Lee, J., McGregor, M., & Li, L. (2014). Characterization of the human oncogene SCL/TAL1 interrupting locus (Stil) mediated Sonic hedgehog (Shh) signaling transduction in proliferating mammalian dopaminergic neurons. *Biochemical and Biophysical Research Communications*, *449*(4), 444–448. <https://doi.org/10.1016/j.bbrc.2014.05.048>
- Sun, X., Meyers, E. N., Lewandoski, M., & Martin, G. R. (1999). Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo. *Genes & Development*, *13*(14), 1834–1846.
- Tian, H., Jeong, J., Harfe, B. D., Tabin, C. J., & McMahon, A. P. (2005). Mouse *Disp1* is required in sonic hedgehog-expressing cells for paracrine activity of the cholesterol-modified ligand. *Development*, *132*(1), 133–142. <https://doi.org/10.1242/dev.01563>
- Ware, M., Hamdi-Rozé, H., & Dupé, V. (2014). Notch signaling and neuronal genes work together to control the neural building blocks for the initial scaffold in the hypothalamus. *Frontiers in Neuroanatomy*, *8*, 140. <https://doi.org/10.3389/fnana.2014.00140>
- Xavier, G. M., Seppala, M., Barrell, W., Birjandi, A. A., Geoghegan, F., & Cobourne, M. T. (2016). Hedgehog receptor function during craniofacial development. *Developmental Biology*, *415*(2), 198–215. <https://doi.org/10.1016/j.ydbio.2016.02.009>
- Yin, J., & Schaaf, C. P. (2017). Autism genetics—An overview. *Prenatal Diagnosis*, *37*(1), 14–30. <https://doi.org/10.1002/pd.4942>

## AUTHOR BIOGRAPHIES



**C. DUBOURG** is a molecular geneticist, with a specific interest in the field of developmental anomalies (intellectual deficiency, holoprosencephaly), and in charge of molecular diagnoses of these diseases in the Laboratory of Molecular Genetics and Genomics at the University Hospital of Rennes.



**A. KIM** is a data scientist specialized in clinical NGS data analysis. He is currently working toward his PhD, which focuses on genetic basis of holoprosencephaly.



**E. WATRIN** holds a PhD in Biological Sciences and is principal investigator at the French CNRS (Centre National de la Recherche Scientifique). His work focuses on chromosome biology and its relevance to developmental disorders with a particular interest for cohesinopathies.



**M. DE TAYRAC** is head of the Laboratory of Bioinformatics and Computational Genetics at the University Hospital of Rennes. Her research activities at the Institute of Genetics and Development of Rennes focus on deciphering the genetic basis of human diseases using bioinformatics and systems biology

approaches.



**S. ODENT** is a Clinical Geneticist, head of the department and is responsible for the teaching of genetics at the Faculty of Medicine of Rennes. Her research projects focus on neurodevelopmental syndromes, with a special interest in brain malformations (holoprosencephaly) and rare diseases of development.

She coordinates an accredited center of expertise in the west of France (CLAD-Ouest) and is a member of the European Reference Network ITHACA.



**V. DAVID** is head of the Laboratory of Molecular Genetics and Genomics at the University Hospital of Rennes. She is also head of the research team « Genetics of pathologies related to development » at the Institute of Genetics and Development of Rennes, aiming at studying the genetics and physiopathology of complex developmental diseases.



**V. DUPÉ** is a research scientist at Inserm (Institut national de la santé et de la recherche médicale). After having acquired a strong background in vertebrate development at the IGBMC (Strasbourg, France) and UCL (London), she has joined the Institute of Genetics and Development of Rennes (IGDR) to study brain development with a special interest in holoprosencephaly.

**How to cite this article:** Dubourg C, Kim A, Watrin E, et al. Recent advances in understanding inheritance of holoprosencephaly. *Am J Med Genet.* 2018;178C:258–269. <https://doi.org/10.1002/ajmg.c.31619>

#### IV. Next Generation Sequencing in medical genomics

1977 marked the beginning of the DNA sequencing era with the publication of two novel sequencing methods. The method developed by Maxam and Gilbert was based on a combination of 4 base-specific chemical cleavages of a terminally-labeled DNA fragment followed by separation of the reaction products by gel electrophoresis (Maxam and Gilbert, 1977). This method presented however several drawbacks including the extensive use of chemical materials for DNA cleavage and a relatively complex set up. The method developed by Sanger and colleagues was based on the detection of the labeled analogs of nucleotide bases (called dideoxynucleotides) which are incorporated during the *in vitro* replication and act as base-specific inhibitors of DNA synthesis (Sanger et al., 1977). Similar to the Maxam and Gilbert method, the Sanger technique involves 4 separate sequencing reactions (one for each base and the corresponding dideoxynucleotide) and the resulting DNA fragments are separated by size using electrophoresis. Due to its speed and relative simplicity, the Sanger method was further refined and commercialized thus becoming the most widely used sequencing method in research and clinical genetics. In particular, the Sanger method was used in the Human Genome Project, the first effort to sequence the entire human genome which took 13 years (completed in April 2003) with an estimated cost of \$3.5 billion (Collins et al., 2003; Drake, 2011).

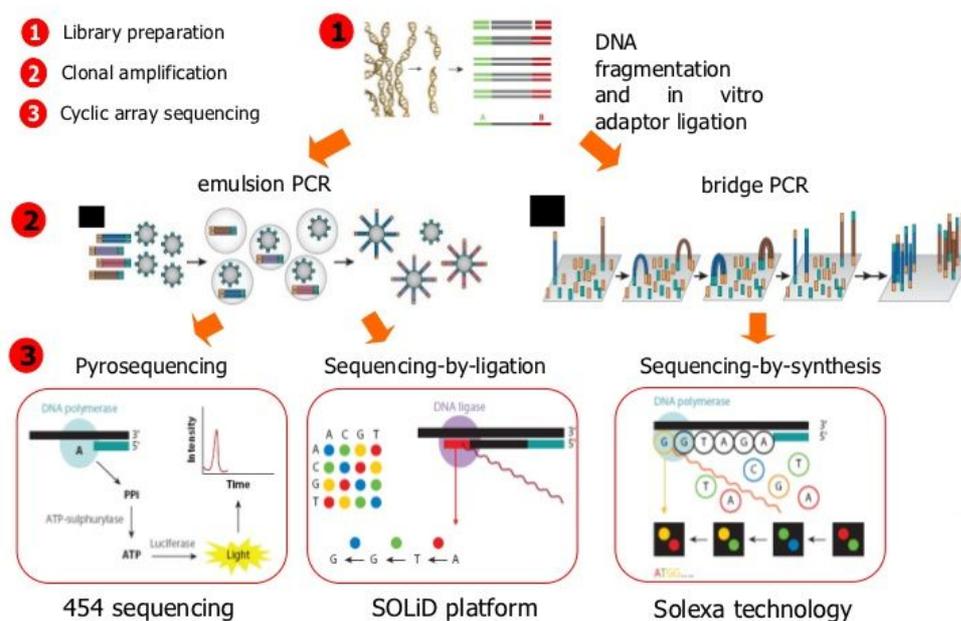
Since 2005, the emergence and commercialization of massively parallel sequencing methods, now referred to as Next Generation Sequencing (NGS), marked the new era of genomic analysis. The rapidly evolving field of NGS technologies has made it possible to perform a high-throughput analysis of genomic sequences from multiple samples, thus offering a much larger throughput than the conventional Sanger approach. The NGS technologies are rapidly developing and technical advances throughout the years resulted in large reductions of both time and cost needed for NGS analysis (Muir et al., 2016; Lightbody et al., 2019). In 2016, an entire human genome could be sequenced in one day. Nowadays, the process can take less than an hour, for a cost of less than

\$1000. The continuously decreasing costs and performance gains of NGS have brought it within the reach of many laboratories and institutions, thus accelerating discoveries and advances in biomedical research.

### **a. Main principles**

NGS technologies share several core steps (Figure 18). Depending on the input biological material and specific technical details, NGS technologies can be applied in a variety of ways to assist biological research (Kulski, 2016). The first major step in all NGS protocols is the fragmentation of the input DNA/RNA, using either mechanical methods (sonication, nebulization) or enzymatic digestion. The preferred fragmentation method is the sonication, which uses ultrasound waves to break chemical bonds and generates more evenly sized DNA fragments than enzyme digestion. The fragmented DNA is then ligated to adapters containing specific sequences designed to interact with the NGS platform. This process is known as library preparation. Ligation products are often separated by gel electrophoresis, and a specific size range is selected, depending on the platform and application. In case of targeted sequencing (*i.e.*, sequencing of a subset of genes / genomic regions), several 'target-enrichment' methods exist to selectively capture specific fragments containing the sequences of interest (Kozarewa et al., 2015). Amplicon-based methods rely on PCR amplification of target regions using sequence-specific primers and are generally preferred in analysis of specific disease gene panels. Hybridization-based methods, used in Whole Exome Sequencing, rely on probes - single-stranded oligonucleotides specific to the sequences of interest - to capture target sequences from the library. Depending on the sequencing technology, the resulting library fragments are amplified either by emulsion PCR or using bridge amplification (Voelkerding et al., 2010). After the library preparation and amplification steps, the DNA fragments are sequenced by the NGS platform. The sequencing step uses specifically labeled nucleotides/oligonucleotides that attach to DNA fragments in a stepwise fashion by complementarity. Each incorporation of a labeled molecule into the DNA chain

generates a luminescent or fluorescent image which is processed by the NGS platform and bioinformatically converted to the corresponding nucleotide. The nucleotide is then added into the sequence read, which is the final output of the NGS platform. Common sequencing techniques include pyrosequencing (detection of pyrophosphate release at each incorporation generating a base-specific luciferase signal), reversible dye terminators (detection of base-specific fluorescent emission) and sequencing by ligation (use of fluorescent oligonucleotide probes).



**Figure 18. Core steps of NGS.**

[www.brendelgroup.org](http://www.brendelgroup.org)

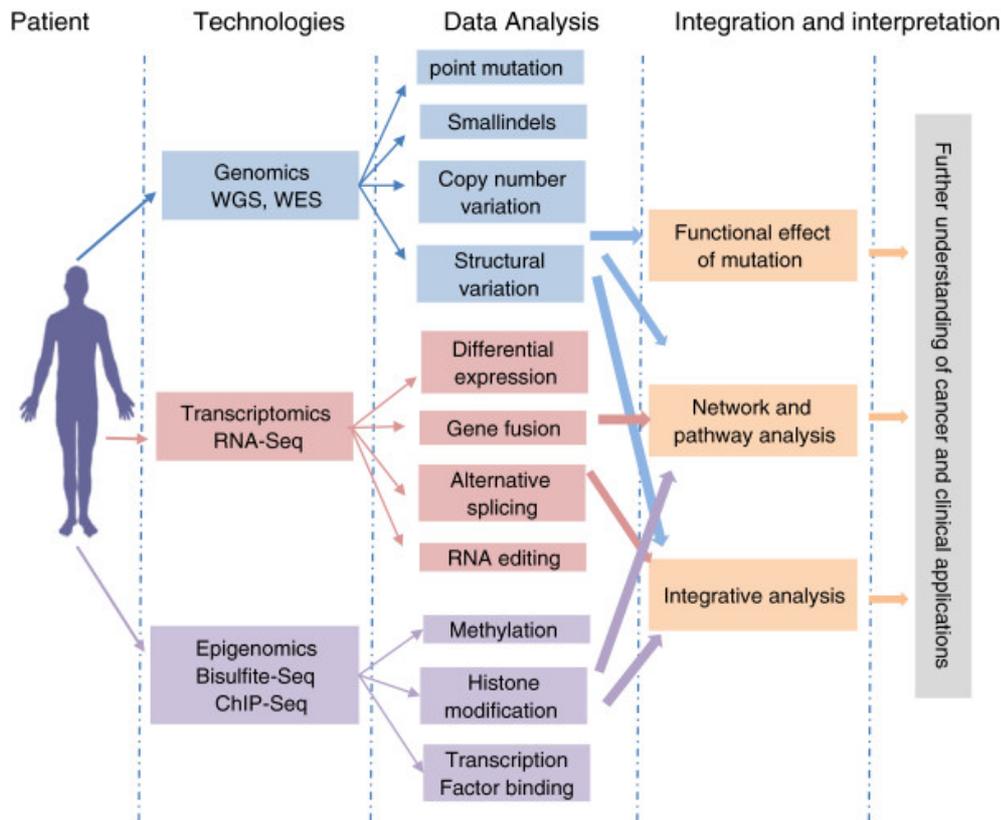
## b. Applications

Various applications of NGS methods, collectively termed as “omics” approaches, have enabled the study of human disease events at multiple levels (Figure 19).

At the genomics level, DNA-Sequencing, comprising Whole Genome, Exome and Targeted Sequencing (WGS, WES, TS) have enabled the identification of genomic alterations, such as point mutations (SNVs, Indels) and CNVs. WGS enables the identification of both non-coding and coding variations and has been used to identify pathogenic variants underlying various disorders, such as cardiomyopathy and

tuberculosis, among others (Cirino et al., 2017; Votintseva et al., 2017). WES, although restricted to variants in coding regions (exome), is a powerful cost-effective technique of disease gene discovery commonly applied in studies of genetic disorders (Bamshad et al., 2011). WES studies contributed to the identification of causal variants in numerous diseases including ocular disorders, rheumatoid arthritis and oral clefts (Gupta et al., 2017; Li et al, 2017 Bureau et al., 2014). TS can be applied in routine diagnostics to investigate pathogenic variants in a panel of known or candidate disease genes. Examples include studies of retinal disorders, NTDs and autism (Chen et al., 2013; Beaumont et al., 2019; O’Roak et al., 2012).

Other omics approaches are also widely used in studies of genetic disorders. Transcriptomics (RNA-Seq), which investigates the gene expression level in a given sample, can be used in a case/control differential expression analysis to aid pinpointing genes and molecular pathways altered by the disease pathogenesis (Ritchie et al., 2015). Moreover, RNA-Seq data can also be used to discover additional genomic alterations, such as gene fusions, or to investigate the splicing impact of DNA variants on gene expression (Vu et al., 2017; Brandão et al., 2019). Epigenomics (Bisulfite-Seq, CHIP-Seq) explores mechanisms that affect gene expression without altering the DNA sequence, such as methylations and histone modifications. Studies of epigenetic mechanisms have been conducted in various disorders including epilepsy, cancer and autism (Roopra et al., 2012; Shamra et al., 2010; Miyake et al., 2012).



**Figure 19. Applications of NGS in biomedical research.**

(Shyr and Liu, 2013)

### c. DNA-Seq and identification of disease-causing variants

DNA-Seq technologies, such as WES and WGS, are widely used in clinical genetics to detect pathogenic variants underlying the disease. However, a typical DNA-Seq experiment results in identification of an extremely large number of genetic variants, most of which either correspond to neutral variation (polymorphisms) or have a functional effect but are not actually pathogenic for the disease under investigation (MacArthur et al., 2014). Additionally, sequencing errors are common and further complicate the identification of genetic events underlying the disease. On average, exome sequencing detects between 20,000 and 50,000 SNVs and indels per individual, while genome sequencing can result in identification of several million variants (Belkadi et al., 2015; Lek et al., 2016). A key challenge in clinical DNA-Seq is, therefore,

a classic ‘needle in the haystack’ problem, *i.e.* identifying few phenotypically causal variants among a large background of non-pathogenic variants and sequencing errors (Bamshad, 2011; Cooper and Shendure, 2011).

Generally, a filtering approach is applied to the DNA-Seq output to exclude non-pathogenic variants and prioritize ‘candidate’ variants presenting arguments for their disease causality. Such approaches can be used at both the variant and the gene level. Common variant prioritization strategies used in clinical setting are described below.

### *Variant-level prioritization*

#### *Population frequency threshold*

Under assumption that genetic disorders are rare, filtering out variants commonly present in the general population is one of the main approaches used in pathogenic variant prioritization. The availability of large reference databases of human genetic variation makes it possible to obtain robust frequency estimates for most variants detected by WES/WGS. Exome Aggregation Consortium (ExAC) and its successor Genome Aggregation Database (gnomAD) are considered as main variant frequency sources, containing information from over 120,000 exomes and 10,000 genomes (Lek et al., 2016; Karczewski et al., 2019). Classical MAF thresholds used in variant prioritization include 1 %, 0.1 % and 0.01 %, depending on the disease inheritance mode and recurrence risk (Rybicki and Elston, 2000; Bamshad, 2011; Kobayashi et al., 2017). However, the frequency cutoffs are often arbitrary and must account for additional factors, such as the prevalence, genetic heterogeneity and penetrance of the studied disease (Whiffin et al., 2017). Although the frequency filtering is a powerful variant prioritization strategy, it presents a risk of eliminating pathogenic variants that are segregating in the general population at low frequencies. For recessive disorders, this approach could potentially discard causal variants that are detected at heterozygous state in healthy carriers. For complex disorders known to involve relatively common (MAF>1 %) variants, proper frequency thresholds remain unclear.

In some cases, the frequency of a variant can be substantially higher in certain subpopulations (phenomenon known as population stratification) thus further complicating interpretation of its disease relevance (Eilbek et al., 2017).

#### *Functional class*

Variants can be further prioritized based on their functional class. As illustrated by several studies, a large fraction of rare variants predicted to have a functional/deleterious impact are located in the protein-coding regions (Kryukov et al., 2007; Bamshad 2011). As the exome is considered to be the primary source of pathogenic variation, genetic investigations tend to prioritize exonic variants over variants located in non-coding regions. This strategy has been proven particularly successful in studies of Mendelian disorders (Gilissen et al., 2011).

Additionally, certain strategies tend to prioritize variants that are predicted to completely disrupt the function of the affected gene. Such variants are called Loss-of-Function (LoF) and include nonsense, frameshift and splicing alterations. LoF variants tend to be prioritized over missense and synonymous variants as they are believed to have a greater functional impact. This assumption can, however, be biologically incorrect. For example, if a nonsense variant occurs sufficiently downstream in a protein coding sequence, the truncated protein may still preserve its function. In contrast, a missense variant leading to an amino acid substitution at a functionally important site may have large deleterious consequences.

#### *Computational predictions*

The deleterious impact of a variant can be further investigated using *in silico* predictions. There exists a wide range of computational algorithms that assess the potential pathogenicity of a given variant (Eilbeck et al., 2017).

Conservation-based approaches rely on phylogenetic conservation to distinguish deleterious variants. Conservation scores are based on an assumption that the evolutionary conservation of the nucleotide/aminoacid reflects its functional importance. As such, variants affecting evolutionary conserved sites are believed to be

more damaging to the protein. Conservation scores exist at both nucleotide (GERP++, SipHy scores) and aminoacid (SIFT score) level (Davydov et al., 2010; Garber et al., 2009; Ng and Henikoff, 2001).

Functional scores use biological data to measure the deleteriousness of a variant. For example, Multivariate Analysis of Protein Polymorphism (MAPP) score prioritizes missense variants based on a physicochemical variation between the wildtype amino acid and its replacement. Polyphen-2 uses protein sequence and structure information to predict the potential impact of missense variants (Adzhubei et al., 2013).

Ensemble methods, such as CADD, MutationTaster2 and VAAST, integrate multiple levels of evidence (conservation, functional and biological data) into a statistical model to produce a single measure pathogenicity score for each variant (Kircher et al., 2014; Schwartz et al. 2014; Hu et al., 2013).

### ***Gene-level prioritization***

Despite the variety and relative success of prioritization techniques described above, hypothesis-free variant filtering (*i.e.*, based purely on variant characteristics without any *a priori* knowledge of genes or disease mechanisms) typically results in hundreds of candidate variants difficult to interpret in clinical context (Thuriot et al., 2018; Martin et al., 2019). In order to limit the number of candidate variants identified through DNA-Seq, various strategies of gene-level prioritization have been developed. As opposed to simply searching the genome for potentially deleterious variants, such strategies aim to identify candidate genes presenting evidence of disease causation (Eilbeck et al., 2017). In simple terms, gene-level prioritization reduces the search space for causal mutations by focusing on genes related to the studied disease. Lists of candidate genes can be defined using the combined expert knowledge of clinicians or biologists, as exemplified by PanelApp - a curated knowledgebase of genes presenting high evidence of disease causation (Martin et al., 2019). Moreover, gene prioritization approaches can use biological and statistical evidence to identify novel disease genes. Examples of such approaches are described below.

### *Protein-protein interactions*

Analysis of Protein-Protein Interaction (PPI) networks is one of the most widely used gene prioritization techniques. The majority of protein functions are performed through cooperation of a set of functionally-related proteins. In disease context, the main assumption is that genes related to the same disease encode proteins that interact with each other and display grouping patterns within specific molecular subnetworks (Ideker and Sharan, 2008). Therefore, candidate genes can be prioritized according to their proximity within the interactome to known disease genes. The data on more than 100,000 experimentally validated PPI are now available for human organism (Schaefer et al., 2013). Initial attempts to prioritize disease genes relied on direct interactions (Oti et al., 2006), but several recent methods, such as the STRING database, also integrate functional associations derived from co-expression data, genomic context and text-mining of scientific literature (Franceschini et al., 2013). PPI-based prioritization approaches have shown promising results in identifying disease-related genes of epilepsy, Parkinson's disease and cancer, among others (Campbell et al., 2013; Ferrari et al., 2018; Li et al., 2017).

### *Gene burden*

Another popular method to identify disease-related genes is to use burden testing. In this approach, rare variants observed in a given gene within the affected individuals are aggregated to calculate a gene-specific sum (burden) score. Genes presenting high burden scores can be prioritized as they are more likely to be damaged, indicating their potential disease implication. Most burden tests apply a case-control approach in which the burden of rare variants in each gene is compared between patients and unaffected individuals. Genes presenting a significant enrichment of rare variants in cases as compared to controls represent candidate genes associated to the disease. Burden testing has led to discovery of a number of candidate genes in various disorders, such as amyotrophic lateral sclerosis (Cirulli et al., 2015). This approach is, however, limited as it requires a large cohort of cases and ethnically matched controls to provide robust statistical estimations. Several approaches have been developed to

circumvent these limitations. Recently published TRAPD tool (Testing Rare vAriants using Public Data) enables the burden testing of rare variants using public control data from gnomAD (Guo et al., 2018). Variant Effect Scoring Tool (VEST) does not require a control cohort and instead employs a supervised machine learning method to identify genes enriched in functional variants across a set of disease exomes (Carter et al., 2013).

#### *Intolerance to variation*

Another approach very similar to burden testing is to use population data of genetic variation to infer genes whose function is more likely to be disrupted by the presence of a rare deleterious variant. This concept was termed as gene tolerance or gene constraint (Harvilla et al., 2019). In simple terms, genes with significantly more common variants observed in the general population than expected are inferred to be ‘mutation-tolerant’, *i.e.*, resistant to the presence of a rare deleterious variant. In contrast, genes showing less variation than expected, termed ‘mutation-intolerant’, are more likely to be damaged by the presence of a rare variant. Several measures exist to measure the gene constraint. Using the data of the NHLBI Exome Sequencing Project (ESP), Petrovski et al. established a gene-intolerance scoring system by comparing the observed number of common LoF and missense variants in each gene to the total number of observed variants (Petrovski et al., 2013). The most widely acclaimed gene constraint statistic is *probability of being loss-of-function intolerant (pLI)*, which makes use of the ExAC data (Lek et al., 2016). pLI calculates gene constraint by deriving a score (0-1 range) from a statistical model (expectation-maximization algorithm) which compares the observed and expected counts of LoF variants in each gene. Prioritizing LoF-intolerant genes (pLI score > 0.9) to search for causal variants has become a popular approach in clinical studies (Singh et al., 2017; Leonenko et al., 2017; Wang et al., 2019). Similar constraint scores are now also available for missense and synonymous variants in the gnomAD database.

For all prioritization strategies described above, no single line of evidence is sufficient to implicate a variant as disease-causing. In clinical context, variants are scored according to the guidelines jointly established by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP, Richards et al., 2015). These guidelines require multiple levels of evidence, both at variant and gene level, in order to classify a variant as pathogenic. Taking all the necessary factors into account (variant population frequency, *in-silico* predictions, gene constraint and burden, available biological knowledge) for all variants identified by DNA-Seq requires computational approaches of automated variant prioritization. As such, recent years marked the emergence of novel bioinformatic tools integrating various clinical and biological data with DNA-Seq output (Smedley et al., 2015; Pagel et al., 2020; Jiang et al., 2019).

Strategies for prioritizing pathogenic variants may vary and depend on additional factors, such as the inheritance pattern and locus heterogeneity of the disease, as well as the size and the type of the studied population (cohorts of unrelated patients or family-based studies). In case of a presumed monogenic disorder caused by a rare, highly penetrant exonic variant, a trio-based WES of the affected family is generally applied. In some cases, even WES of a single proband can be sufficient to identify the causal mutation (Cabral et al., 2012). In contrast, optimal strategies for studying complex disorders remain unclear. Large cohorts with appropriate controls are generally required in order to provide a robust statistical association of a variant to the disease. Moreover, the etiology of complex disorders can involve relatively common variants with a low deleterious effect. Further advances require new strategies implementing novel lines of evidence for pathogenic variant interpretation.

#### d. Novel strategies and future challenges for clinical variant interpretation

##### *Deep and cross-phenotyping*

Recent studies indicate that gene-driven prioritization outperforms hypothesis-free approaches that are purely based on variant deleteriousness (Stark et al., 2017). In particular, the identification of pathogenic variants can be improved by prioritizing genes associated with particular phenotypes found in the affected individual. The precise analysis of clinical abnormalities encountered in human disorders, termed as deep phenotyping, led to the establishment of Human Phenotype Ontology (HPO) - a comprehensive database of known disease-phenotype and gene-phenotype associations. Currently, HPO contains over 13,000 phenotypic terms and approximately 150,000 disease annotations extracted from scientific literature and medical databases. The phenotype-driven approach is used in many computational tools, such as PhenoRank, PhenoTips and Phenomizer, to generate lists of candidate genes associated with clinical phenotypes of interest (Cornish et al., 2018; Girdea et al., 2013; Kohler et al., 2009). Deep phenotyping is also integrated in variant prioritization frameworks such as PHIVE and PhenIX (Robinson et al., 2014; Smedley and Robinson, 2015).

Disease implication of a candidate gene can also be supported by the similarity between the studied pathology and the phenotype obtained in relevant animal models (MacArthur et al., 2014). Phenotype-driven approaches of gene prioritization also rely on gene-phenotype associations extracted from mutant studies of model organisms, such as mouse and zebrafish. For example, hiPHIVE algorithm implemented in the Exomiser tool performs a cross-species analysis to identify similarities between human disease manifestations and phenotypes observed in mice/zebrafish mutants. The resulting gene-phenotype associations are then integrated into a variant prioritization framework (Smedley et al., 2015). Although useful, phenotype-based approaches are mostly based on the existing knowledge, which limits their ability to identify novel gene-disease associations.

### *Integration of RNA-Seq data*

As mentioned above, transcriptomic data can aid pinpointing novel disease genes. In particular, spatiotemporal patterns of gene expression observed in different tissues/development stages can provide a valuable information for deciphering the genetic architecture of human disorders (Feiglin et al., 2017). Consequently, the integration of transcriptome-based knowledge has emerged as a powerful method for prioritizing pathogenic variants (Cummings et al., 2019). The advantage of this approach is that it is not based on any *a priori* knowledge and can uncover novel disease mechanisms. In particular, studies of differentially expressed and co-expressed (*i.e.*, sharing similar expression patterns across different tissues) genes has enabled the identification of novel risk factors in autism and Wilm's tumor, among others (Parikshak et al., 2016; Wang et al., 2019). Transcriptomic studies of gene expression are now facilitated by the existence of public databases which contain large numbers of RNA-Seq samples across a variety of tissues. One such resource is the Genotype-Tissue Expression (GTEx) project, which regroups various genetic and transcriptomic data of nearly 1000 individuals across 54 tissues (GTEx Consortium, 2013).

### *Regulatory and synonymous variants*

It is now clear that many non-coding regions of the genome, such as promoters and enhancers, act as crucial regulators of gene expression. Consequently, many computational tools that attempt to predict the impact of non-coding variants have emerged. Machine-learning approaches, such as FATHMM-MKL, employ a training method based on various genomic features to discriminate known disease and benign variants (Shihab et al., 2015). Tools such as DeepBind and DeepSEA integrate molecular data of histone modifications and chromatin accessibility into a deep learning-based sequence analysis to predict the pathogenic impact of non-coding variants (Alipanahi et al., 2015; Zhou and Troyanskaya, 2015). Finally, several approaches combine functional genomic data with measures of nucleotide-specific evolutionary constraint (Funseq2, Fu et al., 2014; DANN, Quang et al., 2015). However, approaches of non-coding variant prioritization are limited by the availability of training data and our

insufficient understanding of regulatory mechanisms encrypted in the non-coding part of the genome (Eilbeck et al., 2017).

The impact of synonymous single nucleotide variants (sSNVs) also remains underinvestigated in clinical studies. sSNVs can generate different molecular alterations at the mRNA and the protein level and have been shown to lead to pathologies by affecting mRNA splicing (Macaya *et al.*, 2009), mRNA structure (Bartoszewski *et al.*, 2010), miRNA-based regulation (Brest *et al.*, 2011) and protein conformation (Kimchy-Sarfaty et al., 2007). However, despite these discoveries, sSNVs are still largely ignored, unless directly associated with splicing defects (Bartoszewski *et al.*, 2010).

#### *Oligogenic inheritance*

Studies of complex disorders presenting oligogenic inheritance are faced with the challenging task of identifying pathogenic combinations of specific variants across multiple genes. The task is further complicated by the fact that such variants can be hypomorphic (*i.e.*, of small effect) and relatively common in the general population. Moreover, it has been argued that monogenic disorders, such as Long QT syndrome, may be better described by complex inheritance mechanisms (Schaffer, 2013), suggesting that hypomorphic variants may also participate in Mendelian disease pathogenesis. Consequently, deciphering oligogenic inheritance and hypomorphic variants involved in human disease represents one of the major challenges in modern genetics.

Several efforts to address this challenge have already been made. In 2016, Gazzo et al. reported the publication of a novel database, DIDA (Digenic diseases DAtabase), describing genetic variants involved in digenic disorders, the simplest form of oligogenic inheritance (Gazzo et al., 2016). DIDA currently contains 258 digenic combinations implicated in 54 genetic disorders. This resource enabled the development of Variant Combinations Pathogenicity Predictor (VarCoPP) which employs a machine-learning approach based on DIDA data to predict disease-causing

variant combinations (Papadimitriou et al., 2019). DIDA and VarCoPP have been integrated with additional lines of evidence, including PPI networks and pathogenicity predictions, to establish ORVAL - a novel platform for exploration of oligogenic variant combinations in clinical context (Renaux et al., 2019).

# ***OBJECTIVES***

As illustrated in the Introduction, the diagnosis of non-Mendelian disorders such as HPE is challenging due to their complex etiology involving multiple genetic factors. Unraveling the etiology of complex disorders requires novel bioinformatic strategies of variant interpretation integrating appropriate clinical and biological data with NGS output. My thesis work is part of the research conducted in the Genetics of Development-Related Pathologies team (GPLD, Institute of Genetics and Development of Rennes UMR6290 Rennes), which works on deciphering the genetics and physiopathology of HPE. The main objectives of my thesis were to elucidate the genetic basis of this pathology and to propose novel bioinformatic strategies for clinical interpretation of pathogenic variants underlying complex genetic disorders.

In my first study, I hypothesized that genetically unsolved cases of HPE follow oligogenic inheritance and result from a cumulative effect of multiple hypomorphic variants in distinct genes. I analyzed the trio-based exome data of HPE cases for whom no disease etiology was established by conventional diagnostic approaches. To identify variants of clinical interest in these enigmatic cases, I developed a new variant prioritization strategy integrating clinical gene-phenotype associations and transcriptome data of gene expression profiles. Applied to the exome data, this approach allowed the identification of particular combinations of rare hypomorphic variants presenting a strong association with the disease, which were further investigated using deep and cross-species phenotyping. The hypothesis was statistically validated using two independent control cohorts, thus providing evidence for oligogenicity as clinically relevant model in HPE.

In my second study, I further explored the role of hypomorphic variants in HPE by investigating the pathogenic impact of synonymous variants in the *SHH* gene found in HPE patients. Computational analysis indicated that the identified synonymous variants introduced significant changes in synonymous codon usage and were predicted to impact protein translation. Subsequent *in vitro* studies illustrated the impact of synonymous variants on translation, leading to misfolding, degradation and

reduced levels of the resulting SHH protein. Moreover, a significant correlation was observed between the computational and experimental results, thus underlining the relevance of certain *in silico* approaches in predicting the pathogenic impact of synonymous variants. The results of this study indicate that synonymous mutations may play a major role in complex genetic disorders.

During the final months of my thesis, I continued the work previously initiated by our research team and explored the hypothesis of autosomal recessive HPE in consanguineous families. In this study, I was able to establish a variant prioritization method integrating exome analysis with homozygosity mapping, which led to identification of several candidate genes.

Overall, main work of my thesis explores genetic mechanisms that are currently discarded during conventional diagnostic procedures and proposes novel bioinformatics strategies for their analysis. Ultimately, these results should help avoid misdiagnosis and improve patient care in genetic pathologies that remain to be resolved.

# ***RESULTS***

## I. Implication of hypomorphic variants and oligogenic inheritance in HPE.

### a. Background and objectives

HPE has been recently redefined as a complex disorder requiring the joint effect of several genetic variants (Dubourg et al., 2018). Despite numerous advances in understanding the genetic basis of HPE, conventional molecular testing approaches result in a very low diagnostic yield and most familial cases remain unsolved. Recent studies highlighted that non-Mendelian disease phenotypes could present oligogenic inheritance and result from accumulation of hypomorphic variants in multiple genes (Schäffer, 2013; Kousi and Katsanis, 2015). If such variants are inherited from an unaffected parent, such oligogenic events are likely overlooked in clinical studies, which could explain the low diagnostic yield of HPE. Therefore, we explored the possibility that genetically unsolved cases of HPE present an oligogenic origin and result from combined inherited variants in several genes.

### b. Methods

This study comprised familial cases of HPE for which no disease etiology could be established by conventional diagnostic procedures. Initial analyses included targeted panel sequencing of known disease genes, CNV analysis using CGH-array and a trio-based WES study following ACMG guidelines (Richards et al., 2015). As all initial analyses failed to establish the disease etiology, the studied families were considered eligible for the hypothesis of oligogenic inheritance.

A total of 26 HPE families (representing 80 individuals) have been included in this study. The corresponding WES data was analyzed using more permissive settings (described below). To identify disease-related variants, I developed a novel variant prioritization method together with Clara Savary, a Master II student whom I supervised during my first year of Ph.D. This novel approach of variant prioritization restricts the exome search space to candidate genes presenting evidence of disease association. This approach combines classical WES filtering (population frequency and deleteriousness

predictions) with two gene-based prioritization strategies involving clinical ontologies and analysis of gene co-expression networks.

#### *Whole Exome Sequencing and variant filtering*

For data alignment and variant calling, a pipeline using Burrows-Wheeler Aligner (BWA), Genome Analysis toolkit (GATK) and Freebayes was applied to all patients as previously described (Li and Durbin, 2009; Auwera et al., 2013; Garrison and Marth, 2012; Mouden et al., 2016). Briefly, the reads were aligned to the reference genome (hg19, BWA), followed by removal of PCR duplicates and recalibration of quality scores (GATK). Variants were called using 3 different algorithms: *HaplotypeCaller* and *Unified Genotyper* from GATK, complemented by *Freebayes*. The resulting variant calls were combined and low-quality variants were excluded using GATK filters (Table 4).

Filter	Value	Description
LowQD	QD<1.5	Variant Quality by Depth threshold
LowCoverage	DP < 5	Coverage threshold
VeryLowQual	QUAL < 30.0	Phred quality threshold
HARD_TO_VALIDATE	MQ0 >= 4 && ((MQ0 / (1.0*DP)) > 0.1)	Ambiguous alignments - potential artefacts
SnpcCluster	clusterWindowSize=10	More than 3 SNP in 10 bp region

**Table 4. GATK filters used for exome analysis.**

The resulting variants were annotated by ANNOVAR (Wang et al., 2010). The annotation step consists in assigning various information to each variant, such as variant's sequencing quality, genetic location as well as its frequency in population databases and predicted functional impact. A gene-based annotation using RefGene database was used to identify variants functional classes and the affected genes (O'Leary et al., 2016). The population frequency data was retrieved from dbSNP (build 147), 1000 Genomes (August 2015), Kaviar (September 2015), Exome Sequencing Project (ESP, March 2015), Greater Middle East Variome (GME), Exome Aggregation Consortium (ExAC) and its new version - Genome Aggregation Database (gnomAD) (Sherry et al., 2001; 1000 Genomes Project Consortium et al., 2015; Glusman et al., 2011; Scott et al., 2016; Lek et al., 2016). Variant impact was assessed using

deleteriousness prediction algorithms of dbNSFP and dbSCNV databases (Table 5). Complementary annotations were performed using Alamut Visual v2.8.1 (Interactive Biosoftware, Rouen, France) and Human Splicing Finder (HSF) (Desmet et al., 2009).

Score	Categorical prediction	Deleterious cutoff	Type	Source
GERP	higher scores are more deleterious	$\geq 4.4$	Conservation scores	dbNSFP (annovar)
SIPHY	higher scores are more deleterious	$\geq 12.17$		
CADD	Phred-like scaled C-score	$\geq 20$	Ensemble scores	
DANN	0-1. higher scores indicate a higher probability to be damaging	$\geq 0.98$		
METASVM	D: Deleterious; T: Tolerated	D		
METALR	D: Deleterious; T: Tolerated	D		
MCAP	-	$\geq 0.025$		
REVEL	0-1. higher scores indicate a higher probability to be damaging	$\geq 0.5$	Functional predictions	
SIFT	D: Deleterious (sift $\leq 0.05$ ); T: tolerated (sift $> 0.05$ )	D		
Polyphen2-Hdiv	D: Probably damaging ( $\geq 0.957$ ) P: possibly damaging ( $0.453 \leq \text{pp2\_hdiv} \leq 0.956$ ) B: benign ( $\text{pp2\_hdiv} \leq 0.452$ )	D		
Polyphen2-Hvar	D: Probably damaging ( $\geq 0.909$ ) P: possibly damaging ( $0.447 \leq \text{pp2\_hdiv} \leq 0.909$ ) B: benign ( $\text{pp2\_hdiv} \leq 0.446$ )	D		
LRT	D: Deleterious; N: Neutral; U: Unknown	D		
MutationTaster	A: disease_causing_automatic D: disease_causing N: polymorphism P: polymorphism_automatic	D/A		
MutationAssessor	H: high; M: medium; L: low; N: neutral	H/M		
FATHMM	D: Deleterious; T: Tolerated	D		
PROVEAN	D: damaging, N: neutral	D		
VEST3	0-1. higher scores indicate a higher probability to be damaging	$\geq 0.5$		
FATHMM_MKL	D: Deleterious; N: neutral	D		
GENOCANYON	0-1. higher scores indicate a higher probability to be damaging	$\geq 0.5$		
ADA	0-1. higher scores indicate a higher probability to be damaging	$\geq 0.8$	Splicing predictions	
RF	0-1. higher scores indicate a higher probability to be damaging	$\geq 0.8$		

**Table 5. Summary of *in silico* prediction algorithms used in this study.**

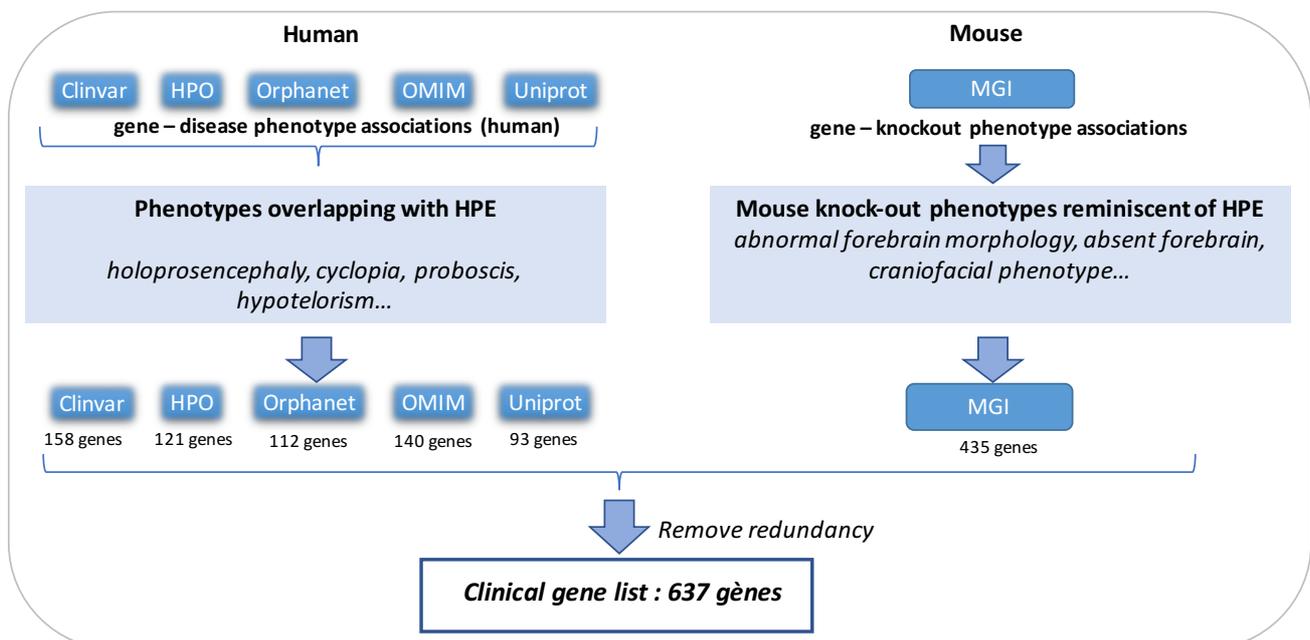
After annotation, a variant filtering step was performed on a family-by-family basis. Specifically, for each family, the following criteria was applied:

- (1) Exclude variants outside of exonic and splicing regions (within 10 bp outside of exons; RefSeq annotation)
- (2) For multiplex families, exclude variants absent in the most severely affected individuals
- (3) Exclude variants with population frequency equal/greater than 1 % (defined as polymorphic frequency threshold for this study) in any of the control databases
- (4) Exclude non-synonymous SNVs with less than 10 deleterious predictions (dbNSFP)

(5) Exclude splicing SNVs without any pathogenic predictions (dbscSNV)

### *Clinically-driven strategy*

The objective of the clinically-driven strategy was to establish a list of candidate genes previously associated with disease phenotypes, *i.e.*, HPE and craniofacial features frequently encountered in HPE context. Two lists of disease-relevant phenotypes were established for human and mouse organisms. Genes associated with the phenotypes of interest were extracted using existing gene-phenotype annotations from public databases and ontologies. For human, disease-relevant genes were extracted from Clinvar, HPO, Orphanet, Online Mendelian Inheritance in Man (OMIM) and Uniprot databases (Landrum et al., 2016; Weinreich et al., 2008; Apweiler et al., 2004; Ambreger et al., 2015; Kohler et al., 2017). For mouse, we used the Mouse Genome Informatics (MGI) database (Blake et al., 2003). The clinically-driven strategy established a list of 659 candidate genes associated to disease phenotypes (Figure 20).



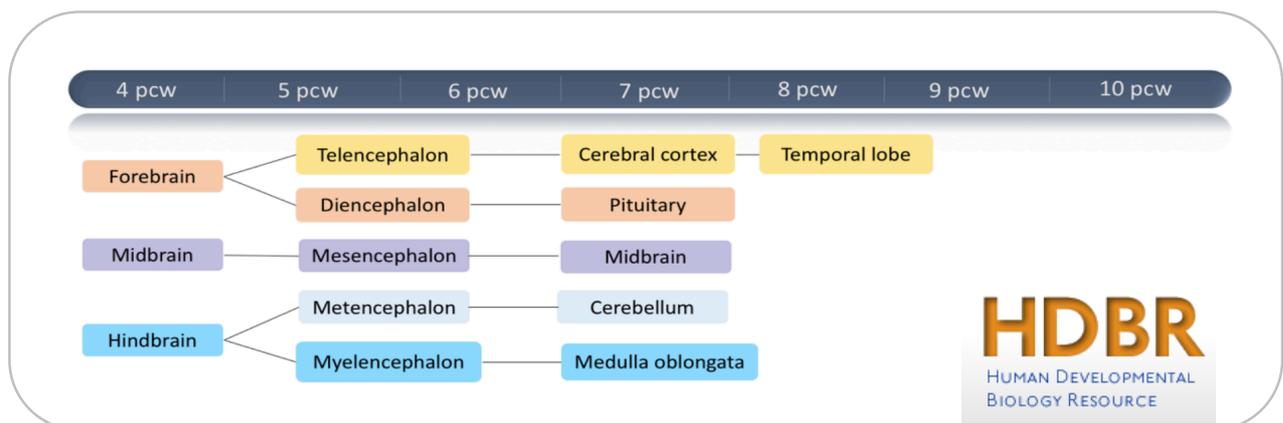
**Figure 20. Schematic representation of clinically-driven gene prioritization.**

### *Transcriptome-driven strategy*

While the clinically-driven strategy is based on existing gene annotations, transcriptome-based analysis can uncover novel disease mechanisms by establishing the first functional link between previously unrelated genes. Our hypothesis was that

genes implicated in HPE pathogenesis are highly co-expressed during the disease susceptibility period, *i.e.*, early brain development. Therefore, genes sharing highly similar expression patterns with major HPE genes (*SHH*, *ZIC2*, *SIX3* and *TGIF1*) during this period can represent novel candidate genes potentially involved in the disease pathogenesis.

To identify such genes, we established a second gene prioritization strategy using the RNA-Seq data of pre-natal human brain from Human Developmental Biology Resource (HDBR, Lindsey et al., 2016). A total of 136 samples were selected and pre-processed according to procedures available at <https://www.ebi.ac.uk/gxa/experiments/E-MTAB-4840>. The resulting dataset contained the expression data of 65217 transcripts across 136 samples corresponding to cerebral structures in early developmental stages (Figure 21).



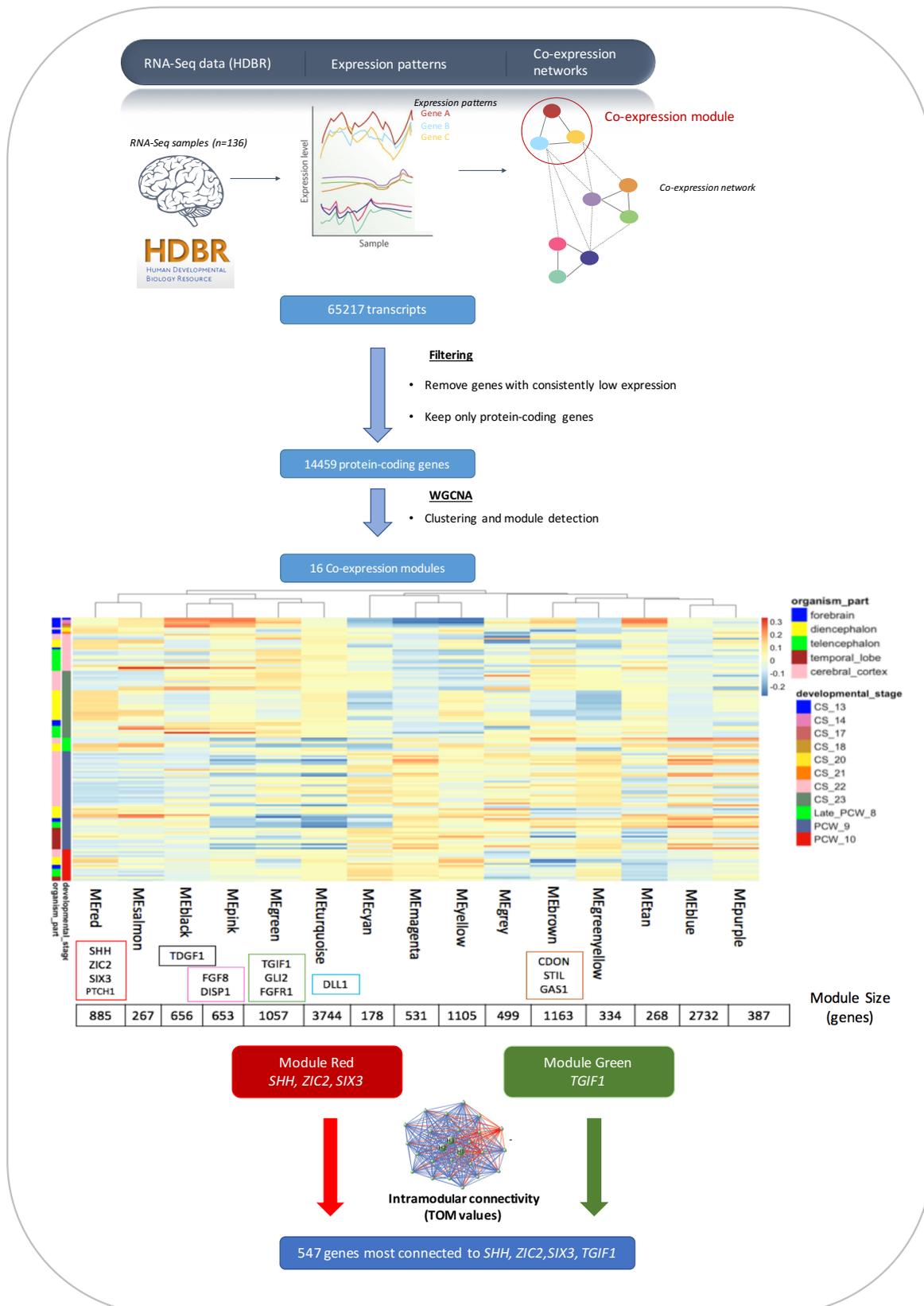
**Figure 21. Sample selection from HDBR.**

Further analysis was performed using Weighted Gene Co-expression Network Analysis (WGCNA), which enables the construction of co-expression networks and identification of groups of genes (gene modules) sharing highly similar expression patterns across the samples of the dataset (Langfelder and Horvath, 2008). WGCNA was performed using R environment (package WGCNA v.1.51). Following the recommendations of WGCNA developers, the HDBR dataset was filtered to exclude transcripts showing consistently low expression (< 10 reads in 90% of samples). For this study, only protein-coding genes were analyzed. Expression values were normalized

using DESeq2 (median normalization) (Love et al., 2014) and clustering by WGCNA removed two outlier samples presenting aberrant expression values. The final dataset contained expression values of 14459 protein-coding genes across 134 samples.

The module construction was based on the expression matrix using the *blockwiseModules* function of the WGCNA package. Briefly, this function calculates a similarity matrix of the expression values based on a Pearson correlation. The similarity matrix is then used to construct a Topological Overlap Matrix (TOM), which reflects the network topology and the relative inter-connectivity between the genes based on the number of common neighbors. TOM values are used to define different modules of highly connected genes using a hierarchical top-down clustering and a dynamic tree cut method.

In our study, a total of 14 co-expression modules were identified labeled by different colors (Figure 22). The 4 major HPE genes were regrouped within the red (*SHH*, *ZIC2*, *SIX3*) and green (*TGIF1*) modules. Genes sharing highly similar expression patterns with major HPE genes were identified using TOM values. Specifically, top 25% most connected partners (top 25% TOM values) of each major HPE gene were retained. This approach established a list of 547 candidate HPE genes based on transcriptomic evidence.



**Figure 22. Overview of the transcriptome-based gene prioritization and WGCNA analysis.**

The heatmap represents expression of each module across all samples of the dataset, as measured by *Module Eigengene*, the first principal component of each module which reflects modules' 'average' expression profile. CS - Carnegie Stage, PCW - Post-conception week.

### Integration

The final approach integrated all strategies described above (Figure 23). Variants identified by WES were filtered and restricted to candidate genes obtained by either transcriptomic ( $n=567$ ) or clinical strategies ( $n=637$ ). Manual analyses were performed on a family-by-family basis to identify oligogenic combinations of strong candidate variants in genes connected in a biologically meaningful way and presenting a significant link with HPE. The manual analyses included deep and cross-phenotyping, segregation study, functional enrichment analysis as well as thorough investigation of variant characteristics and available biological knowledge. Variant enrichment analysis was performed using two control cohorts provided by Genome of the Netherlands (GoNL) and French Exome Project (FREX).

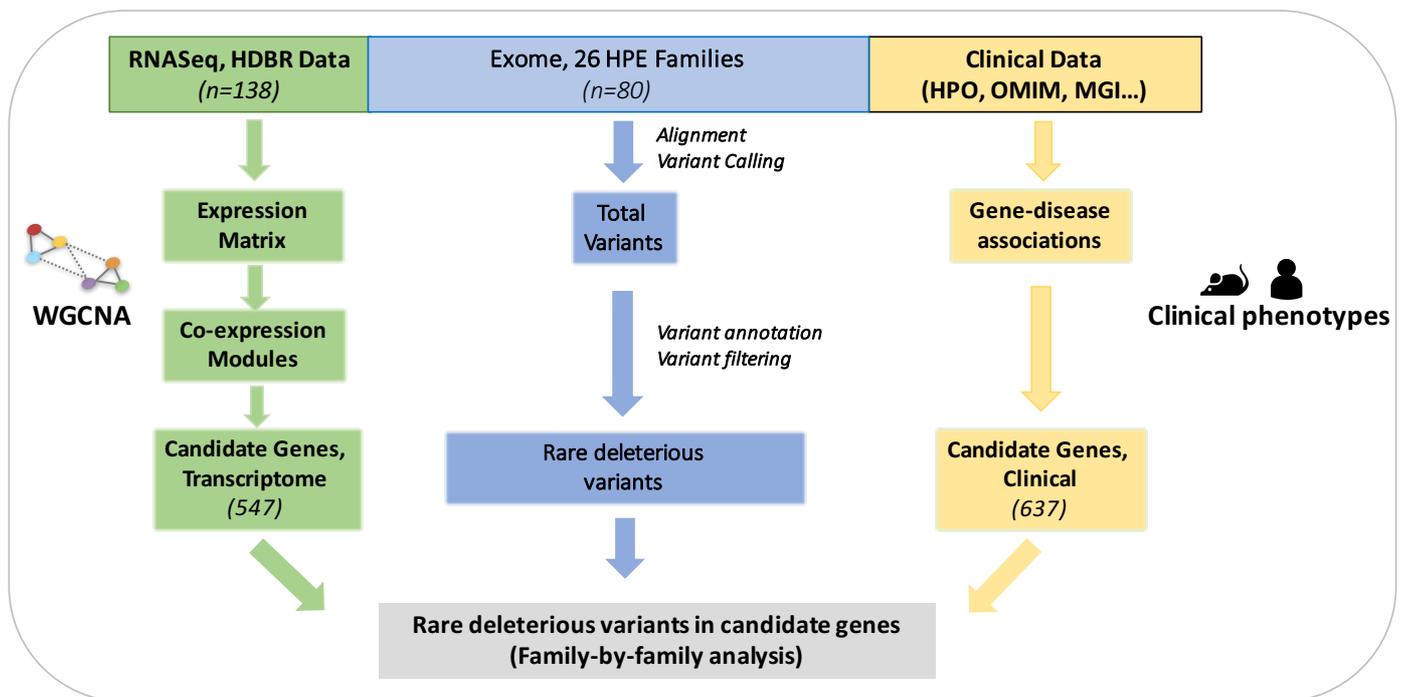


Figure 23. Integrative variant prioritization approach.

## c. Results

### 1. Article II

---

#### Integrated clinical and omics approach to rare diseases: novel genes and oligogenic inheritance in holoprosencephaly

Artem Kim, Clara Savary, Christèle Dubourg, Wilfrid Carré, Charlotte Mouden, Houda Hamdi-Rozé, Hélène Guyodo, Jerome Le Douce, FREX Consortium, GoNL Consortium, Laurent Pasquier, Elisabeth Flori, Marie Gonzales, Claire Bénéteau, Odile Boute, Tania Attié-Bitach, Joelle Roume, Louise Goujon, Linda Akloul, Sylvie Odent, Erwan Watrin, Valérie Dupé, Marie de Tayrac<sup>1,\*</sup>, Véronique David\*

1 - corresponding author: [marie.detayrac@univ-rennes1.fr](mailto:marie.detayrac@univ-rennes1.fr)

\* - equal contribution

Brain, January 2019

---

Applied to the 26 studied families, the integrative approach identified a total of 232 candidate variants in 180 genes significantly associated with key pathways of forebrain development including Sonic Hedgehog (SHH), primary cilia and WNT pathway. Manual analysis of the resulting candidates identified 10 families presenting oligogenic events, defined as combinations of strong candidate variants in  $\geq 2$  genes unique to the affected individuals of each family. The oligogenic events clustered among 19 genes including 15 novel disease genes presenting functional, clinical and statistical evidence of their implication in HPE. Analysis of control cohorts revealed that genes affected by the oligogenic combinations presented a significant burden of combined rare deleterious variants in patients ( $p < 10^{-9}$ ), indicating oligogenic inheritance as clinically relevant model in HPE. Deep phenotyping revealed that unrelated patients harbouring deleterious variants in the same gene present clinical similarities, while cross-species phenotyping revealed phenotypic overlaps between the patients and the mouse mutants for the corresponding gene. The observed overlaps further support the disease implication of the candidate variants and underline that integrating clinical phenotyping in genetic studies will improve the identification of causal genetic factors in complex disorders.

## Integrated clinical and omics approach to rare diseases: novel genes and oligogenic inheritance in holoprosencephaly

Artem Kim,<sup>1</sup> Clara Savary,<sup>1</sup> Christèle Dubourg,<sup>1,2</sup> Wilfrid Carré,<sup>2</sup> Charlotte Mouden,<sup>1</sup> Houda Hamdi-Rozé,<sup>1,2</sup> Hélène Guyodo,<sup>1</sup> Jerome Le Douce,<sup>1</sup> FREX Consortium, GoNL Consortium, Laurent Pasquier,<sup>3</sup> Elisabeth Flori,<sup>4</sup> Marie Gonzales,<sup>5</sup> Claire Bénéteau,<sup>6</sup> Odile Boute,<sup>7</sup> Tania Attié-Bitach,<sup>8</sup> Joelle Roume,<sup>9</sup> Louise Goujon,<sup>3</sup> Linda Akloul,<sup>3</sup> Sylvie Odent,<sup>3</sup> Erwan Watrin,<sup>1</sup> Valérie Dupé,<sup>1</sup> Marie de Tayrac<sup>1,2,\*</sup> and Véronique David<sup>1,2,\*</sup>

\*These authors contributed equally to this work.

Holoprosencephaly is a pathology of forebrain development characterized by high phenotypic heterogeneity. The disease presents with various clinical manifestations at the cerebral or facial levels. Several genes have been implicated in holoprosencephaly but its genetic basis remains unclear: different transmission patterns have been described including autosomal dominant, recessive and digenic inheritance. Conventional molecular testing approaches result in a very low diagnostic yield and most cases remain unsolved. In our study, we address the possibility that genetically unsolved cases of holoprosencephaly present an oligogenic origin and result from combined inherited mutations in several genes. Twenty-six unrelated families, for whom no genetic cause of holoprosencephaly could be identified in clinical settings [whole exome sequencing and comparative genomic hybridization (CGH)-array analyses], were reanalysed under the hypothesis of oligogenic inheritance. Standard variant analysis was improved with a gene prioritization strategy based on clinical ontologies and gene co-expression networks. Clinical phenotyping and exploration of cross-species similarities were further performed on a family-by-family basis. Statistical validation was performed on 248 ancestrally similar control trios provided by the Genome of the Netherlands project and on 574 ancestrally matched controls provided by the French Exome Project. Variants of clinical interest were identified in 180 genes significantly associated with key pathways of forebrain development including sonic hedgehog (*SHH*) and primary cilia. Oligogenic events were observed in 10 families and involved both known and novel holoprosencephaly genes including recurrently mutated *FAT1*, *NDST1*, *COL2A1* and *SCUBE2*. The incidence of oligogenic combinations was significantly higher in holoprosencephaly patients compared to two control populations ( $P < 10^{-9}$ ). We also show that depending on the affected genes, patients present with particular clinical features. This study reports novel disease genes and supports oligogenicity as clinically relevant model in holoprosencephaly. It also highlights key roles of SHH signalling and primary cilia in forebrain development. We hypothesize that distinction between different clinical manifestations of holoprosencephaly lies in the degree of overall functional impact on SHH signalling. Finally, we underline that integrating clinical phenotyping in genetic studies is a powerful tool to specify the clinical relevance of certain mutations.

- 1 Univ Rennes, CNRS, IGDR (Institut de génétique et développement de Rennes) - UMR 6290, F-35000 Rennes, France
- 2 Service de Génétique Moléculaire et Génomique, CHU, Rennes, France
- 3 Service de Génétique Clinique, CHU, Rennes, France
- 4 Laboratoire de Cytogénétique, Cytologie et Histologie Quantitative, Hôpital de Hautepierre, HUS, Strasbourg, France
- 5 Service de Génétique et Embryologie Médicales, Hôpital Armand Trousseau, Paris, France
- 6 Service de Génétique, CHU, Nantes, France

Received June 12, 2018. Revised August 21, 2018. Accepted September 28, 2018

© The Author(s) (2018). Published by Oxford University Press on behalf of the Guarantors of Brain. All rights reserved.

For permissions, please email: journals.permissions@oup.com

7 Service de Génétique, CHU, Lille, France

8 Service d'Histologie-Embryologie-Cytogénétique, Hôpital Necker-Enfants-Malades, Université Paris Descartes, 149, rue de Sèvres, 75015, Paris, France

9 Department of Clinical Genetics, Centre de Référence "AnDDI Rares", Poissy Hospital GHU PIFO, Poissy, France

Correspondence to: Dr Marie de Tayrac

Univ Rennes, CNRS, IGDR (Institut de génétique et développement de Rennes) - UMR 6290, F - 35000 Rennes, France

E-mail: marie.detayrac@univ-rennes1.fr

**Keywords:** exome; holoprosencephaly; oligogenic inheritance; sonic hedgehog; primary cilia

**Abbreviations:** GoNL = Genome of the Netherlands; HPE = holoprosencephaly; WES = whole exome sequencing

## Introduction

Holoprosencephaly (HPE1, OMIM #236100) is a severe developmental defect resulting from incomplete forebrain cleavage. The disease is characterized by incomplete separation of cerebral hemispheres with several anatomical classes ranging from microforms to lobar HPE. Affected individuals present with typical craniofacial midline defects of varying severity including proboscis, cleft lip and palate, ocular hypotelorism and solitary median incisor. HPE occurs in about 1 in 10 000 to 20 000 live births worldwide (Mercier *et al.*, 2011).

The genetic basis of HPE remains unclear and different transmission patterns have been described including autosomal dominant, recessive and digenic inheritance (Dubourg *et al.*, 2018). Most mutations associated with HPE display incomplete penetrance and variable expressivity, i.e. close relatives carrying the same pathogenic variant can be asymptomatic or present distinct HPE-spectrum anomalies (Mercier *et al.*, 2011). Sonic hedgehog (*SHH*) was the first discovered gene implicated in HPE (Roessler *et al.*, 1996) and its variants remain the most common cause of non-chromosomal HPE (Dubourg *et al.*, 2018). In 2011, molecular screening of 645 HPE probands revealed that mutations in the *SHH*, *ZIC2*, *SIX3* and *TGIF1* genes were the most frequent ones and collectively accounted for 25% of cases (Mercier *et al.*, 2011). The following studies reported that *GLI2* might also be considered as a major HPE gene in terms of frequency (Dubourg *et al.*, 2016), although variants in *GLI2* rarely result in classic HPE but instead cause a distinct phenotype that includes pituitary insufficiency and subtle facial features (Bear *et al.*, 2014). Pathogenic variants in *FGF8*, *FGFR1*, *DISP1*, and *DLL1* were also found in ~7% of HPE cases (Dupé *et al.*, 2011; Dubourg *et al.*, 2016). The other HPE genes reported so far are *TDGF1*, *FOXH1*, *TGIF1*, *CDON*, *NODAL*, *GAS1*, *STIL* and *SUFU* whose frequency is not established due to the small number of reported cases (Mouden *et al.*, 2015, 2016; Dubourg *et al.*, 2018; Kruszka *et al.*, 2018).

Clinical genetic testing of HPE has improved, but ~70% of familial cases remain without a clear molecular diagnosis. Most of known HPE genes belong to the *SHH* pathway, which represents the primary pathway implicated in the disease (Mercier *et al.*, 2013; Dubourg

*et al.*, 2016; Kruszka *et al.*, 2018). Therefore, defective *SHH*-related processes are likely to be substantially involved in HPE.

Whole-exome sequencing (WES) has been successful for Mendelian disease-gene discovery and differential diagnosis (Bamshad *et al.*, 2011). WES analysis uses filtering approaches for candidate variant prioritization combined with comprehensive clinical evaluation. A variety of additional strategies has been developed to further improve the performance of WES in clinical settings. Collaborative platforms such as Matchmaker Exchange (Philippakis *et al.*, 2015) are used to search for recurrence in patients affected by similar phenotypes. Integrative variant-prioritization algorithms such as the Exomiser suite (Smedley *et al.*, 2015) combine WES with different phenotype-driven approaches (based on clinical data and cross-species phenotype comparisons) and analysis of protein interactome data. As useful as they are, these strategies are limited: collaborative platforms are not efficient in case of very rare genetic diseases while pipelines such as Exomiser are not designed to study non-Mendelian disorders. Studying HPE faces these two challenges: (i) HPE live-born infants are excessively rare; and (ii) although HPE is considered a Mendelian disorder, the wide range of severity must necessitate strong modifying factors such that a single pathogenic variant may be neither necessary nor sufficient for pathogenesis.

Recent studies have highlighted that non-Mendelian disease phenotypes could present an oligogenic aetiology and result from accumulation of inherited low-penetrance variants in multiple genes (Li *et al.*, 2017). However, such events are likely overlooked in clinical genetic studies if variants are inherited from a clinically unaffected parent.

In this study, we address the additional yield that can be obtained for HPE patients who underwent medical WES evaluation in clinical settings that failed to establish a molecular diagnosis. Given the wide clinical spectrum of the disease, as well as incomplete penetrance and variable expressivity of HPE mutations, we raised the possibility that the low diagnostic yield is partly due to the complex aetiology of HPE and hypothesized that a part of unsolved HPE cases results from oligogenic events, i.e. accumulation of several rare hypomorphic variants in distinct, functionally connected genes.

Our study involved patients for whom no disease aetiology could be determined by conventional diagnostic approaches. Similarly to previous WES studies (Lee *et al.*, 2014; Stark *et al.*, 2017), we used clinically-driven prioritization approach to identify genes associated with specific clinical features as reported in gene-phenotype reference databases and mouse models. Complementarily, we developed and used a prioritization strategy based on gene co-expression networks of the developing human brain to select genes with spatio-temporal expression patterns compatible with those of known HPE genes. Finally, we used in-depth clinical phenotyping together with cross-species similarities to further strengthen the evidence of causality.

This study highlights novel HPE genes and identifies new disease-related pathways including the primary cilia pathway. Our findings also illustrate the high degree of oligogenicity of HPE and suggest that the disease requires a joint effect of multiple hypomorphic mutations.

## Materials and methods

### Patient selection and preliminary genetic analyses

Study protocol was approved by the Ethics Committee of Rennes Hospital. Patients diagnosed with HPE and relatives were recruited using the clinical database of Holoprosencephaly Reference Center of Rennes Hospital. Study participation involved informed written consent, availability of clinical data, and either DNA or peripheral blood sample.

The main selection criterion for this study was the absence of clear genetic cause of HPE after conventional diagnostic procedures. As part of routine diagnosis, all patients were scanned for rare damaging mutations by targeted HPE gene-panel sequencing (Dubourg *et al.*, 2016) and for copy number variants (CNVs) using comparative genomic hybridization (CGH)-array and multiplex ligation-dependent probe amplification (MLPA). Patients for whom no genetic cause of HPE (i.e. a fully-penetrant causal mutation in known HPE gene or a chromosomal aberration/copy number variant explaining the pathology) could be established, underwent trio-based WES for further analysis. WES was performed using standard procedures as previously described (Mouden *et al.*, 2015, 2016). The scheme for variant classification followed the American College of Medical Genetics and Genomics association (ACMG) guidelines (Richards *et al.*, 2015) and included a hypothesis-free analysis of all *de novo* and homozygous variants on a family-by-family basis. Patients for whom no such variants of clinical interest had been detected were considered eligible for the hypothesis of oligogenic inheritance and included in this study.

### Variant selection under oligogenic hypothesis

As discussed in previous studies, ACMG guidelines are useful in identifying variants with strong effect on phenotype but are unhelpful in case of modifier variants (Hong *et al.*, 2017).

Therefore, the ACMG classification was not taken into account for variant selection dedicated to the analysis of oligogenic events. WES trio data were reanalysed using more permissive settings (filtering protocols used in this study are described in the [Supplementary material](#)). The exome analysis was complemented with two gene prioritization strategies based on available clinical knowledge and co-expression networks.

### Clinically-driven approach

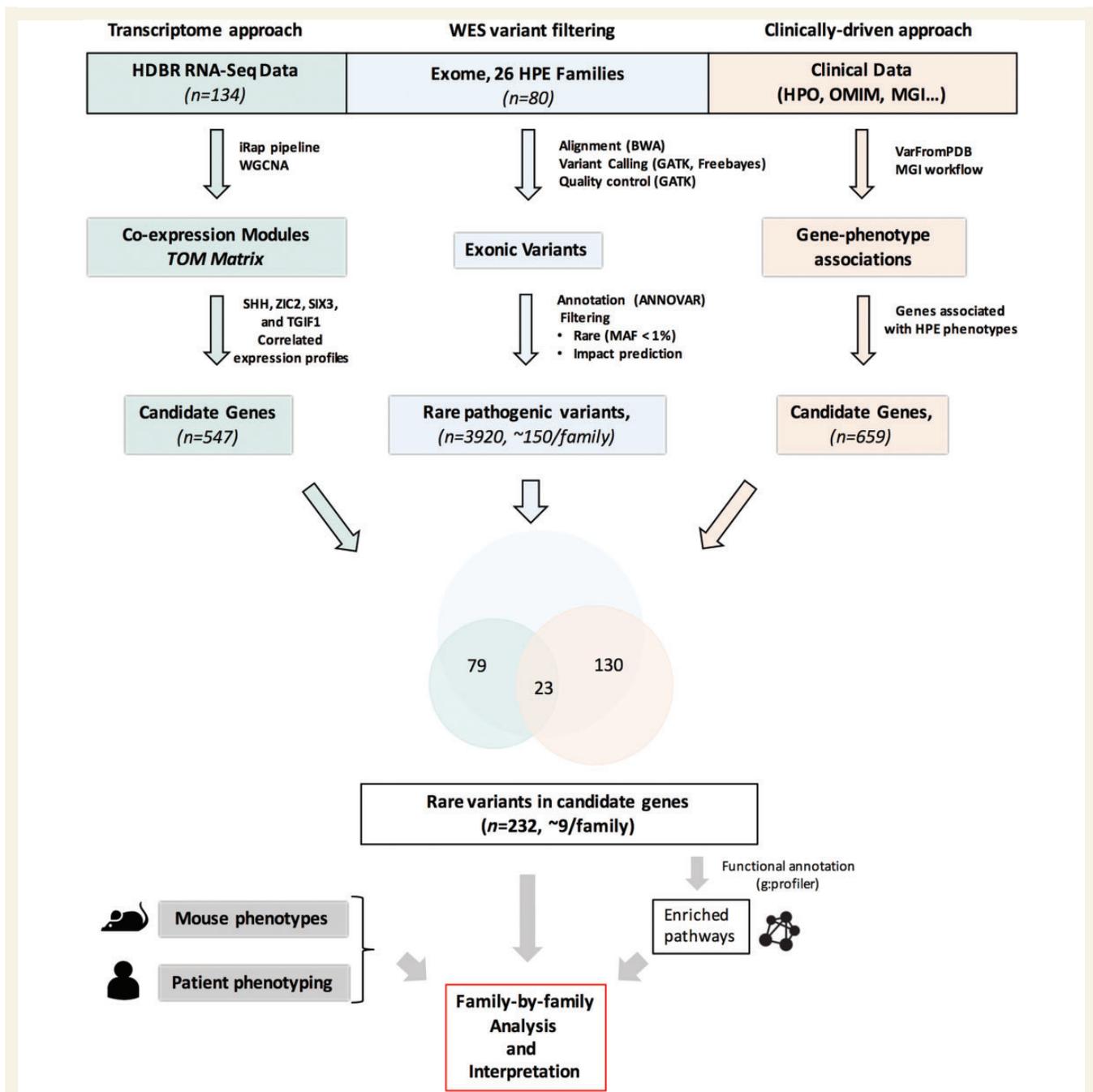
We established two clinician-generated lists of relevant phenotypes reminiscent of HPE in human and mouse models, respectively ([Supplementary Table 3](#)). Genes associated with the phenotypes of interest were identified with publicly available clinical resources and associated ontologies. Human gene-phenotype associations were extracted from relevant databases ([Supplementary Fig. 1](#)) using R package *VarFromPDB* (<https://github.com/cran/VarfromPDB>). The Mouse Genome Informatics (MGI) (Smith *et al.*, 2018) database and a homemade workflow were used to retrieve genes associated with any of the corresponding phenotypes in mouse mutants. Human and mouse results were combined and redundancy was removed to establish a list of clinically-driven candidate genes associated with HPE-related anomalies ([Supplementary Table 4](#)).

### Identification of HPE-related genes by weighted gene co-expression network analysis

We used weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008) on the RNA-Seq data from the Human Development Biology Resource (HDBR) (Lindsay *et al.*, 2016) to identify genes sharing highly similar expression patterns with four classical genes associated with HPE (*SHH*, *SIX3*, *ZIC2* and *TGIF1*) during cerebral development. Data from samples corresponding to forebrain, cerebral cortex, diencephalon, telencephalon and temporal lobe structures taken between the fourth and 10th post-conception weeks were selected ([Supplementary Fig. 9](#)). RNA-seq data were analysed with the iRAP pipeline (<https://github.com/nunofonseca/irap>). We used R package WGCNA to construct co-expression networks and identify modules of co-expressed genes. The detailed protocols for WGCNA analysis are described in the [Supplementary material](#). The Topological Overlap Matrix (TOM) matrix was used to establish a list of transcriptome-driven candidate genes sharing highly similar expression profiles with *SHH*, *ZIC2*, *SIX3* and *TGIF1* ([Supplementary Table 5](#)).

### Integration and identification of oligogenic events

The two gene prioritization schemes were combined with the WES results to identify a restricted list of rare variations located in genes identified by either the transcriptomic or the clinical prioritization approach (Fig. 1). Further analyses of the candidate variants were performed on a family-by-family basis. Oligogenic events were defined as combinations of candidate variants in  $\geq 2$  genes co-segregating with disease, i.e. unique to



**Figure 1** Flow chart illustrating the prioritization strategy. Classical WES analysis was performed (blue) and combined with two prioritization approaches: (i) based on gene co-expression networks (green); and (ii) based on clinical knowledge (salmon). Details of the pipeline are also provided in the [Supplementary material](#). Variant overlaps were selected and further analysed by functional annotation analysis and on a family-by-family basis, by integrating a comprehensive clinical phenotyping of patients and exploration of cross-species similarities.

the affected individuals of each family. Variants could be either inherited from the parents—at least one each from the mother and the father—or occur *de novo* in the affected child.

To evaluate the impact of candidate genes further, we performed deep clinical phenotyping to characterize similarities between unrelated patients and/or published knockout mice. Special attention was given to genes harbouring distinct rare variants in at least two affected patients with striking phenotypic overlap. Phenotypic overlaps between patients and mouse

mutants deficient for the corresponding candidate genes were also examined. The most interesting oligogenic combinations of rare deleterious variants in the affected children were finally discussed during multidisciplinary meetings.

To determine significantly enriched biological processes and pathways, functional annotation was performed by *g:profiler* (<http://biit.cs.ut.ee/gprofiler>) and Bonferroni adjusted *P*-value were considered significant below a value of 0.05 (KEGG, REACTOME and Gene Ontology Biological Processes).

## Control cohorts and validation

To test whether the identified oligogenic combinations were specific to the HPE cohort, we used SNV and INDELS data from 248 healthy trios (744 individuals) provided by Genome of the Netherlands (GoNL) sequencing project as a control cohort (Genome of the Netherlands Consortium, 2014). Additional control cohort consisting of 574 unrelated French individuals was provided by the French Exome Project (FREX).

We applied the same variant filtering approach and the same strategy for selection of oligogenic events. Proportion of families and/or individuals presenting oligogenic events were then compared between HPE cohort and the control cohorts. *P*-values were calculated using two-sided Fisher's exact test (*fisher.test* function in R, version 3.4.2).

## Data availability

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

## Results

### Clinical findings

We assembled a cohort of 26 families representing a total of 80 individuals including 29 affected children diagnosed with lobar ( $n=3$ ), semilobar ( $n=11$ ), alobar ( $n=13$ ) or microform HPE ( $n=2$ ) (Table 1). Common HPE clinical manifestations were observed among the probands and included cleft lip and palate (38%), hypotelorism (34%), microcephaly (31%) and arhinencephaly (31%). Ancestry analysis identified that 24 families were of European descent and two of South East Asia and African descent (Supplementary Fig. 10). Eight parents presented minor signs of midline facial anomalies and three parents were diagnosed with HPE microforms.

The initial targeted sequencing had identified point mutations in known HPE genes in 13 families and a full heterozygous deletion of *SIX3* gene had been detected by CGH-array in one family (Fig. 2 and Supplementary Fig. 8). All anomalies were later confirmed by WES analysis. They were inherited from asymptomatic or mildly affected parents and were considered as insufficient to fully explain the pathogenesis of HPE, suggesting that the presence of additional risk factors was required for the disease to occur.

### HPE variants overview and identification of disease-related pathways

Combined clinically- and transcriptome-driven analysis of the exome data identified a total of 232 rare candidate variants in 180 genes (Fig. 1 and Supplementary Table 6). All variants presented a minor allele frequency below 1% and were predicted to be highly deleterious to protein function (Supplementary material). One hundred and fifty-three variants concerned genes associated with HPE phenotypes

**Table 1 Clinical description of 26 HPE families**

Category and feature	<i>n</i>	%
<b>Proband sex</b>		
Male	6	21
Female	20	69
Unknown	3	10
<b>Total</b>	<b>29</b>	<b>100</b>
<b>Clinical phenotype of the parents</b>		
Unaffected	40	78
Minor sign	8	16
Hypotelorism	4	8
Incomplete iris	1	2
Epicanthus	1	2
Narrow palate	1	2
Nasal anomaly	1	2
HPE microform	3	6
<b>Total</b>	<b>51</b>	<b>100</b>
<b>Clinical characteristics of the probands</b>		
HPE	29	100
Lobar	3	10
Semilobar	11	38
Alobar	13	45
Microform	2	7
Cleft lip/palate	11	38
Hypotelorism	10	34
Microcephaly	9	31
Arhinencephaly	9	31
Agensis of corpus callosum	7	24
Flat head (plagiocephaly)	6	21
Thalami Fusion	6	21
Ventricles Fusion	6	21
Premaxillary agenesis	5	17
Fusion frontal lobes	4	14
Flat nose	4	14
Proboscis	3	10
Cyclopia	2	7
<b>Total</b>	<b>29</b>	<b>100</b>
<b>Families with mutations in HPE genes</b>		
SHH	4	15.4
ZIC2	1	3.8
SIX3	5*	19.2
TGIF1	2	7.7
PTCH1	1	3.3
ZIC2/GLI2	1	3.8
No mutation	12	46.2
<b>Total</b>	<b>26</b>	<b>100.0</b>
<b>Family ethnicity</b>		
European	21	81
African	1	4
South Asian	1	4
Admix	3	12
<b>Total</b>	<b>26</b>	<b>100.0</b>

\*For *SIX3*, point mutations were found in four families (targeted sequencing) and a heterozygous deletion was detected by CGH-array in one family.

among which 32 were located in genes reported to induce HPE-like phenotypes in mutant mice (Supplementary Table 8). One hundred and two variants were located in genes sharing expression profiles highly similar to those of







genes (Fig. 2A). *FAT1* is a protocadherin and its knock-down in mice causes severe midline defects including HPE (Ciani *et al.*, 2003); in *Drosophila* it has been shown to regulate the PCP pathway (Rock *et al.*, 2005). *LRP2*, *NDST1* and *COL2A1* are all functionally relevant to the SHH pathway (Fig. 3); *NDST1* and *COL2A1* mice mutants exhibit HPE phenotype and reduced SHH signalling in the forebrain (Grobe *et al.*, 2005; Leung *et al.*, 2010), while *LRP2* acts as an auxiliary receptor of SHH during forebrain development and its inactivation in mouse similarly leads to HPE phenotype (Christ *et al.*, 2012).

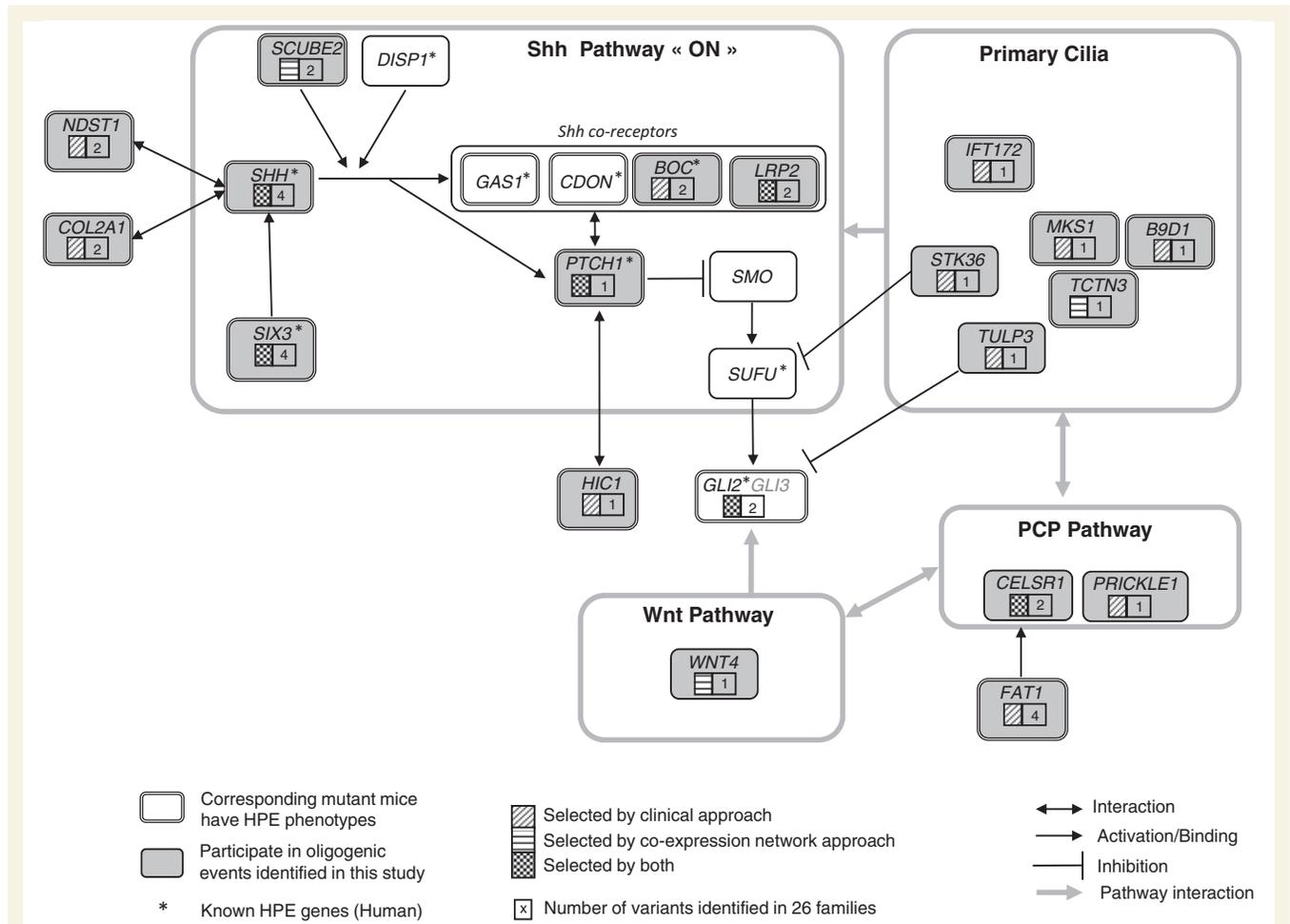
Oligogenic events involved the following combinations: *SHH/FAT1/NDST1* (Family F3), *FAT1/NDST1/COL2A1* (Family F16), *FAT1/COL2A1/PTCH1* (Family F26) and *FAT1/LRP2* (Family F23) (Fig. 2A, Tables 2 and 3). Details are provided in the [Supplementary material](#), Case report 1.

In Family F3, Sanger sequencing of additional family members revealed that the *SHH/FAT1/NDST1*

combination was unique to the affected individuals (Fig. 2A). For Family F16, only the foetus carrying the *FAT1/NDST1/COL2A1* combination was affected by semi-lobar HPE, while the sibling carrying *NDST1/COL2A1* variants presented only a microform (Fig. 2A). These observations are fully consistent with the oligogenic inheritance model where accumulation of multiple variants in genes associated to HPE phenotypes and/or HPE-related molecular pathways is required.

### Recurrent oligogenic events involving *SCUBE2/BOC* implicated in SHH signalling

Two families presented oligogenic events implicating combined variants in the *BOC* and *SCUBE2* genes (Fig. 2B, Tables 2 and 3). *BOC* is an auxiliary receptor of SHH and was recently reported as an HPE modifier in humans (Hong



**Figure 3** Implication of the candidate genes in the signalling pathways involved in HPE. Key affected pathways and genes are presented. Under each gene name, the selection methods (clinical or co-expression networks approach or both) is shown on the left and the number of variants for each gene is shown on the right. Genes known in HPE are marked with an asterisk, and genes for which corresponding mutant mice have HPE phenotypes are surrounded by a double line. The genes implicated in oligogenic events in the study are indicated with a grey background.

*et al.*, 2017). *SCUBE2* shares a highly similar expression pattern with *SHH* and *SIX3* and is implicated in the release of *SHH* from the secreting cell (Jakobs *et al.*, 2014). In Family F4, a combination of *SCUBE2/BOC* variants was associated with additional variants in *SHH*, *STK36* (see below) and *WNT4*, a member of the Wnt pathway, implicated in regulation of SHH signalling (Murdoch and Copp, 2010). In Family F22, the *SCUBE2* variant results in a premature stop codon at position 525 (Supplementary Fig. 7), which results in truncation of its CUB domain and is predicted to directly affect its SHH-related activity (Jakobs *et al.*, 2014). This family presented an additional candidate variant in *HIC1*, which genetically interacts with *PTCH1* (Briggs *et al.*, 2008). Mice deficient for *HIC1* exhibit craniofacial defects including HPE (Carter, 2000).

The reported variant combinations were observed exclusively in the affected probands and were absent in asymptomatic individuals. Altogether, these results reveal recurrent mutations in *SCUBE2/BOC* and further strengthen the oligogenic inheritance model of HPE.

## Implication of primary cilium in HPE

Remarkably, five families presented candidate variants in genes related to the primary cilium: *STK36*, *IFT172*, *B9D1*, *MKS1*, *TCTN3* and *TULP3* (Fig. 2C). Ciliary proteins are known to play essential roles in the transduction of SHH signalling downstream of *PTCH1* during forebrain development (Goetz *et al.*, 2009; Murdoch and Copp, 2010).

*STK36*, also known as ‘fused’, is a ciliary protein implicated in SHH signalling and associated to craniofacial phenotypes (Goetz *et al.*, 2009; Murdoch and Copp, 2010). *IFT172* codes for a core component of intraflagellar transport complex IFT-B required for ciliogenesis and regulation of SHH signal transduction. Moreover, *Ift172*<sup>-/-</sup> mice exhibit reduced expression of *Shh* in the ventral forebrain and severe craniofacial malformations including HPE (Gorivodsky *et al.*, 2009). *B9D1*, *MKS1* and *TCTN3* are all members of the transition zone protein complex implicated in regulation of ciliogenesis (Garcia-Gonzalo *et al.*, 2011). The disruption of *B9d1* and *Mks1* in mouse models causes craniofacial defects that include HPE (Dowdle *et al.*, 2011; Wheway *et al.*, 2013). Although no mouse model is available for *TCTN3*, its expression profile is highly similar to that of *SHH* and disruption of its protein complex partners (*TCTN1*, *TCTN2*, *CC2D2A*, *MKS1*, *B9D1*) leads to HPE in mouse (Dowdle *et al.*, 2011; Garcia-Gonzalo *et al.*, 2011; Wheway *et al.*, 2013). Moreover, *TCTN3* was shown to be necessary for the transduction of SHH signal and *TCTN3* mutations were found in patients affected by ciliopathies (Thomas *et al.*, 2012). Finally, *TULP3* is a critical repressor of *Shh* signalling in mice and is associated with various craniofacial defects (Murdoch and Copp, 2010).

Additional variants observed in these families include a heterozygous deletion of *SIX3*, missense mutations in *SHH*, *SCUBE2*, *BOC* and *LRP2* (described above) as well as two genes implicated in PCP pathway (Fig. 3): *CELSR1* (two

families) and *PRICKLE1*, both associated with craniofacial defects in mouse mutants (Fig. 2C) (Goetz *et al.*, 2009; Murdoch and Copp, 2010; Yang *et al.*, 2014). Similar to previously described cases, the oligogenic events were present exclusively in the affected children.

Given the essential role of the primary cilium in SHH signal transduction, these observations strongly suggest that rare variants in ciliary genes contribute to the disease onset in these families.

## Correspondence between affected genes and secondary clinical features

To provide additional evidence, we performed an in-depth analysis of secondary clinical features associated with HPE in our patients. Deep clinical phenotyping identified clinical similarities between unrelated patients (Tables 2 and 3) as well as overlaps of secondary clinical features between patients and the corresponding mouse mutants.

Interestingly, the two patients with variants in ciliary genes (*IFT172/PRICKLE1* and *SIX3/TCTN3/TULP3*) both presented with polydactyly, a clinical feature commonly associated with ciliopathies (Goetz *et al.*, 2009). Importantly, the patient with the oligogenic combination *IFT172/PRICKLE1* presented with a large set of overlapping clinical features with the corresponding mouse mutants including polydactyly, cleft palate and eye defects (Gorivodsky *et al.*, 2009; Yang *et al.*, 2014).

Of note, the two unrelated patients having variants in *FAT1* and *NDST1* shared a large set of specific secondary clinical features, including mandibular and ear abnormalities. Intrauterine growth restriction was found exclusively in the two patients with *COL2A1* variants. The most severely affected child in Family F16 (*FAT1/NDST1/COL2A1*) presented a strong overlap with *NDST1*-null and *COL2A1*-null mutant mice (HPE, mandibular anomalies, absent olfactory bulb, abnormal nose morphology) (Grobe *et al.*, 2005; Leung *et al.*, 2010). Similarly, proboscis and eye defects were observed in both *FAT1/NDST1/SHH* patient and *FAT1*<sup>-/-</sup> mice (Ciani *et al.*, 2003).

Finally, the two unrelated *SCUBE2/BOC* cases in Families F4 and F22 presented with cebocephaly, a midline facial anomaly characterized by ocular hypotelorism and a single nostril, which was absent in all other patients. Consistently, *SCUBE2* is highly expressed in the nasal septum in mouse (Xavier and Cobourne, 2011), and cebocephaly was previously associated with *CDON*—another known HPE gene sharing highly similar functions and structure with *BOC* (Zhang *et al.*, 2006).

While these clinical features are not specific to HPE, the described overlaps provide additional support for disease implication of the presented candidate variants.

## Statistical validations

The identified oligogenic events were clustered among 19 genes (Fig. 2, Tables 2 and 3). To assess the frequency of

**Table 4 Statistical validations: Fisher's exact test analysis for oligogenic events**

Comparison	HPE	GoNL	FREX	P-value		
				HPE versus GoNL	HPE versus FREX	GoNL versus FREX
Families with oligogenic events	10/26 (38%)	3/248 (1.2%)	NA	$2.301 \times 10^{-9}$	NA	NA
Children harbouring rare deleterious variants in two or more candidate genes	13/29 (45%)	6/248 (2.4%)	NA	$1.902 \times 10^{-10}$	NA	NA
All individuals harbouring rare deleterious variants in two or more candidate genes	21/80 (26%)	14/744 (1.8%)	16/574 (2.7%)	$3.237 \times 10^{-14}$	$1.521 \times 10^{-11}$	0.35

Oligogenic inheritance is defined as presence of combined rare deleterious variants in two or more genes, described in Table 3 and Fig. 2. The proportion of individuals harbouring combined rare deleterious variants in the identified genes is significantly higher in HPE cohort as compared to two control populations GoNL and FREX (Fisher's exact test).

healthy individuals presenting similar variant combinations in these genes, we applied the same family-by-family variant analysis to the 248 control trios provided by GoNL. This control cohort was chosen as 24/26 (92%) of the HPE families included in the study were of European descent (Supplementary Fig. 10).

The approach identified three families among controls presenting variant combinations satisfying the criteria that we established for the oligogenic events (gene, variant and parental inheritance). The three oligogenic events found in the control cohort were *FAT1/B9D1*, *SCUBE2/PTCH1* and *SCUBE2/LRP2/PTCH1/CELSR1* (Supplementary Table 9). Although one *SCUBE2* variant (p.Thr285Met) was found in both the HPE and the control cohort, none of the combinations found among controls corresponded to oligogenic events identified in the HPE cohort. The incidence of oligogenic events was significantly lower in the GoNL families (3/248, 1.2%) as compared to the HPE cohort (10/26, 38%) with a Fisher's exact test *P*-value of  $2.301 \times 10^{-9}$  (Table 4).

Three additional children of the GoNL cohort harboured combinations of rare deleterious variants in two or more candidate genes. However, in these cases, all variants were inherited from the same parent. Therefore, these combinations were not considered as oligogenic events similar to those of HPE patients. Nevertheless, even when taking into account these three additional cases, the proportion of children having variants in two or more candidate genes was significantly different between the HPE cohort (13/29, 45%) and the GoNL cohort (6/248, 2.4%) with a Fisher's exact test *P*-value of  $1.902 \times 10^{-10}$ .

Finally, 14 individuals of the GoNL cohort (parents and children combined) harboured rare deleterious variants in two or more genes. Without taking into account the relatedness between the GoNL individuals, the proportion of individuals having variants in two or more candidate genes remained significantly different between the HPE cohort (21/80, i.e. 26%) and the GoNL control cohort (14/744, 1.8%), as confirmed by Fisher's exact test (*P*-value =  $3.237 \times 10^{-14}$ ).

To assess the frequency of control individuals presenting rare variant combinations in the identified candidate genes

further (Fig. 2, Tables 2 and 3), we analysed a second control cohort. The FREX data were chosen as they consist of 574 unrelated French individuals ancestrally matching the HPE cohort.

Screening of the FREX cohort revealed that 16/574 individuals (i.e. 2.7%) harboured rare deleterious variants in two or more candidate genes. This proportion was statistically different from that observed in the HPE cohort (21/80, 26% versus 16/574, 2.7%; *P*-value =  $1.521 \times 10^{-11}$ , Fisher's exact test).

Additionally, the two control cohorts (GoNL and FREX) did not present statistically significant differences in terms of proportions of individuals having rare deleterious variants in two or more candidate genes: 14/744 (1.8%) for the GoNL cohort versus 16/574 (2.7%) for the FREX (*P*-value = 0.35, Fisher's exact test).

The analysis of the GoNL and FREX cohorts illustrates that the incidence of combined rare deleterious variants in the identified candidate genes is significantly higher in HPE patients as compared to a control population. All performed comparisons showed a statistically significant *P*-value between the cases and the controls (Table 4), thus providing evidence for oligogenicity as clinically relevant model in HPE.

## Discussion

In this study, we addressed the relevance of oligogenic model for unsolved HPE cases. We provide evidence that the onset of HPE arises from the combined effects of hypomorphic variants in several genes belonging to critical biological pathways of brain development. To circumvent the limitations of classical WES analysis in complex rare disorders, we combined clinically-driven and co-expression network analyses with classical WES variant prioritization. This strategy was applied to 26 HPE families and allowed prioritization of 180 genes directly linked to the SHH signalling, cilium and Wnt/PCP pathways (Fig. 3). The analysis of oligogenic events in patients with HPE anomalies revealed 19 genes including 15 genes previously unreported in human HPE patients (Tables 2 and 3). All these genes

are either associated with HPE phenotypes in corresponding mouse models (such as *FAT1*, *NDST1*), present highly similar expression patterns with already known HPE genes in the developing brain (such as *SCUBE2*, *TCTN3*), or both. We observed co-occurrence of mutations in several gene pairs such as *FAT1/NDST1* and *SCUBE2/BOC*, which provides additional arguments towards their implication in HPE. The incidence of oligogenic combinations was significantly higher in HPE patients compared to the GoNL and FREX control populations. We additionally show that in-depth evaluation of secondary clinical features in patients with HPE anomalies and comparison to published mouse knockout models may provide additional arguments for the causality of candidate genes.

The main challenge in disease-gene discovery by WES is to identify disease-related variants among a large background of non-pathogenic polymorphisms (Bamshad *et al.*, 2011; MacArthur *et al.*, 2014). For example, the presented *FAT1* encodes a large protocadherin gene spanning over 139 kb in the human genome and presenting over 2000 missense variants with a minor allele frequency below 1% in the gnomAD database. Despite this high number of variations found in the general population, rare variants in *FAT1* were recently implicated in several genetic disorders including facioscapulohumeral dystrophy-like disease (Puppo *et al.*, 2015). Hence, correct interpretations and conclusions require extremely careful assessment of available biological and clinical knowledge.

To improve the pertinence of our study, we developed a strategy to restrict the potential candidates by targeting genes with biological and clinical arguments for their implication in the disease. Implication of a given gene in a disease is often supported by the similarity between the human pathology and the phenotype obtained in relevant animal models (MacArthur *et al.*, 2014). Accordingly, in this study, the main evidence of causality for candidate genes was that their disruption leads to clinically-defined HPE-related phenotypes in corresponding published mutant mouse models. Unlike other phenotypes, such as reduced body weight (Reed *et al.*, 2008), holoprosencephaly is a rare effect of gene knockout in mice as it is associated with <1% of knockout mice (as reported in the MGI database). Recent exome sequencing studies have applied similar phenotype-driven approaches to identify causal variants in monogenic disorders. Dedicated tools have been developed to that aim (Exomiser, Phive) (Smedley *et al.*, 2015) but none are designed for non-Mendelian traits involving hypomorphic variants with mild effects. We provide a method to specifically address such cases and show that further developments are necessary to improve the diagnosis of genetic disorders, especially by taking into account oligogenic inheritance. Inclusion of carefully defined mouse mutant phenotypes is of powerful value as certain phenotypes like HPE are very informative due to their rarity.

Prioritization tools can also include protein–protein interaction (PPI) network information, which improves

performance in cases where candidate genes do not have an associated knockout mouse model. However, PPI-based prioritization is limited when disease investigation requires incorporation of tissue-specific data. The key process affected by HPE is the elaboration of the forebrain and its dorso-ventral patterning (Fernandes and Hébert, 2008). Deciphering the biological mechanisms involved in the early brain development is therefore necessary to provide relevant information to select disease-related genes. To incorporate tissue-specificity, we performed analysis using the RNA-Seq data of embryonic human brain at the earliest available developmental stages (from 4 to 17 post-conception weeks) as provided by the Human Development Biology Resource (Lindsay *et al.*, 2016). We defined relevant co-expression modules and selected candidate genes of which expression patterns follow those of known HPE genes. Further analysis showed that the resulting candidate genes, such as *SCUBE2* and *TCTN3*, are pertinent as they are equally implicated in the SHH pathway that is the primary HPE pathway (Thomas *et al.*, 2012; Jakobs *et al.*, 2014). Co-expression analysis provides additional insight into disease pathogenesis by establishing the first link between previously unrelated genes. A future challenge will be to generalize this approach, but such a task will face the necessity to incorporate disease relevant co-expression modules that need to be pre-computed.

Patients exhibiting HPE anomalies present enrichment of rare variants in genes related to the SHH pathway, as well as to the Wnt/PCP and primary cilia pathways, which were both shown to functionally interact with and regulate SHH pathway (Goetz *et al.*, 2009; Gorivodsky *et al.*, 2009; Murdoch and Copp, 2010; Wheway *et al.*, 2013). Accumulation of multiple rare variants in genes related to these pathways will likely disrupt the dorso-ventral gradient of the SHH morphogen (Fernandes and Hébert, 2008), leading to an incomplete cleavage of the forebrain and, ultimately, to HPE. In this model, distinction between different manifestations of HPE lies in the degree of overall functional impact on SHH signalling (Mercier *et al.*, 2013). Moreover, depending on the affected genes and pathways, HPE patients would present different secondary clinical features.

The observed overlapping secondary clinical features further support the causality of the reported variants for HPE. As hypomorphic mutations do not have the same impact as the complete inactivation of a gene in most cases, phenotypic overlaps may be challenging to detect and require expert assessment of clinical and biological data. For example, mice deficient in *NDST1* exhibit agnathia (Grobe *et al.*, 2005) (absence of the lower jaw) while unrelated patients presenting candidate variants in *NDST1* exhibit prognathia and retrognathia (abnormal positioning of the lower jaw), respectively. All three phenotypes are part of the same spectrum of mandibular anomalies. From a clinical perspective, overlap of secondary clinical features between the patient and the animal models provides additional critical evidence of a causal relationship between

candidate gene and disease. A key issue here remains the semantic representation of patient's phenotype and the use of a well-established phenotypic ontology during the examination processes. Explorations of secondary clinical features should be performed in future studies of genetic diseases.

Additional molecular screenings in larger populations of HPE patients are necessary to definitely assess the implication of our candidate genes in the disease. Therefore, we propose to include these novel genes into future genetic screenings of HPE patients.

In conclusion, this paper presents novel genes implicated in HPE and illustrates that HPE presents an oligogenic inheritance pattern requiring the joint effect of multiple genetic variants acting as hypomorphic mutations. The proposed inheritance pattern accounts for a wide clinical spectrum of HPE and explains the significant part of cases in which no molecular diagnosis could be established by conventional approaches. It also explains the incomplete penetrance and variable expressivity of inherited causal mutations observed in the reported cases of HPE (Mercier *et al.*, 2011). We propose that in cases of non-Mendelian diseases with variable phenotypes, the possibility of oligogenic inheritance needs to be evaluated. Exploration of such events will improve the diagnostic yield of complex developmental disorders and will contribute to better understanding of the mechanisms that coordinate normal and pathological embryonic development.

## Acknowledgements

We would like to thank the families for their participation in the study, all clinicians who referred HPE cases, the eight CLAD (Centres Labellisés pour les Anomalies du Développement) within France that belong to FECLAD, French centers of prenatal diagnosis (CPDPN) and the SOFFOET for foetal cases, and the 'filière AnDDI-Rares'. We particularly thank all members of the Molecular Genetics Laboratory (CHU, Rennes) and of the Department of Genetics and Development (UMR6290 CNRS, Université Rennes 1) for their help and advice.

## Funding

This work was supported by Fondation Maladie Rares (grant PMO1201204), Agence Nationale de la Recherche (grant ANR-12-BSV1-0007-01) and the Agence de la Biomedecine (AMP2016). This work was supported by La Fondation Maladie Rares and the Agence de la Biomedecine. The authors acknowledge the Centre de Ressources Biologiques (CRB)-Santé (<http://www.crbsante-rennes.com>) of Rennes for managing patient samples. This Work was supported by France Génomique National infrastructure, funded as part of "Investissement d'avenir" program managed by Agence Nationale pour la Recherche

(contrat ANR-10-INBS-09) <https://www.france-genomique.org/spip/spip.php?article158>. This study makes use of data generated by the Genome of the Netherlands Project. Funding for the project was provided by the Netherlands Organization for Scientific Research under award number 184021007, dated July 9, 2009 and made available as a Rainbow Project of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL). Samples were contributed by LifeLines (<http://lifelines.nl/lifelines-research/general>), The Leiden Longevity Study (<http://www.healthy-ageing.nl>; <http://www.langleven.net>), The Netherlands Twin Registry (NTR: <http://www.tweelinge-register.org>), The Rotterdam studies, (<http://www.erasmus-epidemiology.nl/rotterdamstudy>) and the Genetic Research in Isolated Populations program (<http://www.epib.nl/research/geneticipi/research.html#gip>). The sequencing was carried out in collaboration with the Beijing Institute for Genomics (BGI).

## Competing interests

The authors report no competing interests.

## Supplementary material

Supplementary material is available at *Brain* online.

## Appendix I

Full collaborator details are available in the online Supplementary material.

## The FREX Consortium

Emmanuelle Génin, Dominique Campion, Jean-François Dartigues, Jean-François Deleuze, Jean-Charles Lambert, Richard Redon.

### Bioinformatics group

Thomas Ludwig, Benjamin Grenier-Boley, Sébastien Letort, Pierre Lindenbaum, Vincent Meyer, Olivier Quenez.

### Statistical genetics group

Christian Dina, Céline Bellenguez, Camille Charbonnier-Le Clézio, Joanna Gierza.

### Data collection

Stéphanie Chatel, Claude Férec, Hervé Le Marec, Luc Letenneur, Gaël Nicolas, Karen Rouault.

### Sequencing

Delphine Bacq, Anne Boland, Doris Lechner.

## Genome of the Netherlands Consortium

### Steering group

Cisca Wijmenga, Morris A. Swertz, P. Eline Slagboom, Gert-Jan B. van Ommen, Cornelia M. van Duijn, Dorret I. Boomsma, Paul I.W. de Bakker

### Ethical, legal, and social issues

Jasper A. Bovenberg

### Cohort collection and sample management

P. Eline Slagboom, Anton J.M. de Craen, Marian Beekman, Albert Hofman, Dorret I. Boomsma, Gonneke Willemsen, Bruce Wolffenbuttel, Mathieu Platteel.

### Sequencing

Yuanping Du, Ruoyan Chen, Hongzhi Cao, Rui Cao, Yushen Sun, Jeremy Sujie Cao.

### Analysis group

Morris A. Swertz, Freerk van Dijk, Pieter B.T. Neerinx, Patrick Deelen, Martijn Dijkstra, George Byelas, Alexandros Kanterakis, Jan Bot, Kai Ye, Eric-Wubbo Lameijer, Martijn Vermaat, Jeroen F.J. Laros, Johan T. den Dunnen, Peter de Knijff, Lennart C. Karssen, Elisa M. van Leeuwen, Najaf Amin, Vyacheslav Koval, Fernando Rivadeneira, Karol Estrada, Jayne Y. Hehir-Kwa, Joep de Ligt, Abdel Abdellaoui, Jouke-Jan Hottenga, V. Mathijs Kattenberg, David van Enckevort, Hailiang Mei, Mark Santcross, Barbera D.C. van Schaik, Robert E. Handsaker, Steven A. McCarroll, Evan E. Eichler, Arthur Ko, Peter Sudmant, Laurent C. Francioli, Wigard P. Kloosterman, Isaac J. Nijman, Victor Guryev, Paul I.W. de Bakker.

## References

- Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011; 12: 745–55.
- Bear KA, Solomon BD, Antonini S, Arnhold IJP, França MM, Gerkes EH, et al Pathogenic mutations in *GLI2* cause a specific phenotype that is distinct from holoprosencephaly. *J Med Genet* 2014; 51: 413–18.
- Briggs KJ, Corcoran-Schwartz IM, Zhang W, Harcke T, Devereux WL, Baylin SB, et al Cooperation between the *Hic1* and *Ptch1* tumor suppressors in medulloblastoma. *Genes Dev* 2008; 22: 770–85.
- Carter MG. Mice deficient in the candidate tumor suppressor gene *Hic1* exhibit developmental defects of structures affected in the Miller-Dieker syndrome. *Hum Mol Genet* 2000; 9: 413–19.
- Christ A, Christa A, Kur E, Lioubinski O, Bachmann S, Willnow TE, et al LRP2 is an auxiliary SHH receptor required to condition the forebrain ventral midline for inductive signals. *Dev Cell* 2012; 22: 268–78.
- Ciani L, Patel A, Allen ND, French-Constant C. Mice lacking the giant protocadherin *mFAT1* exhibit renal slit junction abnormalities and partially penetrant cyclopia and anophthalmia phenotype. *Mol Cell Biol* 2003; 23: 3575–82.
- Dowdle WE, Robinson JF, Kneist A, Sirerol-Piquer MS, Frints SGM, Corbit KC, et al Disruption of a ciliary B9 protein complex causes meckel syndrome. *Am J Hum Genet* 2011; 89: 94–110.
- Dubourg C, Carré W, Hamdi-Rozé H, Moudén C, Roume J, Abdelmajid B, et al Mutational Spectrum in holoprosencephaly shows that FGF is a new major signaling pathway. *Hum Mutat* 2016; 37: 1329–39.
- Dubourg C, Kim A, Watrin E, de Tayrac M, Odent S, David V, et al Recent advances in understanding inheritance of holoprosencephaly. *Am J Med Genet C Semin Med Genet* 2018; 178: 258–69.
- Dupé V, Rochard L, Mercier S, Le Pétillon Y, Gicquel I, Bendavid C, et al NOTCH, a new signaling pathway implicated in holoprosencephaly. *Hum Mol Genet* 2011; 20: 1122–31.
- Fernandes M, Hébert JM. The ups and downs of holoprosencephaly: dorsal versus ventral patterning forces. *Clin Genet* 2008; 73: 413–23.
- Garcia-Gonzalo FR, Corbit KC, Sirerol-Piquer MS, Ramaswami G, Otto EA, Noriega TR, et al A transition zone complex regulates mammalian ciliogenesis and ciliary membrane composition. *Nat Genet* 2011; 43: 776–84.
- Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet* 2014; 46: 818–25.
- Goetz SC, Ocbina PJR, Anderson KV. The primary cilium as a hedgehog signal transduction machine. *Methods Cell Biol* 2009; 94: 199–222.
- Gorivodsky M, Mukhopadhyay M, Wilsch-Braeuning M, Phillips M, Teufel A, Kim C, et al Intraflagellar transport protein 172 is essential for primary cilia formation and plays a vital role in patterning the mammalian brain. *Dev Biol* 2009; 325: 24–32.
- Grobe K, Inatani M, Pallerla SR, Castagnola J, Yamaguchi Y, Esko JD. Cerebral hypoplasia and craniofacial defects in mice lacking heparan sulfate *Ndst1* gene function. *Development* 2005; 132: 3777–86.
- Hong M, Srivastava K, Kim S, Allen BL, Leahy DJ, Hu P, et al BOC is a modifier gene in holoprosencephaly. *Hum Mutat* 2017; 38: 1464–70.
- Jakobs P, Exner S, Schürmann S, Pickhinke U, Bandari S, Ortmann C, et al Scube2 enhances proteolytic Shh processing from the surface of Shh-producing cells. *J Cell Sci* 2014; 127: 1726–37.
- Kruszka P, Martinez AF, Muenke M. Molecular testing in holoprosencephaly. *Am J Med Genet C Semin Med Genet* 2018; 178: 187–93.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; 9: 559.
- Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, et al Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* 2014; 312: 1880–7.
- Leung AWL, Wong SYY, Chan D, Tam PPL, Cheah KSE. Loss of procollagen IIA from the anterior mesendoderm disrupts the development of mouse embryonic forebrain. *Dev Dyn* 2010; 239: 2319–29.
- Li L, Bainbridge MN, Tan Y, Willerson JT, Marian AJ. A potential oligogenic etiology of hypertrophic cardiomyopathy: a classic single-gene disorder. *Circ Res* 2017; 120: 1084–90.
- Lindsay SJ, Xu Y, Lisgo SN, Harkin LF, Copp AJ, Gerrelli D, et al HDBR expression: a unique resource for global and individual gene expression studies during early human brain development. *Front Neuroanat* 2016; 10: 86. <http://journal.frontiersin.org/article/10.3389/fnana.2016.00086/full>
- MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al Guidelines for investigating causality of sequence variants in human disease. *Nature* 2014; 508: 469–76.
- Mercier S, David V, Ratié L, Gicquel I, Odent S, Dupé V. NODAL and SHH dose-dependent double inhibition promotes an HPE-like phenotype in chick embryos. *Dis Model Mech* 2013; 6: 537–43.

- Mercier S, Dubourg C, Garcelon N, Campillo-Gimenez B, Gicquel I, Belleguic M, et al New findings for phenotype-genotype correlations in a large European series of holoprosencephaly cases. *J MedGenet* 2011; 48: 752–60.
- Mouden C, Dubourg C, Carré W, Rose S, Quelin C, Akloul L, et al Complex mode of inheritance in holoprosencephaly revealed by whole exome sequencing. *Clin Genet* 2016; 89: 659–68.
- Mouden C, Tayrac M de, Dubourg C, Rose S, Carré W, Hamdi-Rozé H, et al Homozygous STIL mutation causes holoprosencephaly and microcephaly in two siblings. *PLoS One* 2015; 10: e0117418.
- Murdoch JN, Copp AJ. The relationship between sonic hedgehog signalling, cilia and neural tube defects. *Birt Defects Res A Clin Mol Teratol* 2010; 88: 633–52.
- Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Brudno M, et al The matchmaker exchange: a platform for rare disease gene discovery. *Hum Mutat* 2015; 36: 915–21.
- Puppo F, Dionnet E, Gaillard M-C, Gaildrat P, Castro C, Vovan C, et al Identification of variants in the 4q35 gene FAT1 in patients with a facioscapulohumeral dystrophy-like phenotype. *Hum Mutat* 2015; 36: 443–53.
- Reed DR, Lawler MP, Tordoff MG. Reduced body weight is a common effect of gene knockout in mice. *BMC Genet* 2008; 9: 4.
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. *Genet Med* 2015; 17: 405–24.
- Rock R, Schrauth S, Gessler M. Expression of mouse *dchs1*, *fjx1*, and *fat-j* suggests conservation of the planar cell polarity pathway identified in *drosophila*. *Dev Dyn* 2005; 234: 747–55.
- Roessler E, Belloni E, Gaudenz K, Jay P, Berta P, Scherer SW, et al Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. *Nat Genet* 1996; 14: 357–60.
- Smedley D, Jacobsen JOB, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat Protoc* 2015; 10: 2004–15.
- Smith CL, Blake JA, Kadin JA, Richardson JE, Bult CJ, Mouse genome database group. Mouse genome database (MGD)-2018: knowledge-base for the laboratory mouse. *Nucleic Acids Res.* 2018; 46: D836–42.
- Stark Z, Dashnow H, Lunke S, Tan TY, Yeung A, Sadedin S, et al A clinically driven variant prioritization framework outperforms purely computational approaches for the diagnostic analysis of singleton WES data. *Eur J Hum Genet* 2017; 25: 1268–72.
- Thomas S, Legendre M, Saunier S, Bessières B, Alby C, Bonnière M, et al TCTN3 mutations cause Mohr-Majewski syndrome. *Am J Hum Genet* 2012; 91: 372–8.
- Wheway G, Abdelhamed Z, Natarajan S, Toomes C, Inglehearn C, Johnson CA. Aberrant Wnt signalling and cellular over-proliferation in a novel mouse model of Meckel–Gruber syndrome. *Dev Biol* 2013; 377: 55–66.
- Xavier GM, Cobourne MT. Scube2 expression extends beyond the central nervous system during mouse development. *J Mol Histol* 2011; 42: 383–91.
- Yang T, Jia Z, Bryant-Pike W, Chandrasekhar A, Murray JC, Fritsch B, et al Analysis of PRICKLE1 in human cleft palate and mouse development demonstrates rare and common variants involved in human malformations. *Mol Genet Genomic Med* 2014; 2: 138–51.
- Zhang W, Kang J-S, Cole F, Yi M-J, Krauss RS. Cdo functions at multiple points in the Sonic Hedgehog pathway, and Cdo-deficient mice accurately model human holoprosencephaly. *Dev Cell* 2006; 10: 657–65.

## 2. Additional candidates

Several additional variants presented clinical and functional evidence of their disease implication but have not been included in the main article.

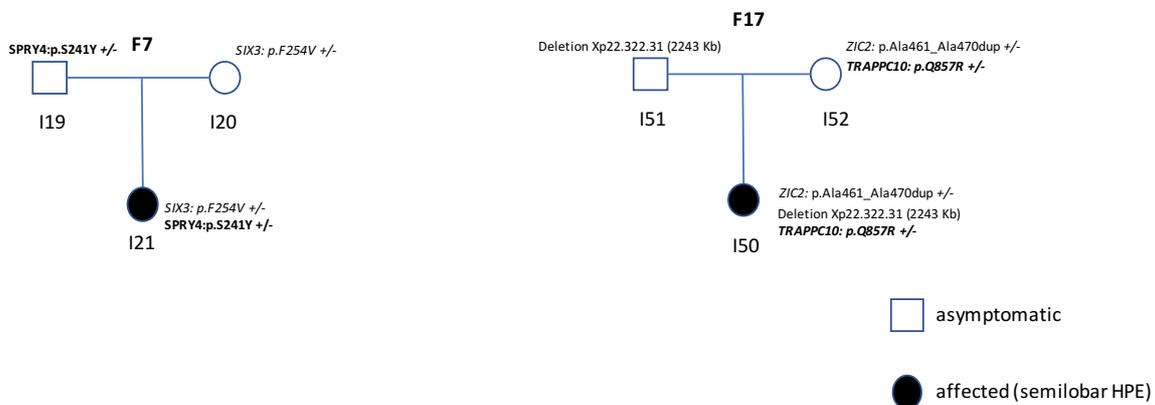
Family F17 presented two unaffected parents and one child affected with semilobar HPE associated with eye defects (hypotelorism) and microcephaly. Initial molecular findings included a variant in known disease gene *ZIC2* (p.Ala461\_Ala470dup) and a 2243 kb deletion at chromosome X (Xp22.322.31) detected by CGH-array. As both anomalies were inherited from asymptomatic parents, they were considered insufficient to explain the disease phenotype in this family (Figure 24). Our integrative approach identified a missense variant in *TRAPPC10*, which was inherited from the mother. The variant is 'novel', *i.e.*, absent in all control databases, and results in a substitution of a highly conserved nucleotide and aminoacid residue, as predicted by GERP and SIFT algorithms, respectively. *TRAPPC10* is also called Epilepsy-Holoprosencehaly Candidate 1 (*EHOC-1*) as it maps to chromosomal region 21q22.3, considered as candidate region for HPE (Yamakawa et al., 1995). This gene encodes a protein involved in vesicular transport and mice mutants exhibit eye and craniofacial defects such as anophthalmia and HPE, according to the MGI database. *TRAPPC10* has been previously considered as candidate gene for HPE and intellectual disability (Savastano et al., 2014; Santos-Cortez et al., 2018). Despite their presence in unaffected parents, the combined effect of variants in *ZIC2*, *TRAPPC10* and Xp22 deletion remains unknown.

Family F7 comprised two unaffected parents and a child affected by semilobar HPE. Initial screening identified a missense variant in known disease gene *SIX3* (p.F254V), which was inherited from asymptomatic mother and insufficient to explain the disease phenotype. Our analysis identified a rare deleterious variant in *SPRY4*, inherited from asymptomatic father. The variant is present in the gnomAD database with a total frequency of 0.4 % and predicted deleterious by 14 algorithms including SIFT, GERP and Polyphen2. In zebrafish and mouse models, dosage of *SPRY4* has been shown to influence disease-related FGF and SHH signaling (Fürthauer et al., 2001; Taniguchi et

al., 2007; Lochovska et al., 2015). In mice, double knockout of *SPRY4* and a closely related gene *SPRY2* results in embryonic lethality associated with severe craniofacial defects including cyclopia and HPE, indicating the essential role of *SPRY* genes in early brain morphogenesis (Taniguchi et al., 2007). Overall, these results suggest that the combined effect of inherited variants in *SPRY4* and *SIX3* could participate in the disease pathogenesis, possibly under digenic inheritance.

Due to the lack of additional evidence for *TRAPPC10* and *SPRY4* (mutational recurrence, variant enrichment in relevant biological pathways, phenotypic overlaps), these families have not been included in the main article. Nevertheless, *TRAPPC10* and *SPRY4* represent candidate genes for HPE and should be investigated in future studies.

Family	Gene	Mutation	Inheritance	Frequency	Candidate	Nature, deleteriousness
F7	<i>SPRY4</i>	c.722C>A (p.S241Y)	Paternal	0.004	Clinical	Missense, 14 deleterious predictions
	<i>SIX3</i>	c.760T>G (p.F254V)	Maternal	NA	Clinical	Missense, 16 deleterious predictions
F17	<i>ZIC2</i>	c.1377_1406dup (p.Ala461_Ala470dup)	Maternal	0.0000226	Clinical	Known pathogenic variant (ClinVar)
	<i>TRAPPC10</i>	c.2570A>G (p.Q857R)	Maternal	NA	Clinical	Missense, 12 deleterious predictions
	Xp22.322.31del	-	Paternal	NA	CGH-array	-



**Figure 24. Additional candidates.**

### 3. The role of HSPGs in SHH signaling

Among the oligogenic combinations reported in the article, two of them involved variants in the *NDST1* gene. *NDST1* interacts genetically with *SHH* as demonstrated by mouse studies: *Ndst1* knockout mutants exhibit reduced Shh signalling in the forebrain and certain *Ndst1*<sup>+/-</sup>; *Shh*<sup>+/-</sup> compound heterozygous mice exhibit craniofacial defects

reminiscent of HPE (Grobe et al., 2005). *NDST1* encodes a N-acetylglucosamine N-deacetylase-N-sulfotransferase enzyme involved in synthesis of heparan sulfates (HS) and plays a crucial role in synthesis of HSPGs (Pan et al., 2006). As discussed in the Introduction, HSPGs play a role in the release of SHH from the secreting cell. Moreover, HSPG are involved in the reception of SHH by PTCH1. Specifically, SHH interacts with HSPG via a series of positively charged aminoacids (known as Cardin-Weinberg motif) which bind to the negatively charged HS and enhance the SHH-PTCH1 interaction. HSPGs have been shown to act as essential regulators of SHH function in flies and mice (Desbordes and Sanson, 2003; Grobe et al., 2005). Moreover, HSPG are involved in the regulation of other HPE-related pathways. HSPG play an essential role in regulating FGF signaling by acting as FGF co-receptors during embryonic development (Lin et al., 1999; Yayon et al., 1991). *EXT1*, a gene involved in HSPG biosynthesis similar to *NDST1*, is required for *FGF8* function during brain development (Inatani et al., 2003). Glypican 3 (*GPC3*), encoding a HSPG core-protein, interacts with BMP pathway during limb and skeletal development, while mice deficient for *Gpc3* exhibit loss of Wnt signaling (Paine-Saunders et al., 2000; Song et al., 2005). As HSPGs regulate multiple signaling pathways implicated in HPE, mutations in HSPG genes could contribute to the disease etiology. Additional studies are, however, needed to elucidate this hypothesis.

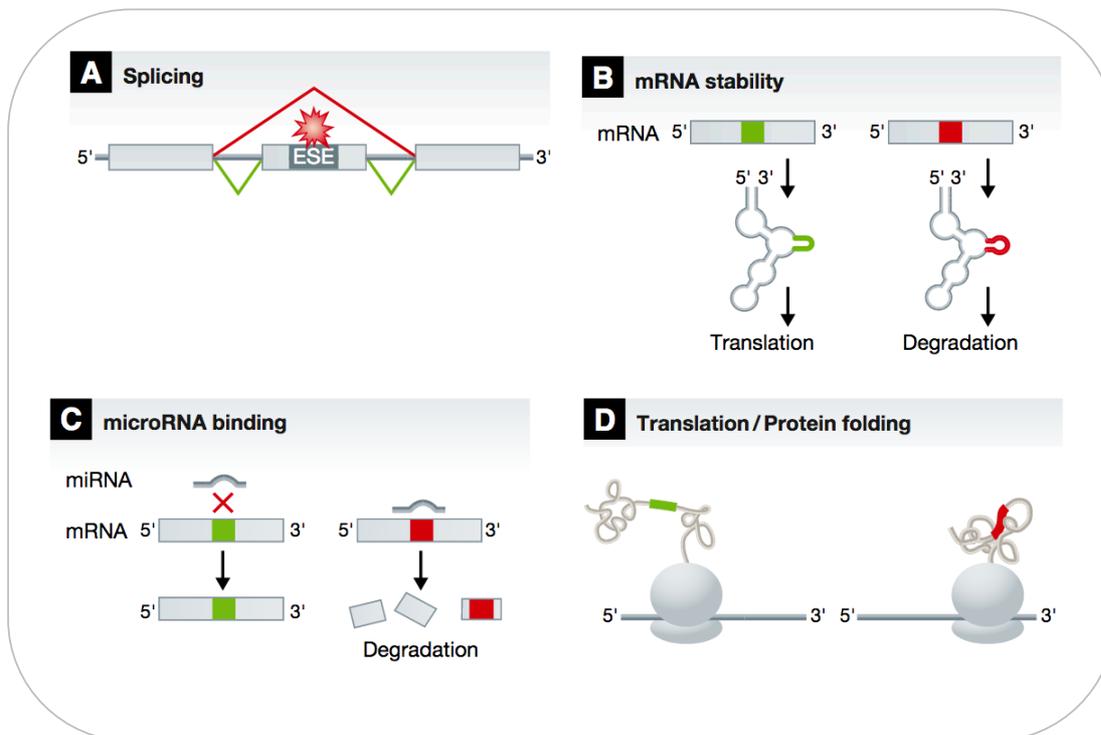
## II. Pathogenic impact of synonymous variants in *SHH* gene

### a. Background

Due to the degeneracy of the genetic code, synonymous mutations, also called synonymous single nucleotide variants (sSNVs), result in substitution of a protein-coding nucleotide without changing the amino acid composition of the resulting protein. Such variants have long time been referred to as 'silent', *i.e.*, without any protein- or phenotype-altering effect. Historically, sSNVs have been considered neutral, *i.e.*, not affected by natural selection and resulting from random genetic drift (Kimura, 1977). In evolutionary biology, approaches for detection of sites under natural selection at the DNA level, such as dN/dS and McDonald-Kreitman methods, have relied on assumption that all sSNVs are neutral (Yang 1998, McDonald and Kreitman 1991). Nonetheless, if the synonymous substitutions are strictly neutral, then codons encoding the same amino acid - referred to as synonymous codons - would appear randomly and equally distributed along the genes. This is not the case, as synonymous codons are not used at equal frequencies in protein coding regions, a phenomenon called codon usage bias (Table 6). The non-random use of synonymous codons is now recognized as a crucial process involved in fine-tuning of gene expression and protein function (Plotkin and Kudla, 2011). Consequently, sSNVs can influence protein biogenesis and phenotype and, thus, contribute to human pathologies (Sauna and Kimchi-Sarfaty, 2011). A recent survey of 21 429 variants statistically associated with human disease determined that nonsynonymous and synonymous variants have a similar probability of disease association (Hunt et al., 2014). Overall, synonymous variants have now been implicated in over 50 genetic disorders (Chamary et al., 2006). Their clinical impact can result from different types of molecular alterations including aberrant splicing, modification of miRNA binding sites, changes in mRNA structure and translation dynamics (Figure 25).

Codon	AA	Fraction									
UUU	F	0.46	UCU	S	0.19	UAU	Y	0.44	UGU	C	0.46
UUC	F	0.54	UCC	S	0.22	UAC	Y	0.56	UGC	C	0.54
			UCA	S	0.15	UAA	*	0.30	UGA	*	0.47
UUA	L	0.08	UCG	S	0.05	UAG	*	0.24	UGG	W	1.00
UUG	L	0.13									
CUU	L	0.13	CCU	P	0.29	CAU	H	0.42	CGU	R	0.08
CUC	L	0.20	CCC	P	0.32	CAC	H	0.58	CGC	R	0.18
CUA	L	0.07	CCA	P	0.28	CAA	Q	0.27	CGA	R	0.11
CUG	L	0.40	CCG	P	0.11	CAG	Q	0.73	CGG	R	0.20
AUU	I	0.36	ACU	T	0.25	AAU	N	0.47	AGU	S	0.15
AUC	I	0.47	ACC	T	0.36	AAC	N	0.53	AGC	S	0.24
AUA	I	0.17	ACA	T	0.28	AAA	K	0.43	AGA	R	0.21
AUG	M	1.00	ACG	T	0.11	AAG	K	0.57	AGG	R	0.21
GUU	V	0.18	GCU	A	0.27	GAU	D	0.46	GGU	G	0.16
GUC	V	0.24	GCC	A	0.40	GAC	D	0.54	GGC	G	0.34
GUA	V	0.12	GCA	A	0.23	GAA	E	0.42	GGA	G	0.25
GUG	V	0.46	GCG	A	0.11	GAG	E	0.58	GGG	G	0.25

**Table 6. Synonymous codon usage bias in *Homo sapiens*.** For each codon, the corresponding aminoacid (AA) and the relative usage frequency in protein-coding sequences (Fraction) are represented. Codon usage frequencies were retrieved from the Kazusa database (Nakamura et al., 2000). \* - stop codon



**Figure 25. Possible mechanisms of sSNV impact on biological function.**

Adapted from Diederichs et al., 2016

### *Splicing*

The first molecular mechanism that explained the clinical impact of synonymous mutations was disruption of the spliceosome: disease-associated sSNVs have been shown to alter existing splicing-control elements resulting in aberrant mRNA splicing and leading to human disease (Cartegni et al., 2002). Such sSNVs can alter a consensus splice site (typically, GT/AG sequence motifs) located near exon-intron boundaries or auxiliary elements such as Exonic Splicing Silencers and Enhancers (ESS/ESE). Additionally, a sSNV can activate a 'dormant' cryptic splice site located within the exon. Such changes lead to incorrect recognition of the splice sites, resulting in aberrantly-spliced transcripts and non-functional proteins. Common disease-associated splicing defects include complete/partial exon skipping (Chao et al., 2001; De Meirleir et al., 1994; Liu et al., 1997) and, rarely, intron retention (Yadegari et al., 2016).

### *mRNA structure*

The mRNA forms a complex molecule with various structural features that can influence the resulting protein output (Faure et al., 2016). In particular, mRNA contains secondary structure elements, such as stems and loops, of varying thermodynamic stability that can modulate protein expression. Altered stability of the mRNA structure can lead to incorrect folding and degradation of the mRNA, resulting in lower protein levels (Sauna and Kimchi-Sarfaty, 2011). Several studies illustrated that sSNVs can distort the secondary structure of mRNA, resulting in mRNA misfolding and pathogenic consequences. A recent example of such impact concerns the  $\Delta$ F508 mutation in the *CFTR* gene.  $\Delta$ F508 is the most common disease-causing variant of cystic fibrosis causing a reduced expression of the resulting protein.  $\Delta$ F508 corresponds to a three-nucleotide deletion (CTT) resulting in synonymous substitution at isoleucine 507 (ATC > ATT) and the deletion of phenylalanine at position 508 (Figure 26). For a long time, the research has been focused on the consequences of the phenylalanine deletion, until a recent study demonstrated that the synonymous ATC>ATT substitution alters the secondary structure of the *CFTR* mRNA, leading to mRNA misfolding (Bartoszewski et al., 2010). Moreover, keeping the phenylalanine 508 deletion but restoring the original

synonymous codon of the Isoleucine 507 leads to correct mRNA folding and higher protein levels. These results indicate that the disease-causing impact of  $\Delta F508$  is a consequence of the synonymous isoleucine substitution rather than the phenylalanine deletion.



**Figure 26. Impact of synonymous isoleucine variant in  $\Delta F508$  mutation.**

Adapted from Bartoszewski et al., 2010

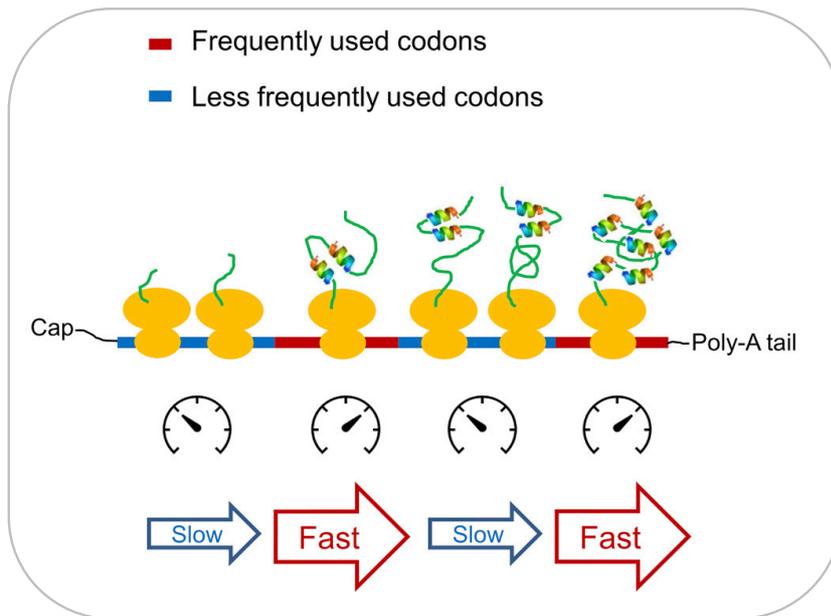
#### *miRNA-based regulation*

sSNVs can also impact the post-transcriptional regulation of gene expression by modifying the binding sites of its miRNA regulators. A common sSNV in *IRGM* has been associated with increased risk of Crohn's disease (McCarroll et al., 2008). Subsequent study illustrated that this sSNV alters the sequence of the binding site for miR-196, resulting in altered expression of *IRGM* and deregulation of *IRGM*-dependent xenophagy (Brest et al., 2011).

#### *Codon-mediated translation dynamics*

The codon usage bias has been largely linked to the abundance of the corresponding transfer RNA (tRNA) molecules and has been shown to influence the translational optimality: rare codons are serviced by low-abundance tRNAs and tend to slow down the translation, while frequently used codons are recognized by abundant tRNAs and result in faster translation rate (Sauna and Kimchi-Sarfaty, 2011; Buhr et al., 2016; Rogozin et al., 2018). Specific combinations of rare and frequent codons in a protein-coding sequence modulate the translational speed and protein folding (Quax et al., 2015) (Figure 27). Consequently, the impact of synonymous codon changes on translational dynamics has been suggested. Depending on the codon change and the

positional context, sSNVs can alter the translational efficiency, influencing the amount of protein produced per mRNA and folding of the nascent protein (Rodnina, 2016). Kimchi-Sarfaty and colleagues were the first to pay attention to that in clinical context with their discovery that a synonymous polymorphism in the Multidrug Resistance gene (*MDR1/ABCB1*) introduces a very rare codon instead of a frequent one resulting in altered conformation of the produced protein (Kimchi-Sarfaty et al., 2007).



**Figure 27. Impact of codon usage on translational dynamics.**

Yu et al., 2015

## Objective

Although clinically relevant, sSNVs are still largely overlooked in studies of genetic disorders, unless directly associated with splicing defects. The under-prioritization of sSNVs is mainly due to the lack of reliable *in-silico* prioritization and simply the vast number of such variants present in every individual's genome. However, the pathogenic impact of sSNVs can be explored for a particular gene already known to be causative for the disease phenotype. In this study, we investigated the contribution of sSNVs in *SHH* to the disease mechanisms of HPE.

## B. Strategy

We performed a retrospective analysis of the sequencing data of 931 patients diagnosed with HPE. The screening identified eight sSNVs in the *SHH* gene, initially discarded as benign during molecular diagnosis. As sSNVs can affect protein and cell functions via a variety of mechanisms mentioned above, assessing their pathogenic impact requires a multi-hypothesis approach. In this study, we performed a series of *in silico* and *in vitro* analyses to investigate the impact of the identified sSNVs at both the mRNA and the protein level.

### *Splicing-related defect*

*In silico* prediction tools are widely used in clinical genetics to assess the impact of variants on splicing. The basic premise of such tools is to determine whether a given variant alters an existing splice site or creates a new one. Commonly used algorithms for splice site recognition, such as SSF-like, employ Position Weight Matrices (PWM), which are based on nucleotide sequences of experimentally validated splice sites (Shapiro and Senapathy, 1987). Specifically, sequences of experimentally validated splice sites are aligned to generate a consensus motif and a count matrix. The count matrix is used to algorithmically scan the genome and identify sequences matching the splice motif. Sequences with high matching scores are considered as candidate splice sites. Consequently, if a variant falls within the candidate splice site it is predicted to alter the splicing pattern. Tools such as MaxEntScan use maximum entropy principle to additionally take into account the dependencies between nucleotide positions of the splicing motif (Yeo and Burge, 2004). Several tools, such as SPiCe and Human Splicing Finder (HSF), integrate multiple methods into a statistical framework to derive a single splicing pathogenicity score (Leman et al., 2018; Desmet et al., 2009).

Splicing impact can be further investigated *in vitro*, using the minigene approach (Soukariéh et al., 2016). In this approach, the variant-containing exon and the adjacent intronic regions (typically the first 200–300 nucleotides of the 5' and 3' flanking introns) are cloned within a plasmid vector containing a transcriptional promoter and

additional segments necessary for mRNA formation. The resulting vector is used to investigate the variant impact on the splicing outcome by comparing the PCR products of the wildtype and the mutant constructs.

#### *mRNA structure*

Commonly used algorithms for RNA secondary structure prediction are based on the Minimum Free Energy (MFE) method (Mathews et al., 1999; Mathews and Turner, 2002). For a given mRNA sequence, every possible secondary structure has a global free energy ( $\Delta G$ ) depending on the nucleotide composition and pre-determined energy parameters. The basis behind the MFE approach is that the native secondary structure is the most optimal and thermodynamically stable, with the minimum global free energy. Tools such as mFold and RNAfold use dynamic programming to compute the minimum  $\Delta G$  and predict an optimal secondary structure of a given mRNA sequence (Zucker, 2003; Lorenz et al., 2011). Consequently, it is possible to estimate an impact of a given variant on mRNA structure by comparing the predictions for the wildtype and the mutated sequences. In case of observed differences in predicted structures or high variations of estimated minimum free energy ( $\Delta\Delta G$ ) between the wildtype and the mutant sequence, the variant can be predicted to impact on mRNA folding (Bartoszewski et al., 2010; Wang et al., 2015).

#### *miRNA - RegRNA*

A variety of computational programs exist to identify regulatory RNA motifs in the genome. Tools such as RegRNA process an input DNA sequence and identify candidate regulatory elements matching the experimentally validated regulatory RNA motifs extracted from scientific literature and databases (Huang et al., 2006). RegRNA integrates several computational tools of motif identification, such as miRanda and GeneSplicer (Enright et al., 2003; Pertea et al., 2001). Consequently, if a variant alters a predicted regulatory motif it can disrupt the miRNA-mediated regulation.

#### *Codon usage and translational output*

As mentioned above, synonymous codon changes have been shown to impact the translational dynamics. In this study, we used multiple *in silico* algorithms to measure

the impact of *SHH* sSNVs on codon usage and estimate their impact on codon-mediated translational dynamics. The *in-silico* results were verified by experimentally measuring the quantity of the produced SHH protein *in vitro*. Computational and *in vitro* approaches used in our study are fully described in the resulting paper (part C). The *in-silico* investigation of impact of *SHH* sSNVs on codon usage has been done in collaboration with Luis Diambra from Regional Center of Genomic Studies (CREG, University of La Plata, Argentina).

## c. Results.

### 1. Article III.

---

#### Synonymous variants in holoprosencephaly alter codon usage and impact the Sonic Hedgehog protein.

Artem Kim, Jerome Le Douce, Farah Diab, Monika Ferovova, Christele Dubourg, Sylvie Odent, Valérie Dupé, Véronique David, Luis Diambra, Erwan Watrin, Marie de Tayrac<sup>1</sup>

1 - corresponding author: [marie.detayrac@univ-rennes1.fr](mailto:marie.detayrac@univ-rennes1.fr)

Brain, July 2020

---

Our retrospective analysis of the diagnostic data identified eight different sSNVs in the *SHH* sequence significantly enriched in HPE patients as compared to control populations (gnomAD, FReX, GoNL), suggesting their role in the disease etiology. We further demonstrated that five out of the eight identified sSNVs are associated with a significant reduction of protein levels, up to 23 %. Computational and functional validations excluded splicing-, mRNA folding- and miRNA-related defects but indicate that the protein reduction results from alterations of codon-mediated translational control, leading to protein misfolding and degradation. Considering the observed protein reduction and the critical role of SHH morphogen in early brain development, our findings underline the clinical relevance of *SHH* sSNVs in HPE and the necessity for deeper investigation of ‘silent’ substitutions in complex genetic disorders. Moreover, we observed that the experimental measures of protein reduction are significantly correlated with computational variations of codon usage, indicating the relevance of certain *in silico* models for predicting the impact of sSNVs on translation. Our study thus provides an analytical framework which can be applied in future investigations of sSNVs and their pathogenic impact on translation.

# Synonymous variants in holoprosencephaly alter codon usage and impact the Sonic Hedgehog protein

Artem Kim,<sup>1</sup> Jérôme Le Douce,<sup>1</sup> Farah Diab,<sup>1</sup> Monika Ferovova,<sup>1</sup> Christèle Dubourg,<sup>1,2</sup> Sylvie Odent,<sup>1,3</sup> Valérie Dupé,<sup>1</sup> Véronique David,<sup>1,2</sup> Luis Diambra,<sup>4</sup> Erwan Watrin<sup>1</sup> and Marie de Tayrac<sup>1,2</sup>

Synonymous single nucleotide variants (sSNVs) have been implicated in various genetic disorders through alterations of pre-mRNA splicing, mRNA structure and miRNA regulation. However, their impact on synonymous codon usage and protein translation remains to be elucidated in clinical context. Here, we explore the functional impact of sSNVs in the Sonic Hedgehog (*SHH*) gene, identified in patients affected by holoprosencephaly, a congenital brain defect resulting from incomplete forebrain cleavage. We identified eight sSNVs in *SHH*, selectively enriched in holoprosencephaly patients as compared to healthy individuals, and systematically assessed their effect at both transcriptional and translational levels using a series of *in silico* and *in vitro* approaches. Although no evidence of impact of these sSNVs on splicing, mRNA structure or miRNA regulation was found, five sSNVs introduced significant changes in codon usage and were predicted to impact protein translation. Cell assays demonstrated that these five sSNVs are associated with a significantly reduced amount of the resulting protein, ranging from 5% to 23%. Inhibition of the proteasome rescued the protein levels for four out of five sSNVs, confirming their impact on protein stability and folding. Remarkably, we found a significant correlation between experimental values of protein reduction and computational measures of codon usage, indicating the relevance of *in silico* models in predicting the impact of sSNVs on translation. Considering the critical role of *SHH* in brain development, our findings highlight the clinical relevance of sSNVs in holoprosencephaly and underline the importance of investigating their impact on translation in human pathologies.

- 1 Univ Rennes, CNRS, IGDR (Institut de génétique et développement de Rennes)—UMR 6290, F—35000 Rennes, France
- 2 Service de Génétique Moléculaire et Génomique, CHU, Rennes, France
- 3 Service de Génétique Clinique, CHU, Rennes, France
- 4 CREG, CONICET-Universidad Nacional de La Plata, La Plata, CP 1900, Argentina

Correspondence to: Marie de Tayrac  
IGDR CNRS UMR 6290, Campus santé de Villejean, 2 avenue du Professeur Léon Bernard  
CS 34317, France  
E-mail: marie.detayrac@univ-rennes1.fr

**Keywords:** brain development; genetics; clinical practice; synonymous variants; codon usage

**Abbreviations:** RSBU = relative synonymous bicodon usage; RSCU = relative synonymous codon usage; sSNV = synonymous single nucleotide variant

## Introduction

Synonymous single nucleotide variants (sSNVs) generate different molecular alterations at the mRNA and the protein level and, accordingly, can lead to pathologies by affecting mRNA splicing (Macaya *et al.*, 2009), structure (Bartoszewski *et al.*, 2010) and miRNA-based regulation (Brest *et al.*, 2011). The impact of sSNVs on translation has also been proposed (Kimchi-Sarfaty *et al.*, 2007; Sauna and Kimchi-Sarfaty, 2011; Poliakov *et al.*, 2014) but remains to be explored in clinical context.

The degeneracy of the genetic code allows most amino acids to be encoded by more than one codon. Codons encoding the same amino acid—referred to as synonymous codons—are not used at equal frequencies in protein coding regions, a phenomenon called codon usage bias (Quax *et al.*, 2015). The non-random use of synonymous codons is largely linked to the abundance of the corresponding tRNA molecules and has been shown to participate in translational optimality: rare codons, serviced by low-abundance tRNAs, tend to slow down the translation (Dana and Tuller, 2014), while frequently used codons are recognized by abundant tRNAs and result in a faster translation rate (Berg and Kurland, 1997; Gustafsson *et al.*, 2004). The influence of codon usage on translational output has also been demonstrated for adjacent codon pairs, also called bicodons (Gamble *et al.*, 2016; Diambra, 2017), indicating that the ability of a codon to be translated also depends on its sequence environment, termed codon context (Saunders and Deane, 2010; Komar, 2016). Depending on the codon change and context, sSNVs can alter translational rate and, thereby, the folding and stability of the nascent protein, ultimately leading to a reduced protein amount (Buhr *et al.*, 2016; Brule and Grayhack, 2017; Chaney *et al.*, 2017). Variants with such an impact on translation can conceivably underlie disorders driven by haploinsufficiency of specific dose-sensitive genes.

A complex interplay between different regulatory pathways orchestrates human brain development (Jiang and Nardelli, 2016). During embryonic development, the Sonic Hedgehog (SHH) pathway regulates forebrain dorso-ventral patterning in a morphogen-dependent manner (Fuccillo *et al.*, 2006), involving the SHH protein gradient. SHH is initially synthesized as a 45-kDa precursor, which then undergoes molecular processing in the endoplasmic reticulum (Bumcrot *et al.*, 1995). After removal of the signal peptide, SHH is auto-proteolytically cleaved into two secreted peptides: a 19-kDa N-terminal domain (SHH-N), possessing the signalling activity, and a 26-kDa C-terminal domain (SHH-C), responsible for the auto-processing and maturation of the protein. The precise spatiotemporal regulation of SHH activity is essential for normal development, and even slight variations of SHH gradient lead to brain malformations (Roessler *et al.*, 2009; Zhao *et al.*, 2012). Accordingly, pathogenic variants in *SHH* itself and its signalling effectors have been implicated in holoprosencephaly, a severe developmental disorder caused by SHH signalling

deficiency and characterized by incomplete separation of the two cerebral hemispheres (Mercier *et al.*, 2011). Although pathogenic variants have been reported in 17 genes, the genetic aetiology of holoprosencephaly remains elusive for up to 70% of patients (Dubourg *et al.*, 2018). Recently, we reported that hypomorphic variants in SHH-related genes are important risk factors and contribute to the holoprosencephaly phenotype through oligogenic inheritance (Kim *et al.*, 2019), which implies that a variant with a small effect on SHH activity can participate in the disease pathogenesis. Therefore, we sought to evaluate the contribution of *SHH* sSNVs to the pathogenesis of holoprosencephaly by investigating their effect on *SHH* mRNA and SHH protein.

## Materials and methods

### Patients, clinical information and sequencing data

The study protocol was approved by the Ethics Committee of Rennes Hospital. The full retrospective cohort was constituted of all the cases referred to our laboratory for molecular diagnosis during a 10-year period (2009–19) from eight French Centers Labeled for Developmental Anomalies (CLADs), centres of prenatal diagnosis and several European centres. Study participation involved informed written consent as well as the availability of clinical information and associated molecular data (whole exome, gene panel and Sanger sequencing with comparative genomic hybridization array/multiplex ligation-dependent probe amplification) for known/suspected genes associated with holoprosencephaly and related midline defects (Supplementary Table 6). For the 931 patients involved in the present study, we retrieved all the quality checked sSNVs located in the coding sequence of *SHH* (NM\_000193.3).

### Bioinformatic annotation of *SHH* sSNVs

Patient-associated SHH sSNVs were analysed through a semiautomated bioinformatics pipeline (genome version hg19) using Variant Effect Predictor (release 97), ANNOVAR, SpiCE (v2.1.3) and SilVA (v1.1.1) software, combined with manual annotations by Human Splicing Finder, TrAP and RegRNA 2.0 web servers. Alamut Visual v2.8.1 was used to retrieve splicing predictions of MaxEntScan and SSF-like models. gnomAD database (v2.1, *controls* dataset) was used to assess variant population frequency. Eighty-nine sSNVs of *SHH* (NM\_000193.3) found in the gnomAD *controls* dataset are provided in Supplementary Table 3.

### Secondary mRNA structure prediction

To assess the effect of each sSNV on the secondary structure of *SHH* mRNA, we performed simulations using the Mfold web server. RNA fragments of 75 and 401 nucleotides centred around each sSNV were queried to Mfold using default parameters. All predicted wild-type/mutated structures were considered

for each case. Mfold results of 401 nucleotide fragments were used to compare the average free energies ( $\Delta G$ ) predicted for the wild-type and the mutated sequences using Wilcoxon rank-sum test (statistical significance was set as  $P < 0.05$ ), while the 75 nucleotide fragments were used to compare the resulting optimal (most thermodynamically stable) structures. Other predicted structures are available in [Supplementary Fig. 4](#). Additionally, the ss-count values (propensities of a base to be single stranded) were compared between wild-type and mutated sequences ([Supplementary Fig. 2](#)).

## Minigene splicing assay

*In vitro* minigene assays were carried out as previously described ([Soukariéh et al., 2016](#)). Briefly, wild-type exon 2 of *SHH* and 150 bp of the 5' and 3' flanking introns were amplified from whole DNA blood extraction with Phusion® High-Fidelity DNA Polymerase (New England Biolabs). The insert was ligated into the pCAS-2 plasmid and the construct was sequenced by Sanger. Mutagenesis was performed for the *SHH* c.522C>T, c.552G>C and c.562+1G>A variants. The c.562+1G>A variant known to induce a strong defect of splicing was used as positive control. Wild-type or mutant vectors were transfected into HeLa cells in triplicate. Total RNA was harvested 24 h post-transfection and cDNA was transcribed using High Capacity cDNA Reverse Transcription kit with random primers (Applied Biosystems). PCR using primers specific to the 5' and 3' native exons of the pCAS-2 vector was performed and products were visualized on an agarose gel. Gel products were extracted and Sanger sequenced.

## Measures of codon and bicodon usage

The codon usage analysis was based on the relative synonymous codon usage (RSCU) index. The bicodon usage analysis was performed using the pause propensity indexes (PPI) and the relative synonymous bicodon usage (RSBU). Tricodon usage and global codon usage of *SHH* were estimated using %MinMax. Algorithms are briefly described below.

### Relative synonymous codon usage

RSCU index is a relative measure of the codon rareness and is defined as

$$\text{RSCU}_i = g_i / N_a \quad (1)$$

where  $f_i$  is the frequency of the codon  $i$ ,  $g$  is the degree of degeneracy of codon  $i$  (i.e. the number of codons which are synonymous to codon  $i$ ), and  $N_a$  is the frequency of the amino acid encoded by codon  $i$ .

### %MinMax

The %MinMax measure compares actual codon usage of a given sequence (*fact*) to hypothetical sequences encoding the same amino acids using either the most common (*fmax*) or most rare (*fmin*) synonymous codons, accounting for the average usage frequency (*favg*). The %Min and %Max are computed as follows:

$$\% \text{Min} = 100 \times (\text{fact} - \text{favg}) / (\text{fmax} - \text{favg}) \quad (2)$$

$$\% \text{Max} = 100 \times (\text{favg} - \text{fact}) / (\text{favg} - \text{fmin}) \quad (3)$$

For each residue, only the positive values are reported (either %Min or %Max, not both). The resulting values are averaged on a sliding window of size  $N$ . In this report, we used  $N = 9$  to analyse the global usage profile of the *SHH* sequence and  $N = 3$  to assess the effect of each sSNV by accounting for the codon context.

### Pause propensity index

This index is defined by the relationship between bicodon usage frequencies in two different samples of coding sequences: one sample is associated with highly abundant proteins (denoted by H), while the other sample is associated with lowly abundant proteins (denoted by L). Thus, the pause propensity of a bicodon<sub>*ij*</sub> (where  $i$  indicates the codon corresponding to P-site, while  $j$  indicates the one corresponding to the A-site) is defined as:

$$\pi_{ij} = q(\text{fijL} - \text{fijH}) / N_{ap} \quad (4)$$

where,  $\text{fij}^L$  and  $\text{fij}^H$  are the frequencies of the bicodon<sub>*ij*</sub> computed over the sequence samples  $L$  and  $H$ , respectively;  $q$  is the degree of degeneracy of the bicodon, and  $N_{ap}$  is the frequency of the corresponding amino acid pair computed over both samples or over the genomes.

### Relative synonymous bicodon usage index

Some bicodons are scarcely used in both samples. In these cases, the number of occurrences of a bicodon observed in one sample of sequences is not significantly different than the number of occurrences observed in the other sample of sequences, thus leading to  $\pi \sim 0$ . However, they are associated to a very low significance to estimate the pause propensity, as in the case of the variant c.897G>C. For that reason, we used a complementary measure for such cases, which is the RSBU index, computed over both samples of sequences. Mathematically, it is similar to the RSCU index and is defined as:

$$\text{RSBU}_{ij} = q(\text{fijL} + \text{fijH}) / N_{ap} \quad (5)$$

Similar to the RSCU index, RSBU is a measure of a relative rareness of a given bicodon.

Codon/bicodon variations were calculated for patient-associated sSNVs of *SHH* and compared to those of the sSNVs present in the gnomAD control database. All corresponding measures are available in [Supplementary Table 7](#). Importantly, each sSNV is associated with two bicodon variations, as one codon can be part of two different bicodons. For each sSNV, bicodons showing the largest change were taken into account.

## Generation of the mutant *SHH* cDNAs

Site directed mutagenesis was carried out on the *SHH* cDNA sequence integrated in the pRK5 expression vector to induce patient sSNVs using the Phusion® High-Fidelity DNA Polymerase (New England Biolabs) with primers designed according to QuikChange® Site-Directed Mutagenesis protocol (Agilent). The resulting *SHH* wild-type, c.522C>T, c.552G>C, c.570G>A, c.630C>T, c.885C>T, c.897G>C, c.1206C>T and

c.1354C>T constructs were confirmed by Sanger sequencing using the BigDye<sup>®</sup> Terminator v3.1 kit (Applied Biosystems).

## XTAPA plasmid constructs

The XTAPA plasmid constructs are based on a modification of the episomal pRTS-1 vector (Supplementary Fig. 5), which comprises a bidirectional promoter  $P_{tet}$ -bi-1, driving expression of two genes in a coordinated fashion, encoding mCherry and mEGFP, flanked by unique Xho and SmaI sites, respectively. These unique restriction sites allow insertion of cDNAs of interest in phase with the mCherry and mEGFP sequences. Five glycines were added as flexible linkers. The vector backbone also comprises the ampicillin and hygromycin B resistance gene. For each XTAPA construct, two versions of the *SHH* gene were inserted: the wild-type *SHH* sequence tagged by mCherry and either the wild-type *SHH* sequence or the mutant *SHH* sequence containing one of the sSNVs tagged by mEGFP. *SHH* cDNAs were generated by PCR and amplified with pRK5 specific primers (Table 1). PCR products were separated from matrix plasmids on a 0.8% agarose gel and purified (NucleoSpin<sup>®</sup> Gel and PCR Clean-up, Macherey-Nagel). *SHH* cDNAs were inserted using the Gibson Assembly method. The constructs were transformed into DH5 $\alpha$  competent cells, amplified, and confirmed using restriction enzyme digestion (BglIII) as well as conventional Sanger sequencing.

## Cell culture

HeLa cells were cultured in Dulbecco's Modified Eagle Medium 1X (DMEM, Gibco, ThermoFisher Scientific), with addition of 10% v/v of foetal bovine serum (FBS), 100 U/ml of penicillin and 100  $\mu$ g/ml of streptomycin. All cell lines were maintained at 37°C with 5% CO<sub>2</sub>. At 60% confluency, the cells were transfected with XTAPA constructs using jetPRIME<sup>®</sup> transfection reagent (Polyplus transfection) and selected by adding 100  $\mu$ g/ml of hygromycin B (InvivoGen). After selection, stable HeLa transduced cell lines were maintained in a complete medium with 25  $\mu$ g/ml hygromycin B. For experiments, cells were grown onto glass coverslips and treated with 2  $\mu$ g/ml doxycycline for 24 h to induce expression of *SHH*. For proteasome inhibition assays, MG132 was added to medium at 2  $\mu$ mol/l 7 h after doxycycline addition.

## Fluorescence microscopy and analysis

Medium was removed, and cells were fixed in 4% formaldehyde in phosphate-buffered saline for 20 min. Cells were permeabilized with 0.2% Triton<sup>™</sup> X-100 for 10 min, DNA was counterstained with DAPI (Invitrogen, ThermoFisher). Coverslips were mounted to microscope slides with ProLong<sup>®</sup> Gold (Life Technologies). Images were finally acquired using an Olympus IX71 inverted fluorescence microscope (DeltaVision, MRic platform). Quantification of the fluorescence intensities mEGFP (green,  $I_{meGFP}$ ) and mCherry (red;  $I_{mCherry}$ ) for each cell was done with ImageJ software. After background intensity subtraction ( $I_{meGFP}^{BGD}$  and  $I_{mCherry}^{BGD}$ ), signal intensities were expressed as an  $I_{meGFP}/I_{mCherry}$  ratio:  $R = (I_{meGFP} - I_{meGFP}^{BGD}) / (I_{mCherry} - I_{mCherry}^{BGD})$ . The resulting measurements obtained by condition (cell line) were expressed as mean and dispersion

value, normalized to the control condition (wild-type/wild-type) for each experiment.

## Quantitative real-time polymerase chain reaction

Total RNA was isolated from stable HeLa cell lines 24 h after doxycycline induction using NucleoSpin<sup>®</sup> RNA kit (Macherey-Nagel). To generate cDNA, 2  $\mu$ g total RNA per sample was reversed transcribed using the High Capacity cDNA Reverse Transcription kit. Quantitative real-time PCR analyses were performed using SYBR<sup>®</sup> Green in 7900HT Fast Real-Time PCR system (Applied Biosystems) with primer sequences as shown in Table 1. The PCR reactions were carried out under the following conditions: 95°C for 10 min followed by 40 cycles of 95°C for 15 s, 60°C for 1 min, then incubation at 95°C for 15 s, 60°C for 15 s and finally 95°C for 15 s for the dissociation stage. The  $\Delta\Delta C_t$  method was used to calculate the fold-change in gene expression using the housekeeping genes *36B4* and *SNAPIN* and control sample for normalization. Specificity of amplification and absence of primer dimers were confirmed by melting curve analysis at the end of each run. The quality and efficiency of the primers were assessed prior to their usage. The SDS 2.4 software (Applied Biosystems) was used for the quantification analysis.

## Statistical analysis

Statistical analyses were performed using R (v3.5.1). All the statistical tests used are described in the relevant sections of the article.  $P$ -values < 0.05 were considered statistically significant. GraphPad prism 8 was used for graphical presentations. Confidence intervals: \* $P$  < 0.05; \*\* $P$  < 0.01; \*\*\* $P$  < 0.001.

## Data availability

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

**Table 1** List of primers used in real-time qPCR

Gene name		Primer sequence (5' → 3')
<i>GFP</i>	Forward	AAGCTGACCCTGAAGTTCATCTGC
	Reverse	CTTGATGTTGCCGTGCTCCTTGAA
<i>mCherry</i>	Forward	GAACGGCCACGAGTTCGAGA
	Reverse	CTTGAGCCGTACATGAACTGAG
<i>36B4</i>	Forward	CAGCAAGTGGGAAGGTGTAATCC
	Reverse	CCCATTCTATCATCAACGGGTACAA
<i>SNAPIN</i>	Forward	GGGAGCCGTACAGTTTGTT
	Reverse	GTCCTTTGGGAACCTCATTCT
<i>SHH mEGFP</i>	Forward	GAATTCCTGCAGCATTTAAAT ATGCTGCTGCTGGCGAGATGT
	Reverse	CTCCTCCCCCATTAAAT GGCTGGACTTGACCGCCATGC
<i>SHH mCherry</i>	Forward	GTCAGCAAGCTTTTCGGCTCACCT CGAGATGCTGCTGCTGGCGAGATGT
	Reverse	CACTCCCCCTCCTCCCCCTCGAGGCT

## Results

### *SHH* sSNVs are more prevalent in holoprosencephaly patients compared to controls

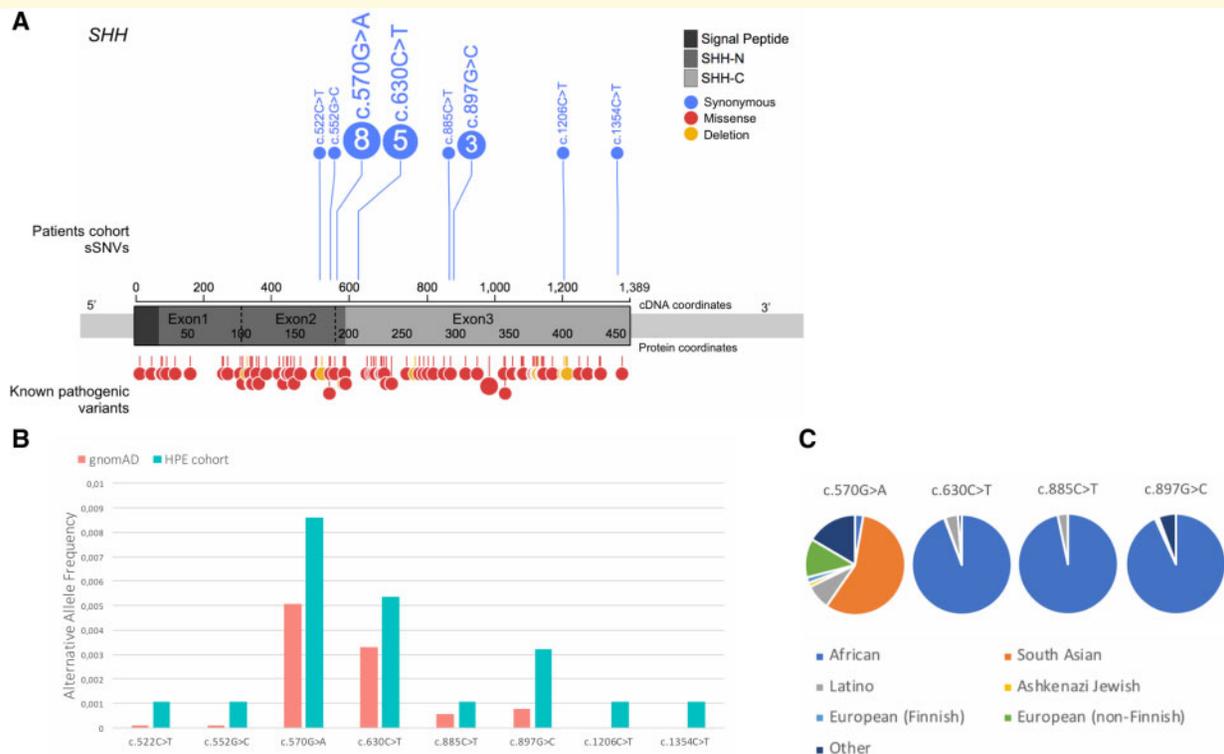
We performed a retrospective analysis of clinical sequencing data obtained from 931 patients affected by holoprosencephaly and associated disorders who were referred to our laboratory for molecular screening of *SHH* and related genes. Variant analysis identified eight different sSNVs of *SHH* (Fig. 1A) in 21 patients (Supplementary Table 1). Among the identified sSNVs, two variants (c.1206C>T and c.1354C>T) were never reported among 60 146 healthy individuals present in Genome Aggregation Database (gnomAD) (Karczewski *et al.*, 2019). Two variants (c.522C>T and c.552G>C) were detected only once in gnomAD. Three variants (c.630C>T, c.885C>T and c.897G>C) were rare in gnomAD [minor allele frequency (MAF) < 0.5%] and were more prevalent in individuals of African ancestry (Fig. 1B, C and Supplementary Table 2), consistent with studies suggesting a higher prevalence of holoprosencephaly in this ethnicity

(Orioli and Castilla, 2010; Vaz *et al.*, 2012). Finally, c.570G>A (MAF = 0.5%) was the most frequent sSNV of *SHH* in both gnomAD and patient cohort, questioning its polymorphic nature (Supplementary Table 3). This exception aside, all patient-associated sSNVs of *SHH* were absent in two additional control cohorts: Genome of the Netherlands (Genome of the Netherlands Consortium, 2014) (GoNL, 498 individuals) and French Exome Consortium (FrEX, 574 individuals) (Supplementary Table 4).

Altogether, these results indicate that *SHH* sSNVs are more prevalent in holoprosencephaly patients as compared to healthy individuals, suggesting their contribution to the disease mechanism.

### Identified sSNVs do not affect splicing, mRNA folding or miRNA regulation of *SHH*

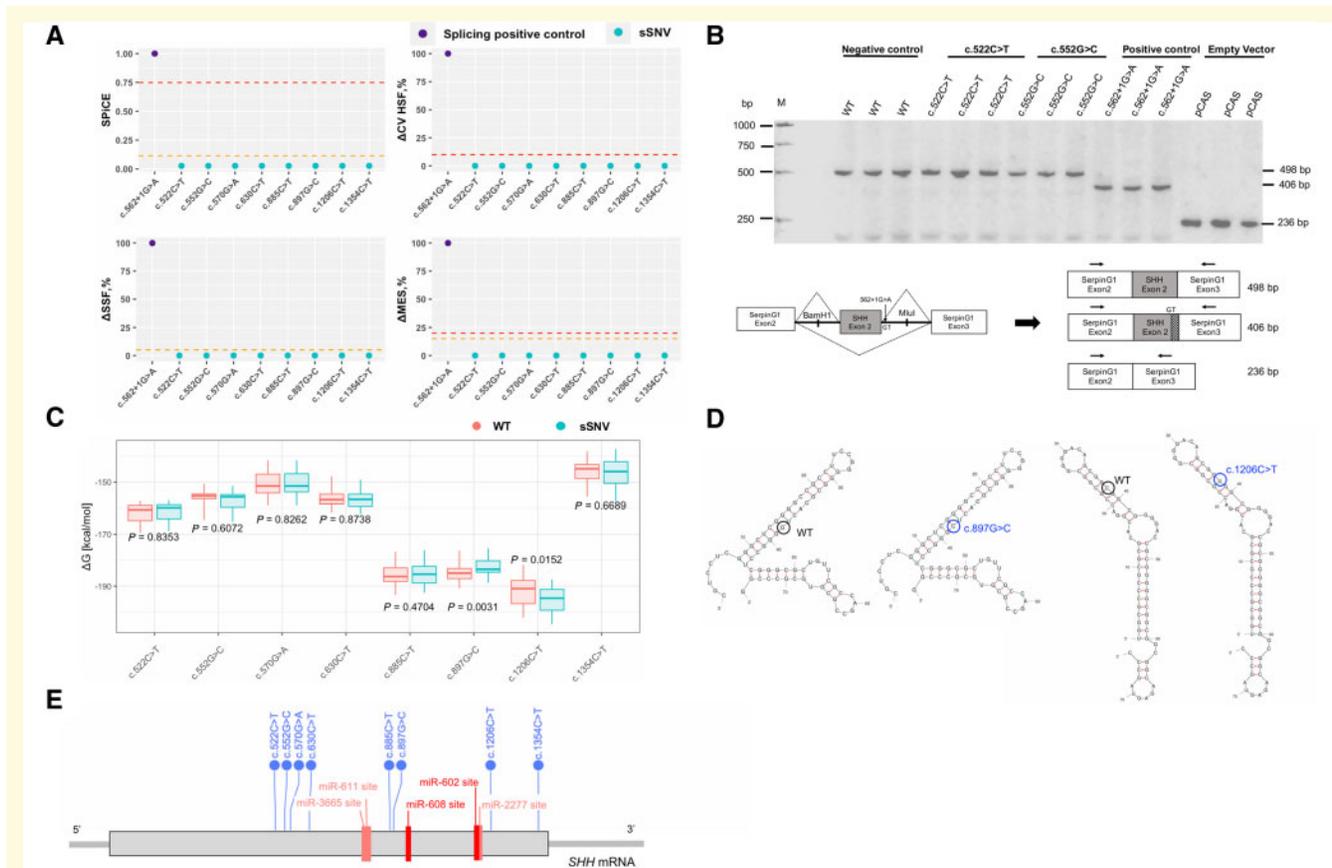
We next investigated the deleterious impact the identified *SHH* sSNVs may have on mRNA by using a series of *in silico* and *in vitro* assays.



**Figure 1 Identification of *SHH* sSNVs in patients with holoprosencephaly.** (A) Schematic representation of the *SHH* gene with the localization of patient-associated sSNVs represented by blue circles. Numbers inside the circles indicate the number of patients harbouring each variant (absent numbers indicate that only one variant was identified). Known pathogenic variants reported in manually-curated UniProt database are shown below (non-synonymous and stop variants in red and deletions in yellow). UTR and exonic regions are represented by, grey lines and squares, respectively. Regions coding for the signal peptide, SHH-N, and SHH-C domains are indicated. (B) Comparison of sSNV frequencies between gnomAD and patients cohort. (C) Ethnicity distribution of the four rare *SHH* sSNVs present in the gnomAD database. c.630C>T, c.885C>T and c.897G>C were more frequent in individuals of African ancestry, while c.570G>A was more prevalent in South Asian ethnicity.

First, we investigated the effect of each sSNV on splicing in comparison to the known pathogenic variant c.562+1G>A (Roessler *et al.*, 2009), which is located at a canonical splice site of *SHH* and used here as a positive control. No alteration of canonical splice sites nor splicing regulatory features was predicted *in silico* for any of the identified sSNVs by SpiCe (Leman *et al.*, 2018), Human Splicing Finder (Desmet *et al.*, 2009), SSF-like (Shapiro and Senapathy, 1987) and MaxEntScan (Yeo and Burge, 2004) models (Fig. 2A and Supplementary Fig. 1). We further evaluated the impact on splicing of the two rare sSNVs of

uncertain significance that were located near exon-intron boundaries—c.522C>T and c.552G>C—by using dedicated *in vitro* assay based on a minigene reporter. Four constructs were produced and analysed with sequences containing either wild-type *SHH*, the above-mentioned positive control, or one of the two sSNVs (Fig. 2B). For the wild-type construct, a splicing product of expected size (498 bp) was observed. The positive control resulted in a smaller product (406 bp), indicative of a partial exon skipping. Analysis of the minigenes containing each of the two tested sSNVs did not reveal any differences when compared to the



**Figure 2** The identified sSNVs are not associated with defects of splicing, mRNA folding or miRNA-based regulation. **(A)** Computational predictions for *SHH* sSNVs by SpiCE, Human Splicing Finder ( $\Delta$ HSF), SSF-like ( $\Delta$ SSF) and MaxEntScan ( $\Delta$ MES) models. Pathogenicity thresholds are represented by yellow (potentially deleterious) and red (probably deleterious) dashed lines. sSNVs prediction values are represented by cyan circles and compared to the positive control c.562+1G>A (purple). **(B)** Minigene splicing assay for c.522C>T and c.552G>C sSNVs. The scheme of the *SHH* exon 2 minigene is shown with the exons of splicing reporter (Serp1nG1), enzyme restriction sites (BamHI, MluI) and the localization of the positive control c.562+1G>A. Bands at sizes 498 bp and 236 bp correspond to a normal splicing product and empty vector, respectively. In the positive control, the c.562+1G>A variant (406-bp band) impairs the splicing by altering the canonical donor site (GT) and activating a cryptic exonic site (GT'), located 92 bp upstream, which results in partial exon 2 skipping of *SHH*. The full-length unmodified gel is available in the [Supplementary material](#). **(C)** Effects of identified sSNVs on mRNA folding were assessed by comparing the average free energy ( $\Delta$ G) between secondary structures predicted by Mfold for wild-type (salmon) and mutated (cyan) sequences. Four hundred and one nucleotide fragments centred around each sSNV were used. Box plots represent the distribution of  $\Delta$ Gs calculated by Mfold for each condition. Statistical comparison was performed using Wilcoxon rank-sum test and corresponding *P*-values are reported. Statistical significance was set as *P* < 0.05. **(D)** Schematics of the most thermodynamically stable secondary structures predicted by (75 nucleotide fragments) for c.897G>C and c.1206C>T variants. **(E)** Schematic presentation of *SHH* mRNA and known/predicted miRNA binding sites. Two target sites in *SHH* were reported by Akhtar *et al.* (2015) for miR-602 and miR-608, and three sites for miR-3665, miR-611 and miR-2277 were predicted by the RegRNA software. All identified sSNVs (blue) are located outside of known (red) and predicted (light red) miRNA binding sites. Additionally, the sSNVs did not result in creation of new miRNA response elements (MREs, data not shown).

wild-type, thus confirming the predicted absence of splicing-related defects.

Next, we examined whether the sSNVs could alter the mRNA folding of *SHH* by comparing the secondary structures predicted by Mfold (Zuker, 2003) for the wild-type and for the different sSNVs alleles. Statistically significant differences in average predicted free energies ( $\Delta G$ ) were observed only for c.897G>C and c.1206C>T variants when compared to the wild-type (Fig. 2C); however, further inspection of optimal mRNA structures (Fig. 2D and Supplementary Fig. 4) and base-pairing probabilities (Supplementary Fig. 2) did not predict any changes. All sSNVs were classified as benign by SilVA (Buske *et al.*, 2013) and TrAP (Gelfman *et al.*, 2017) models (Supplementary Fig. 3 and Supplementary Table 5) further arguing against any effect of sSNVs on mRNA folding.

Finally, none of the sSNVs altered any of the reported (Akhtar *et al.*, 2015) or predicted miRNA response elements in *SHH*, thus ruling out their possible effect on miRNA-based regulation (Fig. 2E).

Taken together, these results indicate that the identified sSNVs of *SHH* do not cause any molecular alterations at the RNA level. Thus, we interrogated their impact at the protein level by investigating whether the identified sSNVs could modify the translation efficiency of SHH through their impact on codon usage.

## Codon usage analysis reveals the translational dynamics of *SHH*

We first calculated the global codon usage profile of *SHH* using the %MinMax algorithm (Rodriguez *et al.*, 2018), which evaluates frequency patterns of synonymous codon usage from a given DNA coding sequence. %MinMax frequency scores were computed for the whole coding sequence of *SHH* and averaged over a sliding window of nine codons. This analysis revealed four clusters of rare codons (%MinMax < 1), three of which being in proximity of the identified sSNVs (Fig. 3A). Interestingly, one cluster overlaps with the boundary between the two main functional domains of SHH (SHH-N and SHH-C) (Choudhry *et al.*, 2014), which is consistent with previous studies suggesting that rare codon clusters occupy strategic positions slowing down the translation at interfaces between important protein domains (Thanaraj and Argos, 1996), albeit with conflicting evidence (Chaney *et al.*, 2017). Globally, this analysis indicates that *SHH* mRNA consists of fast and slow translating regions and that the introduction of certain patient-associated sSNVs could affect its translational efficiency through alterations of codon usage.

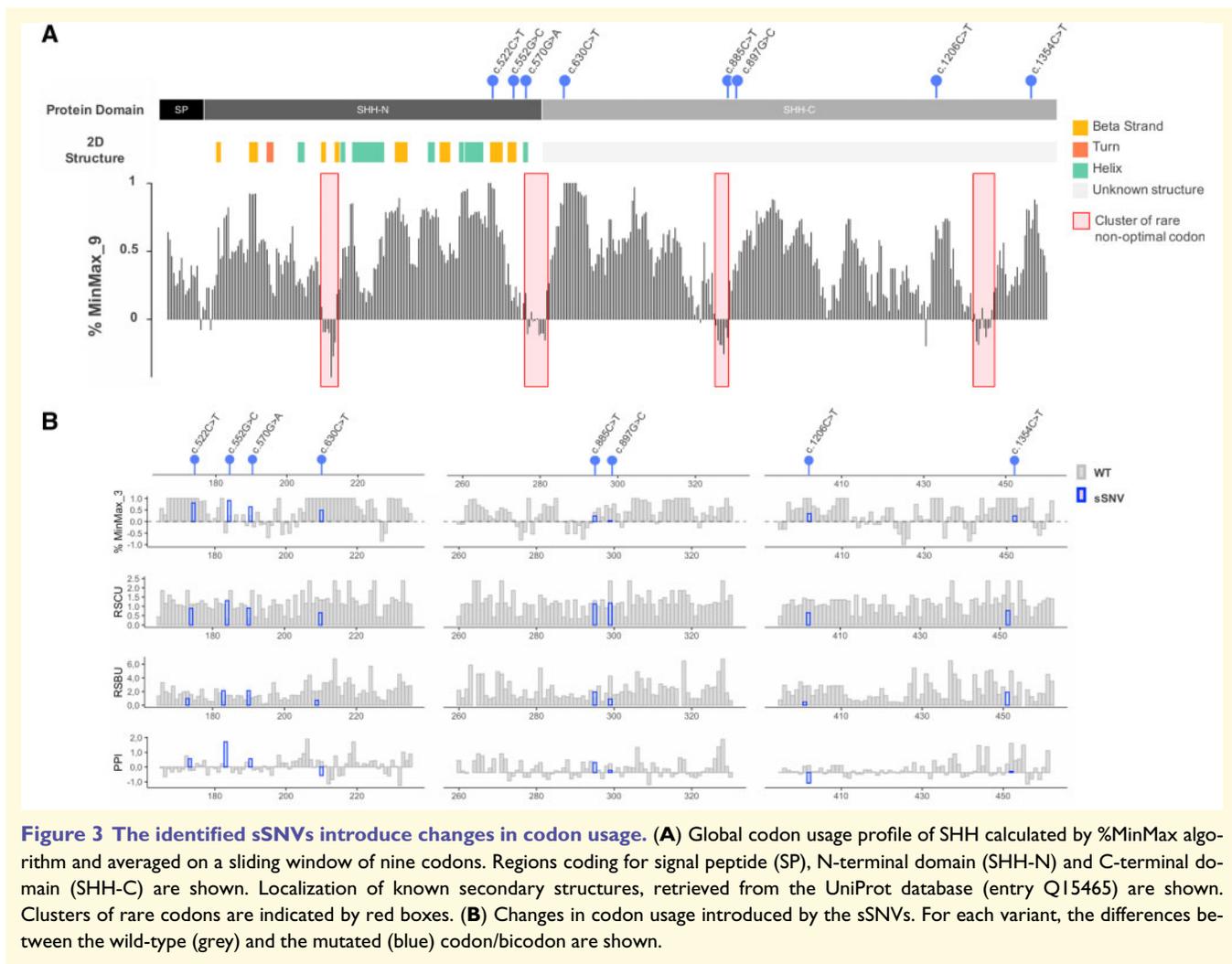
## Five of eight identified sSNVs induce substantial changes in codon usage

We next evaluated the impact of each identified sSNV on codon usage by using a series of established algorithms that

measure the frequency of each codon. First, RSCU (Sharp *et al.*, 1986) was used to measure the variations at the single codon level. To account for the codon context around each sSNV, we extended this analysis to bicodons (RSBU; Diambra, 2017) and the Pause Propensity Index (McCarthy *et al.*, 2017) measures and tricodons (%MinMax scores averaged over a sliding window of three codons) (Fig. 3B, see the 'Materials and methods' section for details). For each algorithm and each sSNV, the changes in codon usage frequency between the wild-type and mutated alleles were defined as  $\Delta$ -values ( $\Delta = |\text{mutated} - \text{wild-type}|$ ). To assess the significance of each  $\Delta$ -value, we performed the same calculations for 89 *SHH* sSNVs reported in the general population (gnomAD), which were used as control dataset. Five of eight patient-associated sSNVs (c.552G>C, c.630C>T, c.897G>C, c.1206C>T and c.1354C>T) presented at least one significant  $\Delta$ -value (above the 3rd quartile of the control dataset) and were predicted to impact the translation (Table 2). The c.1354C>T variant introduced the largest decrease in codon usage (i.e. replacement of a very frequent codon by a very rare one), indicating a slowdown in an otherwise fast translating region (Fig. 3A and B). The largest increase was observed for c.552G>C, indicating an accelerated translation rate in a slow translating region. c.522C>T and c.885C>T were associated with the smallest changes among the studied variants and showed conflicting results between different  $\Delta$  measures, and were therefore considered of low to null impact. Overall, these results indicate that substantial changes in codon usage were observed for c.552G>C, c.630C>T, c.897G>C, c.1206C>T and c.1354C>T variants. Consequently, these five variants are predicted to impact the translational program of SHH, ultimately leading to reduced protein amount.

## The identified sSNVs reduce the protein amount of SHH and confirm *in silico* predictions of codon usage

To test these predictions, we assessed the impact of each sSNV on SHH protein translation using a plasmid-based assay XTAPA, which allows the simultaneous expression of both the wild-type *SHH* tagged by mCherry and of different mEGFP-tagged versions of *SHH* harbouring sSNVs of interest (Supplementary Fig. 5). The use of a bidirectional promoter results in highly correlated expressions of the mEGFP and mCherry-tagged proteins at the single cell level ( $R^2 = 0.82$ ;  $P < 2.2 \times 10^{-16}$ ; Supplementary Fig. 6), thereby allowing for the precise comparison of produced protein amounts. Results are shown in Fig. 4A and B. Five sSNVs, which were previously predicted to impact translation, resulted in a statistically significant reduction of protein amount. The strongest effect was observed for variants c.1206C>T and c.1354C>T, associated with, respectively, 23% and 19% reduction of SHH amount. c.552G>C and c.897G>C variants resulted in a 17% and 13% decrease, respectively.



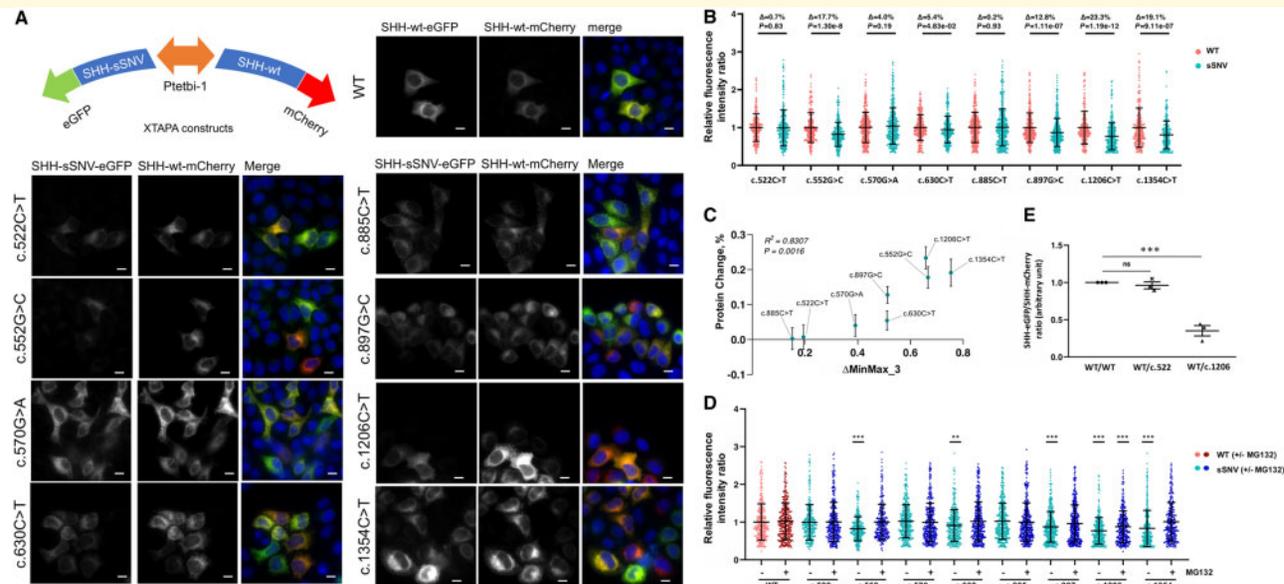
**Table 2** Summary of changes in codon/bicodon usage introduced by the sSNVs

Variant	Codon		$\Delta$ RSCU	$\Delta$ MinMax_3	$\Delta$ RSBU	$\Delta$ PPI
	From	to				
c.522C>T	TAC	TAT	~	~	~	~
<b>c.552G&gt;C</b>	<b>TCG</b>	<b>TCC</b>	↑	~	~	↑
c.570G>A	TCG	TCA	~	~	~	~
<b>c.630C&gt;T</b>	<b>GGC</b>	<b>GGT</b>	~	~	~	↓
c.885C>T	TCC	TCT	~	~	~	~
<b>c.897G&gt;C</b>	<b>CTG</b>	<b>CTC</b>	↓	~	↓	~
<b>c.1206C&gt;T</b>	<b>GGC</b>	<b>GGT</b>	~	~	~	↓
<b>c.1354C&gt;T</b>	<b>CTG</b>	<b>TTG</b>	↓	↓	↓	↓

For each variant, we calculated differences in usage between the wild-type and the mutated codon/bicodon using RSCU, MinMax, RSBU and PPI algorithms. For this analysis, %MinMax values were averaged over a sliding window of three codons to assess with better resolution the effect of each sSNV. To assess the significance of changes, the resulting  $\Delta$ -values were compared to those of known SHH sSNVs reported in gnomAD control database. Tilde (~) and downward and upward arrows (↓) indicate low (below 3rd quartile) and high (above 3rd quartile)  $\Delta$ -values, respectively. For RSBU and PPI, each sSNV is associated with two bicodon variations, since one codon can be part of two different bicodons. For each sSNV, bicodons showing the largest change were taken into account. Bold text indicates sSNVs predicted to alter the translational efficiency.

The c.630C>T variant was associated with a small albeit significant decrease of ~5%. No significant alteration of protein expression was observed for c.522C>T, c.570G>A and c.885C>T variants, consistent with codon usage

predictions. Remarkably, a significant correlation ( $R^2 = 0.83$ ;  $P = 0.0016$ ) between experimental values of protein reduction and computational measures of  $\Delta$ MinMax was observed (Fig. 4C), providing strong arguments for the



**Figure 4** The identified sSNVs affect the translation of the SHH protein. (A) Stable XTAPA cell lines were established for all eight sSNVs together with a cell line expressing mCherry- and mEGFP-tagged wild-type versions of SHH that is used as control condition (see 'Materials and methods' section). XTAPA construct scheme is shown. The images are (from left to right) views in mEGFP, mCherry, and merged. Merged analysis displaying overlapping mEGFP and mCherry signals. Scale bars = 10  $\mu$ m. (B) Upon ectopic protein expression, background-corrected fluorescence intensities ( $I_{mEGFP}$  and  $I_{mCherry}$ ) were measured by fluorescence microscopy and expressed as  $I_{mEGFP}/I_{mCherry}$  ratios for each sSNV. These ratios were further normalized using the wild-type/wild-type version condition. For each sSNV, dots represent measures of SHH protein amount produced by cells containing wild-type (salmon) or mutated cDNA (cyan). Black lines represent means with standard deviation. (C) Correlation between experimental measures of protein change and computational measures of  $\Delta$ MinMax. Each point represents one sSNV with the corresponding standard deviation. Linear regression ( $R^2$ ) coefficient is presented with the corresponding  $P$ -value. (D) Proteasome inhibition assay rescues four of five sSNVs associated with reduced protein amount. For each condition, dots represent measures of SHH protein amount produced by cells containing wild-type (salmon and red) or mutated cDNA (cyan and blue) from three independent experiments. Normal (–) and MG132-treated (+) conditions are indicated. Black lines represent means with standard deviation. Asterisks indicate a significant difference compared to the wild-type (WT) normal condition.  $**P < 0.01$ ;  $***P < 0.001$ , Wilcoxon unpaired two-tailed test. (E) Histogram representing qPCR results of the WT/WT, WT/c.522C > T and WT/c.1206C > T. Black lines represent mean fold change ( $\pm$ SEM,  $n = 3$ ).  $***P < 0.001$ .

relevance of this model in predicting the impact of sSNVs on protein translation.

These results demonstrate the actual impact of five sSNVs on protein levels of SHH and provide evidence as to how changes in codon usage can alter protein translation.

### Inhibition of proteasomal degradation rescues the effects of four of five sSNVs

From a molecular point of view, the observed reduction of SHH amount can be accounted for by either reduced synthesis (Ahat *et al.*, 2019), increased proteasome-mediated degradation (Mohanraj *et al.*, 2019) or both. To test degradation mediated by the proteasome, we measured in the same manner the protein levels for each condition in the presence of the proteasome inhibitor MG132. While the expression of the wild-type SHH remained unaffected, proteasome inhibition completely rescued the levels of SHH amount for c.552G > C, c.630C > T, c.897G > C and

c.1354C > T variants (Fig. 4D), indicating that these mutated versions of SHH are selectively degraded by the proteasome. By contrast, only a partial restoration was observed for c.1206C > T, supporting that proteasome degradation is only in part responsible for the observed reduced protein amount. To investigate the molecular mechanism underlying this variant further, we measured its impact on mRNA levels using a qPCR assay. This analysis revealed a significantly reduced mRNA levels for the c.1206C > T variant compared to wild-type SHH (Fig. 4E), suggesting a combined effect of this variant at both transcriptional and translational levels.

Altogether, these results highlight that the observed protein reduction of SHH is mostly accounted for by proteasome-mediated degradation, leading to a production of an unstable SHH protein.

## Discussion

Our study demonstrates the impact of SHH sSNVs on codon usage, leading to significant reduction of the resulting

protein, for up to 23% for the c.1206C>T variant. Considering the critical role of SHH gradient in early brain development (Fuccillo *et al.*, 2006; Roessler *et al.*, 2009; Choudhry *et al.*, 2014), our findings indicate that such sSNVs can participate in the pathogenesis of holoprosencephaly and of related disorders.

In almost all instances, the observed protein reduction of SHH is accounted for by proteasome-mediated degradation. Considering that the proteasome is involved in the degradation of misfolded proteins (Mohanraj *et al.*, 2019), our results suggest protein misfolding as the main pathogenic impact of the identified sSNVs. To our knowledge, these are the first experimental and computational evidences that simultaneously link sSNVs to alterations of codon usage, protein misfolding and degradation in a clinical context. Future studies will help to elucidate whether this translation-related misfolding can impair the function of SHH protein and the overall morphogenic activity of the SHH gradient. For example, previous explorations of the multidrug resistance 1 gene (*ABCB1*) showed that a common sSNV altered the conformation and substrate specificity of the resulting protein, albeit without causing detectable misfolding or degradation (Kimchi-Sarfaty *et al.*, 2007; Tsai *et al.*, 2008). Bartoszewski *et al.* (2010) showed that an sSNV in the *CFTR* gene reduces the translational activity by altering mRNA folding. In the case of *SHH*, it would be interesting to test whether the identified sSNVs impair the molecular processing and the downstream signalling activity of the SHH-N peptide (Bumcrot *et al.*, 1995).

We demonstrate that one sSNV—c.1206C>T—is associated with a reduction of both the mRNA and the protein levels. Decreased mRNA levels could originate from a reduced transcriptional activity (Helmlinger *et al.*, 2006), although the precise mechanism by which this sSNV could impact on transcription remains unknown. Alternatively, this reduction in mRNA level could be a consequence of a negative feedback loop from suboptimal translation leading to mRNA degradation, as described recently (Presnyak *et al.*, 2015). In both cases, further investigations will be required to elucidate the impact of this particular sSNV.

We also show that sSNVs influencing the SHH protein amount are present in the general population. The c.630C>T and c.897G>C variants (associated with 5% and 13% protein reduction, respectively) were found in 389 individuals of the gnomAD database, including seven homozygous carriers for the c.630C>T variant observed in the African ethnicity (Supplementary Tables 2 and 3). Considering that human brain development is highly influenced by genetic factors (Tarailo-Graovac *et al.*, 2017), these sSNVs could contribute to phenotypic spectrum of brain and midline structures in healthy individuals.

In holoprosencephaly cases, the presence of additional pathogenic variants in the same patients (Supplementary Table 1) raises the possibility that these sSNVs participate in the pathogenesis but are not sufficient by themselves to cause the disease. In that regard, one can envision two disease mechanisms, considering the impact of *SHH* sSNVs on

the expression of the resulting protein. First, certain *SHH* sSNVs could act as modifiers and modulate the disease phenotype induced by another pathogenic mutation. Second, sSNVs could participate in oligogenic inheritance, where a combination of several pathogenic variants, including a *SHH* sSNV, is necessary for the disease to occur, as we previously reported (Kim *et al.*, 2019). In most cases, if not all, the presence of additional genetic factors besides *SHH* sSNVs is certainly necessary for the manifestation of the disease. We, therefore, conclude that *SHH* sSNVs are important risk factors for holoprosencephaly that need to be investigated further. Additional molecular screenings in larger populations of patients with holoprosencephaly combined with dedicated functional studies will help in refining the implication of *SHH* sSNVs in pathogenesis of holoprosencephaly and related disorders.

Importantly for future studies, we show that deleterious impacts of the sSNVs on protein expression can be accurately predicted using *in silico* estimations of codon usage. In particular, the measure based on a three-codon window analysis ( $\Delta$ MinMax) has a strikingly high correlation with the experimental measures of protein reduction, underlining the importance of considering the codon context when predicting the impact of sSNVs on translation. Although successfully applied to *SHH* and holoprosencephaly in this study, additional investigations of sSNVs in other genes and pathologies are needed to evaluate the predictive power of this measure and its potential use in diagnosis of genetic disorders. Overall, we underline the relevance of our *in silico* estimations in predicting the effects of sSNVs on translational efficiency and thus recommend its use when screening known disease genes.

In conclusion, our study underlines the clinical relevance of *SHH* sSNVs in holoprosencephaly and highlights the necessity to more systematically investigate the impact of sSNVs on codon usage in genetic diseases.

## Web resources

RegRNA, <http://regrna2.mbc.nctu.edu.tw/>

HSF, <http://www.umd.be/HSF3/HSF.shtml>

TrAP, <http://trap-score.org/>

UniProt, <https://www.uniprot.org/uniprot/Q15465>

FREX, <http://lysine.univ-brest.fr/FrExAC/>

## Acknowledgements

The authors acknowledge the Centre de Ressources Biologiques (CRB)-Santé (<http://www.crbsante-rennes.com>) of Rennes for managing patient samples. We thank Marie-Dominique Galibert for carefully reading this manuscript and for scientific discussion. We thank Alinoe Lavillaureix for retrieving the clinical data. We thank Charlotte Andrieu for helping with figure design. We thank Alexandra Martins for her assistance with the minigene. We thank Luc Paillard

and Microscopy Rennes Imaging Center (MRIC) for their assistance with this study. We would like to thank the families for their participation in the study, all clinicians who referred patients with midline cerebral anomalies, the eight CLAD (Centres Labellisés pour les Anomalies du Développement) within France that belong to FECLAD, French centres of pre-natal diagnosis (CPDPN) and the SOFFOET for foetal cases, and the 'filière AnDDI-Rares'. We particularly thank all members of the Molecular Genetics Laboratory (CHU, Rennes) and of the research team 'Genetics of Development-related pathologies' (UMR6290 CNRS, University Rennes 1) for their help and advice.

## Funding

This work was supported by Fondation Maladie Rares (grant PMO1201204), Agence Nationale de la Recherche (grant ANR-12-BSV1-0007-01), the Agence de la Biomédecine (AMP2016) and the French National Cancer Institute (PRTK 2016). A.K. is a PhD student funded by Fondation Recherche Médicale (FRM, grant FDT201904008099).

## Competing interests

The authors report no competing interests.

## Supplementary material

Supplementary material is available at *Brain* online.

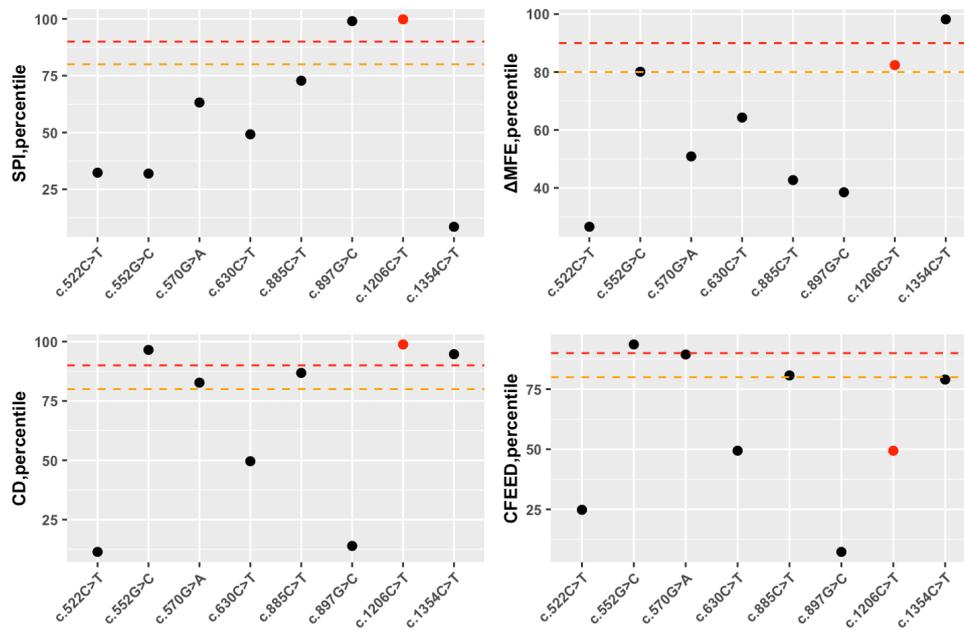
## References

- Ahat E, Xiang Y, Zhang X, Bekier ME, Wang Y. GRASP depletion-mediated Golgi destruction decreases cell adhesion and migration via the reduction of  $\alpha 5 \beta 1$  integrin. *Mol Biol Cell* 2019; 30: 766–77.
- Akhtar N, Makki MS, Haqqi TM. MicroRNA-602 and MicroRNA-608 regulate Sonic Hedgehog expression via target sites in the coding region in human chondrocytes. *Arthritis Rheumatol* 2015; 67: 423–34.
- Bartoszewski RA, Jablonsky M, Bartoszezowska S, Stevenson L, Dai Q, Kappes J, et al. A synonymous single nucleotide polymorphism in  $\Delta F508$  *CFTR* alters the secondary structure of the mRNA and the expression of the mutant protein. *J Biol Chem* 2010; 285: 28741–8.
- Berg OG, Kurland CG. Growth rate-optimised tRNA abundance and codon usage 1 Edited by J. H. Miller. *J Mol Biol* 1997; 270: 544–50.
- Brest P, Lapaquette P, Souidi M, Lebrigand K, Cesaro A, Vouret-Craviari V, et al. A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nat Genet* 2011; 43: 242–5.
- Brule CE, Grayhack EJ. Synonymous codons: choose wisely for expression. *Trends Genet* 2017; 33: 283–97.
- Buhr F, Jha S, Thommen M, Mittelstaet J, Kutz F, Schwalbe H, et al. Synonymous codons direct cotranslational folding toward different protein conformations. *Mol Cell* 2016; 61: 341–51.
- Bumcrot DA, Takada R, McMahon AP. Proteolytic processing yields two secreted forms of sonic hedgehog. *Mol Cell Biol* 1995; 15: 2294–303.
- Buske OJ, Manickaraj A, Mital S, Ray PN, Brudno M. Identification of deleterious synonymous variants in human genomes. *Bioinformatics* 2013; 29: 1843–50.
- Chaney JL, Steele A, Carmichael R, Rodriguez A, Specht AT, Ngo K, et al. Widespread position-specific conservation of synonymous rare codons within coding sequences. *PLoS Comput Biol* 2017; 13: e1005531.
- Choudhry Z, Rikani AA, Choudhry AM, Tariq S, Zakaria F, Asghar MW, et al. Sonic hedgehog signalling pathway: a complex network. *Ann Neurosci* 2014; 21: 28–31.
- Dana A, Tuller T. The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Res* 2014; 42: 9171–81.
- Desmet F-O, Hamroun D, Lalande M, Collod-Bérout G, Claustres M, Bérout C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009; 37: e67.
- Diambra LA. Differential bicodon usage in lowly and highly abundant proteins. *PeerJ* 2017; 5: e3081.
- Dubourg C, Kim A, Watrin E, de Tayrac M, Odent S, David V, et al. Recent advances in understanding inheritance of holoprosencephaly. *Am J Med Genet C Semin Med Genet* 2018; 178: 258–69.
- Fuccillo M, Joyner AL, Fishell G. Morphogen to mitogen: the multiple roles of hedgehog signalling in vertebrate neural development. *Nat Rev Neurosci* 2006; 7: 772–83.
- Gamble CE, Brule CE, Dean KM, Fields S, Grayhack EJ. Adjacent codons act in concert to modulate translation efficiency in yeast. *Cell* 2016; 166: 679–90.
- Gelfman S, Wang Q, McSweeney KM, Ren Z, La Carpia F, Halvorsen M, et al. Annotating pathogenic non-coding variants in gene regions. *Nat Commun* 2017; 8: 236.
- Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat Genet* 2014; 46: 818–25.
- Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. *Trends Biotechnol* 2004; 22: 346–53.
- Helmlinger D, Hardy S, Abou-Sleymane G, Eberlin A, Bowman AB, Gansmüller A, et al. Glutamine-expanded ataxin-7 Alters TFTC/STAGA recruitment and chromatin structure leading to photoreceptor dysfunction. *PLoS Biol* 2006; 4: e67.
- Jiang X, Nardelli J. Cellular and molecular introduction to brain development. *Neurobiol Dis* 2016; 92: 3–17.
- Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans [Internet]. *Genomics*; 2019. Available from: <http://biorxiv.org/lookup/doi/10.1101/531210>.
- Kim A, Savary C, Dubourg C, Carré W, Mouden C, Hamdi-Rozé H, et al. Integrated clinical and omics approach to rare diseases: novel genes and oligogenic inheritance in holoprosencephaly. *Brain* 2019; 142: 35–49.
- Kimchi-Sarfaty C, Oh JM, Kim I-W, Sauna ZE, Calcagno AM, Ambudkar SV, et al. A 'Silent' polymorphism in the MDR1 gene changes substrate specificity. *Science* 2007; 315: 525–8.
- Komar AA. The Yin and Yang of codon usage. *Hum Mol Genet* 2016; 25: R77–R85.
- Leman R, Gaildrat P, Gac GL, Ka C, Fichou Y, Audrezet M-P, et al. Novel diagnostic tool for prediction of variant spliceogenicity derived from a set of 395 combined in silico/in vitro studies: an international collaborative effort. *Nucleic Acids Res* 2018; 46: 7913–23.
- Macaya D, Katsanis SH, Hefferon TW, Audlin S, Mendelsohn NJ, Roggenbuck J, et al. A synonymous mutation in *TCOF1* causes Treacher Collins syndrome due to mis-splicing of a constitutive exon. *Am J Med Genet A* 2009; 149A: 1624–7.
- McCarthy C, Carrea A, Diambra L. Bicodon bias can determine the role of synonymous SNPs in human diseases. *BMC Genomics* 2017; 18: 227.

- Mercier S, Dubourg C, Garcelon N, Campillo-Gimenez B, Gicquel I, Belleguic M, et al. New findings for phenotype-genotype correlations in a large European series of holoprosencephaly cases. *J Med Genet* 2011; 48: 752–60.
- Mohanraj K, Wasilewski M, Benincá C, Cysewski D, Poznanski J, Sakowska P, et al. Inhibition of proteasome rescues a pathogenic variant of respiratory chain assembly factor COA7. *EMBO Mol Med* 2019; 11: e9561.
- Orioli IM, Castilla EE. Epidemiology of holoprosencephaly: prevalence and risk factors. *Am J Med Genet C Genet* 2010; 154C: 13–21.
- Poliakov E, Koonin EV, Rogozin IB. Impairment of translation in neurons as a putative causative factor for autism. *Biol Direct* 2014; 9: 16.
- Presnyak V, Alhusaini N, Chen Y-H, Martin S, Morris N, Kline N, et al. Codon optimality is a major determinant of mRNA stability. *Cell* 2015; 160: 1111–24.
- Quax TEF, Claassens NJ, Söll D, van der Oost J. Codon bias as a means to fine-tune gene expression. *Mol Cell* 2015; 59: 149–61.
- Rodriguez A, Wright G, Emrich S, Clark PL. %MinMax: a versatile tool for calculating and comparing synonymous codon usage and its impact on protein folding: codon analysis with %MinMax. *Protein Sci* 2018; 27: 356–62.
- Roessler E, El-Jaick KB, Dubourg C, Vélez JI, Solomon BD, Pineda-Alvarez DE, et al. The mutational spectrum of holoprosencephaly-associated changes within the SHH gene in humans predicts loss-of-function through either key structural alterations of the ligand or its altered synthesis. *Hum Mutat* 2009; 30: E921–935.
- Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet* 2011; 12: 683–91.
- Saunders R, Deane CM. Synonymous codon usage influences the local protein structure observed. *Nucleic Acids Res* 2010; 38: 6719–28.
- Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl Acids Res* 1987; 15: 7155–74.
- Sharp PM, Tuohy TMF, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucl Acids Res* 1986; 14: 5125–43.
- Soukariéh O, Gaildrat P, Hamieh M, Drouet A, Baert-Desurmont S, Frébourg T, et al. Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using in silico tools. *PLoS Genet* 2016; 12: e1005756.
- Tarailo-Graovac M, Zhu JYA, Matthews A, van Karnebeek CDM, Wasserman WW. Assessment of the ExAC data set for the presence of individuals with pathogenic genotypes implicated in severe Mendelian pediatric disorders. *Genet Med* 2017; 19: 1300–8.
- Thanaraj TA, Argos P. Ribosome-mediated translational pause and protein domain organization. *Protein Sci* 1996; 5: 1594–612.
- Tsai C-J, Sauna ZE, Kimchi-Sarfaty C, Ambudkar SV, Gottesman MM, Nussinov R. Synonymous mutations and ribosome stalling can lead to altered folding pathways and distinct minima. *J Mol Biol* 2008; 383: 281–91.
- Vaz SS, Chodirker B, Prasad C, Seabrook JA, Chudley AE, Prasad AN. Risk factors for nonsyndromic holoprosencephaly: A Manitoba case-control study. *Am J Med Genet A* 2012; 158A: 751–8.
- Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004; 11: 377–94.
- Zhao L, Zevallos SE, Rizzoti K, Jeong Y, Lovell-Badge R, Epstein DJ. Disruption of SoxB1-dependent Sonic Hedgehog expression in the hypothalamus causes septo-optic dysplasia. *Dev Cell* 2012; 22: 585–96.
- Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003; 31: 3406–15.

## 2. The double effect of c.1206C>T variant

As described in our article, the reduced protein levels were mostly accounted for by proteasome-mediated degradation of the mutated protein, indicating protein misfolding as the main pathogenic impact of the identified sSNVs. For the c.1206C>T variant, however, proteasome inhibition resulted in only a partial restoration of the protein levels. Additionally, qPCR experiments revealed that this particular variant is associated with significantly reduced mRNA levels as compared to the wildtype *SHH*. These results indicate that, in addition to protein misfolding and degradation, the c.1206C>T sSNV also has a deleterious impact at the mRNA level. Indeed, mFold simulations revealed statistically significant (albeit, borderline) differences in average predicted free energies ( $\Delta G$ ) for the c.1206C>T variant when compared to the wildtype *SHH* ( $P=0.01$ ). To further investigate the impact of this variant on mRNA folding, we performed additional computational analysis using the recently introduced RNA-stability pipeline (Gaither et al., 2019). RNA-stability estimates the impact of sSNVs on mRNA structure by combining four different metrics measuring the change in mRNA stability and evolutionary constraint: dMFE (Minimum Free Energy), CFEED (Centroid Free Energy Edge Distance), dCD (delta Centroid Distance) and SPI (Structural Predictivity Index). Similar to CADD score, calculations are performed for all possible SNVs of the human genome and variants presenting relatively high scores across the four metrics (>80 percentile, *i.e.* top 20% highest scores) are considered pathogenic. The results are presented in the Figure 28. Although no *SHH* sSNV has been predicted as pathogenic by the four algorithms, c.1206C>T variant has been predicted by 3/4 algorithms to have a deleterious impact on mRNA structure. These results are, however, inconclusive, as another *SHH* sSNV (c.552G>C) presented similar predictions. Although rejected in our study due to insufficient evidence, the hypothesis of mRNA-folding related defect cannot be fully excluded for the c.1206C>T variant without additional functional validations.



Variant	dMFE_abs_perc	CFEEED_perc	dCD_abs_perc	SPI_abs_perc	Number of deleterious predictions (Perc>80)
c.522C>T	26,6	24,8	11,4	32,3	0
c.552G>C	80,1	93,6	96,5	31,9	3
c.570G>A	50,9	89,4	82,7	63,2	2
c.630C>T	64,3	49,4	49,6	49,2	0
c.885C>T	42,7	80,7	86,8	72,8	2
c.897G>C	38,5	7,3	13,9	99	1
<b>c.1206C&gt;T</b>	<b>82,4</b>	<b>49,4</b>	<b>98,8</b>	<b>99,8</b>	<b>3</b>
c.1354C>T	98,2	79	94,7	8,5	2

**Figure 28. In silico predictions of RNA-stability algorithms for *SHH* sSNVs.**

Red and yellow dashed lines indicate, respectively, 80 and 99 percentile thresholds.

### 3. Rare codon clusters

In our study, we analyzed the global codon usage profile of *SHH* and identified four clusters of rare codons. As rare codons are associated with slow and non-optimal translation, such clusters are believed to cause translational pausing and reduce the local translation rate at strategic points beneficial for protein biogenesis (Clarke and Clark, 2008). Clusters of rare codons have previously been found in boundaries of protein functional domains and in regions rich in secondary mRNA structures, such as helices and beta strands (Clarke and Clark, 2010). Consistent with previous findings, our codon usage analysis of the *SHH* sequence revealed 2 clusters of rare codons in the



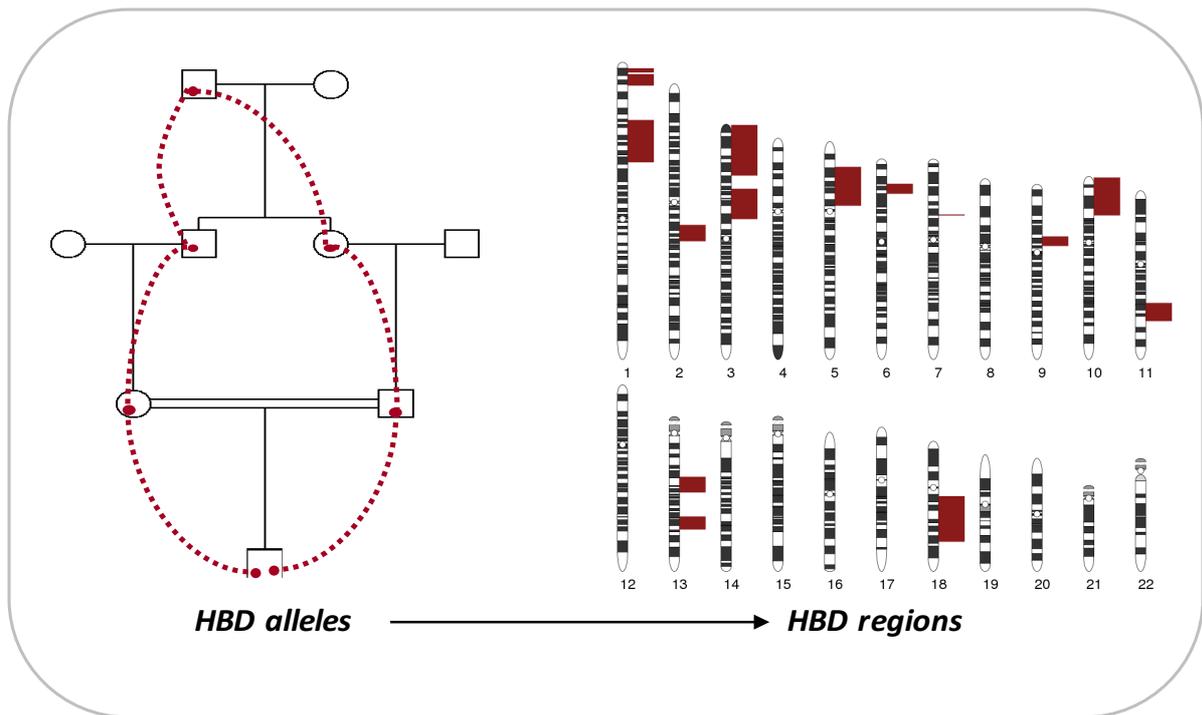
### III. Autosomal recessive holoprosencephaly in consanguineous families

#### a. Background

In 2016, Charlotte Mouden, the previous PhD student of our research team, identified a homozygous pathogenic mutation in *STIL* causing HPE in a consanguineous family (Mouden et al., 2016). Rare cases of homozygous variants underlying HPE have also been reported elsewhere (Kakar et al., 2015; McCabe et al., 2011). The existence of such cases suggests that autosomal recessive inheritance may account for a certain fraction of the disease etiology and can aid discovering novel disease genes. Hence, as the final part of my PhD, I continued the work initiated by Charlotte and further explored the hypothesis of autosomal recessive HPE in consanguineous context.

#### b. Strategy

In case of a recessive disorder and known consanguinity, the initial assumptions are that the disease-causing variant is likely to reside within a homozygous region in which both alleles of each locus are identical and inherited from the same common ancestor (Figure 30). Identification of such regions, called autozygous or Homozygous-by-Descent (HBD), can thus improve variant prioritization by pinpointing genomic segments likely to contain the causal mutation. HBD regions are typically identified by homozygosity mapping. This approach uses family pedigree and genotype information, obtained from exome or SNP-array data, to analyze inbreeding across the genome and identify HBD segments shared between the affected individuals (Keller et al., 2011). Coupled with WES analysis, homozygosity mapping has proven to be a powerful tool for discovery of genes underlying recessive disorders in consanguineous families (Jinks et al., 2015; Botstein and Risch, 2003).

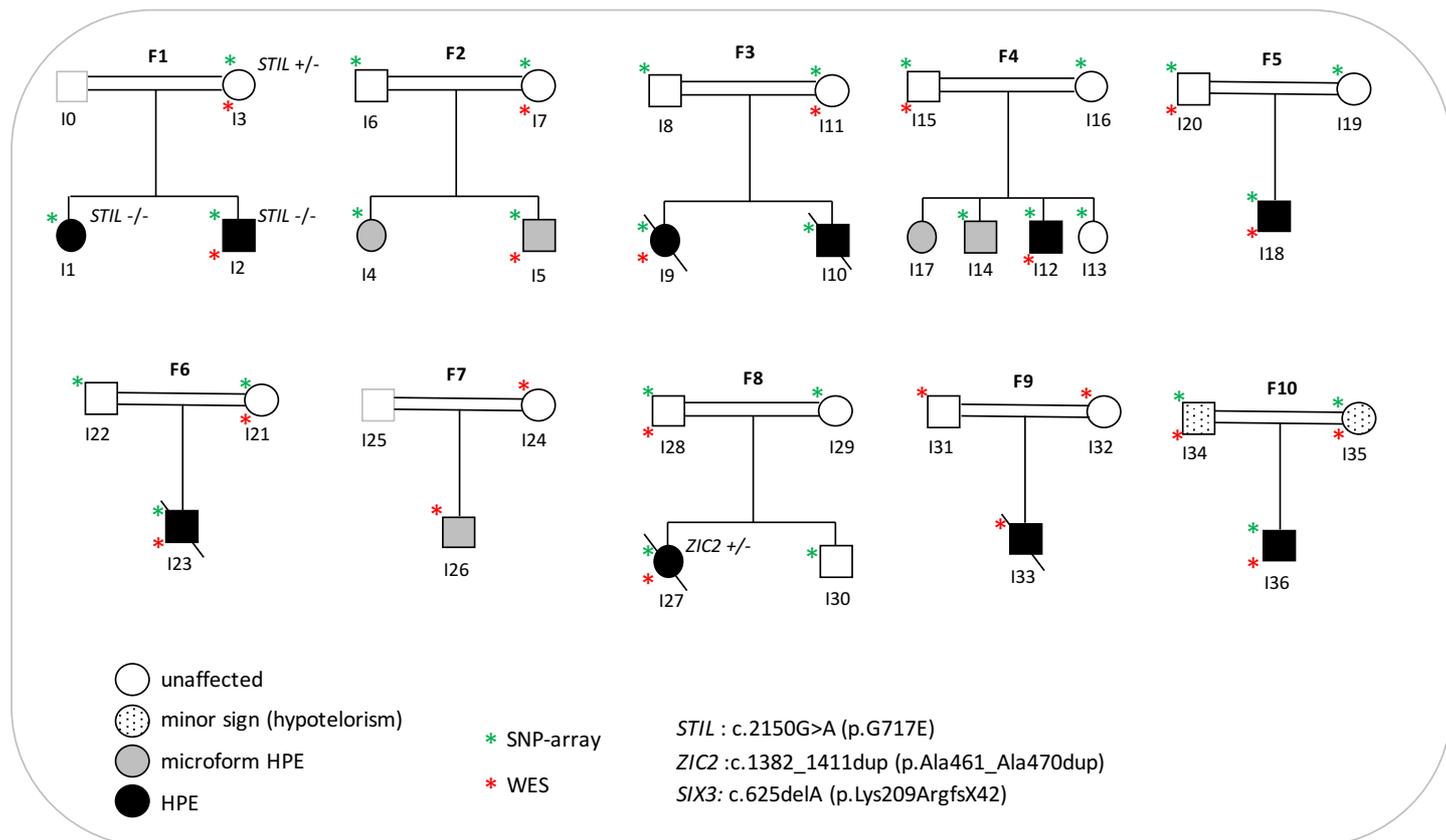


**Figure 30. HBD alleles and HBD regions.**

The aim of this study was to identify novel candidate genes implicated in HPE by exploring homozygous variants located in HBD segments of affected consanguineous individuals. We first performed homozygosity mapping and identified candidate HBD regions segregating with the disease in each family. Using the WES data, pathogenic variants underlying the disease were then searched in the identified candidate regions.

#### *Patients*

The study included 10 consanguineous families affected by HPE described in Figure 31. Family F1 was previously reported by Mouden *et al.* (F1, pathogenic variant in *STIL*) and included in this study as a positive control, to evaluate the ability of homozygosity mapping to identify regions harboring disease-causing variants. Variants in known disease genes were identified in two additional families (F8, F10), but were concluded to be insufficient to completely explain the disease phenotype. For each family, at least 1 affected and 1 unaffected individual were Whole Exome Sequenced for variant analysis. SNP-array genotyping data was available for 7/10 families.



**Figure 31. Summary of non-consanguineous families analyzed in this study.** Barred symbols represent deceased individuals. For families F1 and F7, paternal data (individuals I10 and I25) was unavailable.

### *WES and SNP-array genotyping*

WES analysis was conducted as previously described in Part I. SNP-array genotyping was performed using Human CytoSNP-12 (Illumina) technique, which includes 244,097 SNP markers distributed along the genome. The resulting genotype calls were quality controlled using plink v1.9 (Slifer, 2018). Specifically, only autosomal markers with genotyping rate of at least 99% across all individuals have been conserved. Additionally, Hardy-Weinberg equilibrium and Mendelian error controls were performed to exclude potentially erroneous markers. A total of 140797 markers have been conserved for further analysis.

### *Homozygosity mapping*

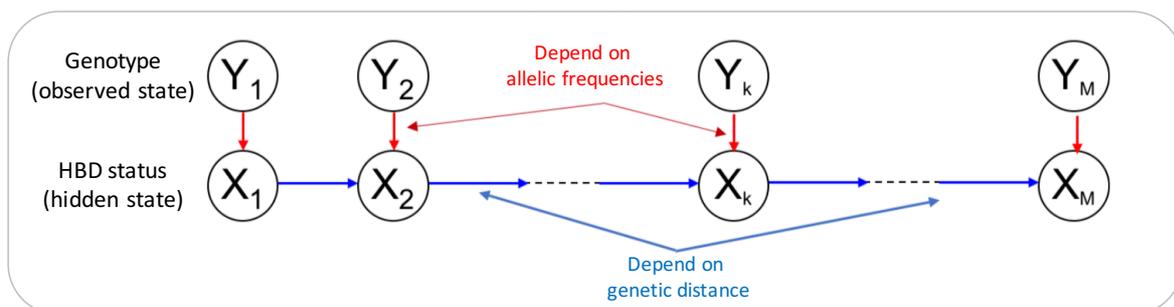
Depending on data availability, homozygosity mapping was performed using either SNP-array (8 families) or WES data (2 families). For this analysis, we used the FSuite

pipeline which allows identifying HBD regions from genomic data obtained through SNP array or WES experiments (Gazal et al., 2014). Identifying HBD regions by FSuite requires the estimation of the inbreeding coefficient and control for linkage disequilibrium. These steps are described below.

### *FEstim model*

Homozygosity mapping requires estimating the inbreeding coefficient  $f$  of each individual. Used in population studies to characterize mating habits,  $f$  reflects the individual's degree of consanguinity and corresponds to the estimated proportion of HBD segments in the genome (Gazal et al., 2016). For example, a coefficient  $f = 0.01$  corresponds to 1 % of genome being HBD. While  $f$  is classically calculated using genealogical information, several methods exist to estimate  $f$  from genomic data. FSuite is based on FEstim method described below (Leutenegger et al., 2003).

The challenge in homozygosity mapping is to distinguish the homozygous segments that are truly HBD (*i.e.*, linked to consanguinity and inherited from a common ancestor) from those that have arisen by chance (homozygosity by chance, HBC). FEstim is based on a Hidden Markov Model (HMM) which defines HBD regions by analyzing each SNP marker (denoted  $k$ ) and estimating its HBD status (denoted  $X_k$ ) (Figure 32).



**Figure 32. HMM model employed in FEstim.** FEstim models the dependencies between the genotypes of the SNP markers (observed state) and their HBD statuses (hidden state). Emission probabilities depend on allelic population frequencies; transition probabilities depend on genetic distances (in cM) separating adjacent markers.

Specifically, FEstim employs a maximum likelihood approach to calculate the probability of each marker  $k$  of being HBD ( $P(X_k = 1)$ ) or non-HBD ( $P(X_k = 0)$ ). These probabilities depend on 2 parameters (Table 7):

- 1) the genotype of the marker (denoted as  $Y_k$ ) which can be either homozygous ( $A_i A_i$ , 2 identical alleles  $i$ ) or heterozygous ( $A_i A_j$ , 2 different alleles  $i$  and  $j$ ).
- 2) the population frequencies of the corresponding alleles ( $p_i, p_j$ ), estimated directly from the samples or by using an external reference (e.g. GnomAD or ENSEMBL).

$Y_k$	$X_k = 0$	$X_k = 1$
$A_i A_i$	$p_i^2$	$(1 - \varepsilon)p_i + \varepsilon p_i^2$
$A_i A_j$	$2p_i p_j$	$\varepsilon 2p_i p_j$

**Table 7. Emission probabilities of HBD status.**

For example, a SNP with a homozygous genotype but presenting a very common allele is likely to be homozygous by chance. In contrast, a homozygous SNP with a rare allele is more likely to be homozygous due to consanguinity (*i.e.*, HBD). The model also takes into account the genotyping error rate (denoted  $\varepsilon$ ).

Thus, FEstim assigns to each SNP marker its probability of being HBD. HBD segments are then defined as regions with at least 5 consecutive SNP markers presenting a high HBD probability ( $P(X_k = 1) > 0.5$ ). Finally, the inbreeding coefficient  $f$  corresponds to the proportion of the resulting HBD segments in the genome. An individual with at least 0.1 % of its genome estimated HBD ( $f > 0.001$ ) is considered consanguineous by FEstim.

#### *Controlling the linkage disequilibrium*

As mentioned in the Introduction, Linkage Disequilibrium (LD) corresponds to the nonrandom association between different alleles driven by evolutionary forces such as genetic drift, natural selection and recombination events (Slatkin, 2008). As FEstim model assumes that there is no genetic interference, it requires SNP markers to be in linkage equilibrium (*i.e.*, without any preferential associations), which is not the case

when using dense SNP-array or WES data. To minimize the LD bias, FSuite generates random genomic sub-maps separating SNP markers presenting high LD. FEstim analysis is then performed separately on each submap, and the inbreeding coefficient  $f$  is calculated as the median value of estimations obtained across all sub-maps (denoted F\_MEDIAN).

#### *Identification of candidate regions*

As recommended by the authors, FSuite analysis was performed on a set of 100 random submaps with one marker every 0.5 cM to reduce the LD bias. Allelic population frequencies were obtained from the ENSEMBL database (GRCh37, release 100). The HBD segments identified by homozygosity mapping were further analyzed on a family-by-family basis to identify candidate regions likely to contain the pathogenic variant.

#### *Variant filtering and integration of homozygosity mapping*

For each family, we retrieved variants located in the candidate regions obtained via homozygosity mapping (Figure 33). Variants were annotated using ANNOVAR, as described in Part I. A variant filtering step was performed on a family-by-family basis. For this study, we retained only rare (GnomAD MAF < 1%) variants located in exonic/splicing regions and corresponding to the autosomal recessive inheritance pattern (homozygous in affected children; heterozygous in unaffected parent(s)). We manually explored the resulting candidates using genetic, genomic and biological data from literature reviews, genes associated with human disorders (OMIM) and mouse phenotypes (MGI database). Additionally, *in silico* algorithms were used to assess the deleterious impact of candidate variants.

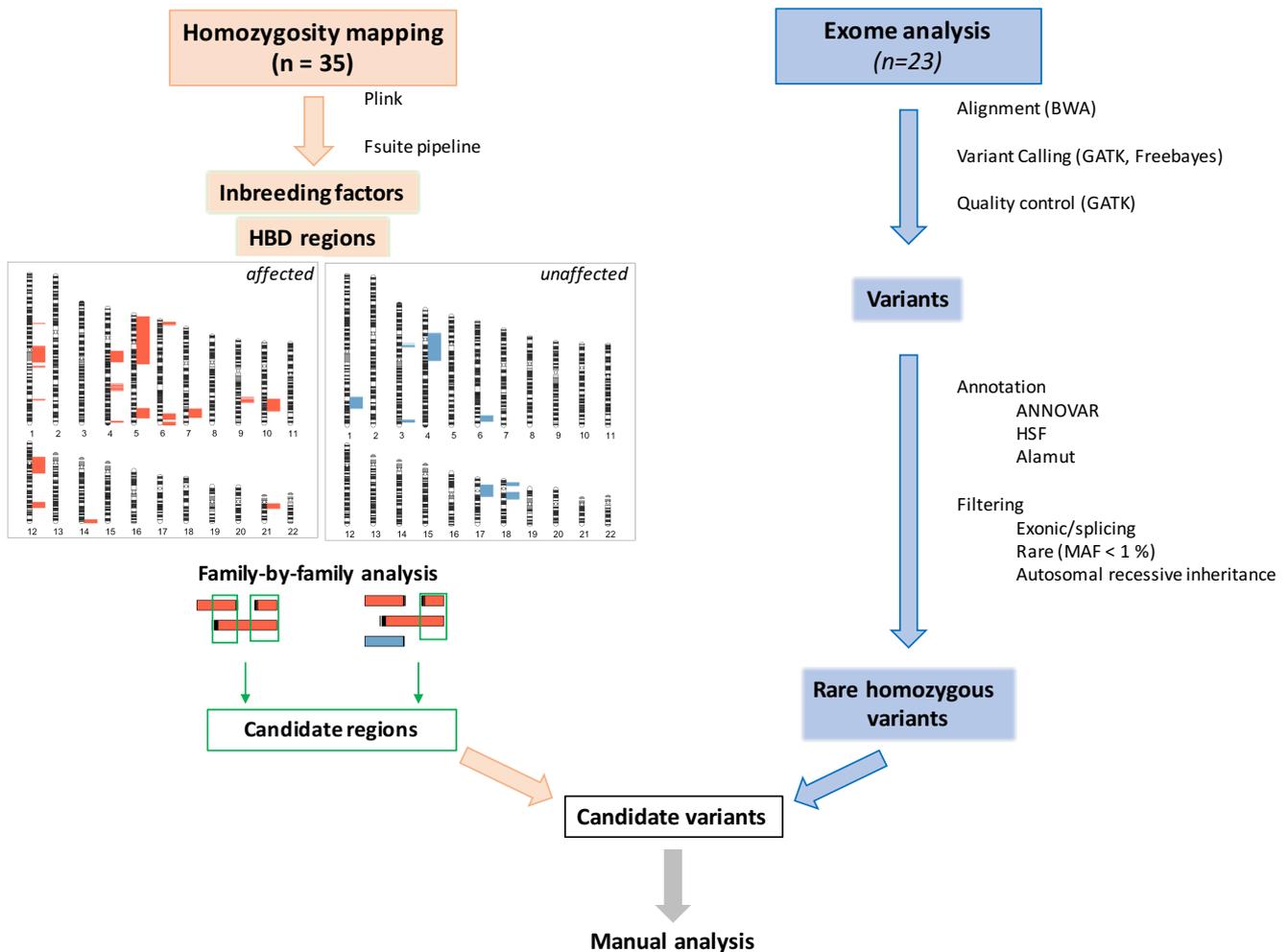


Figure 33. Workflow combining homozygosity mapping and WES analysis.

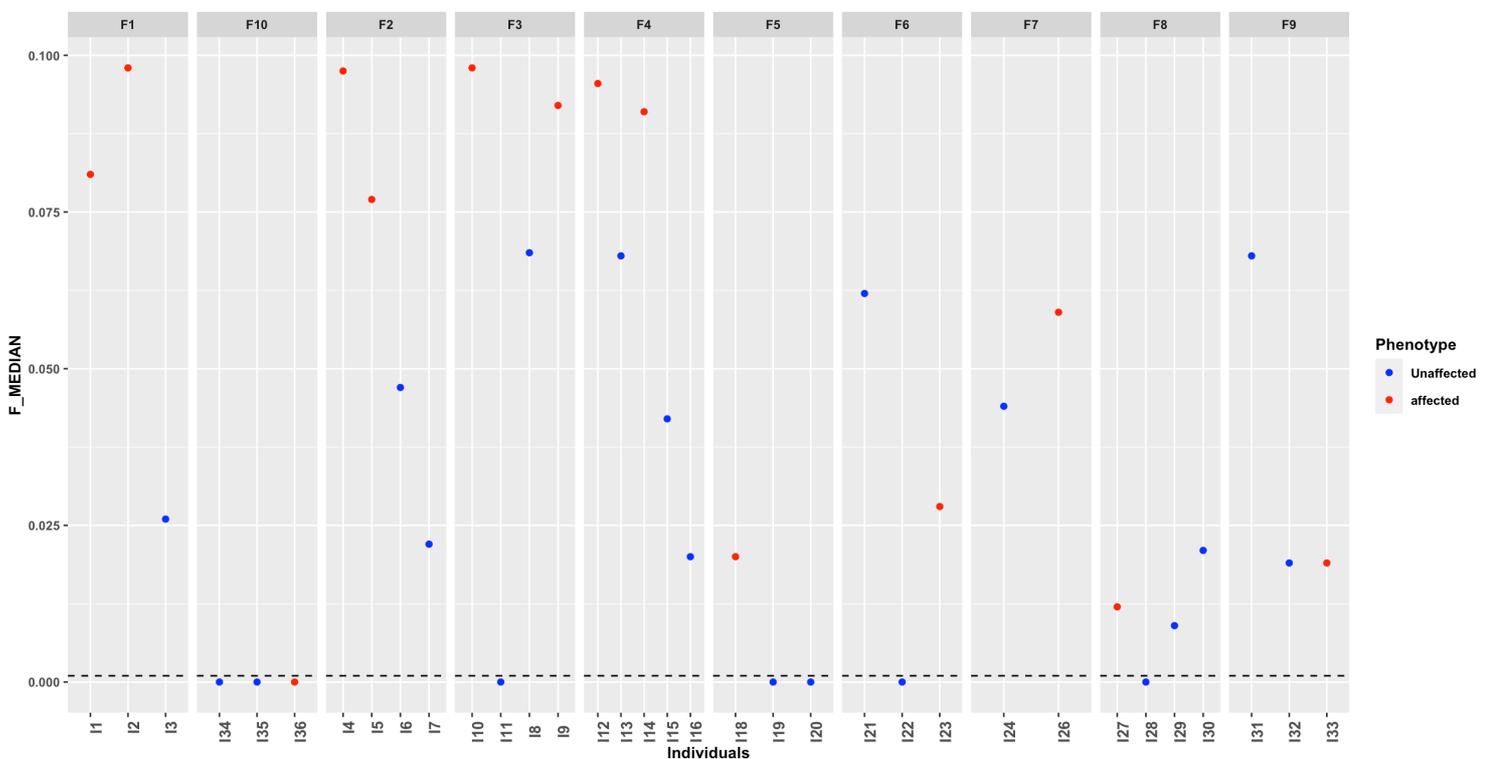
## c. Results

### 1. Consanguinity

We first calculated the inbreeding coefficient ( $f$ ) of each individual using the FEstim model (Figure 34). Inbreeding coefficient varied from 0 (a non-consanguineous individual) to 0.098 (an individual with ~10 % HBD fraction of the genome).

In family F5, only the affected child was estimated consanguineous ( $f > 0.001$ ), indicating a marriage between two related individuals without any prior consanguinity in the family. In families F2, F4 and F9, both children and their parents were estimated consanguineous, indicating a past history of consanguineous marriages and the

existence of additional ancestral inbreeding loops (Knight et al., 2008). Families F3, F6 and F8 correspond to ‘intermediate’ cases in which the children and only one of the parents were estimated consanguineous. For families F1 and F7, while the paternal data was not available, all genotyped individuals were estimated consanguineous. Family F10 was excluded from further analysis, as no individual was estimated consanguineous ( $f < 0.001$ ). Further inspection of family data revealed that family F10 corresponds to a marriage between very distant biological relatives, which could explain the absence of consanguinity estimated by FEstim.

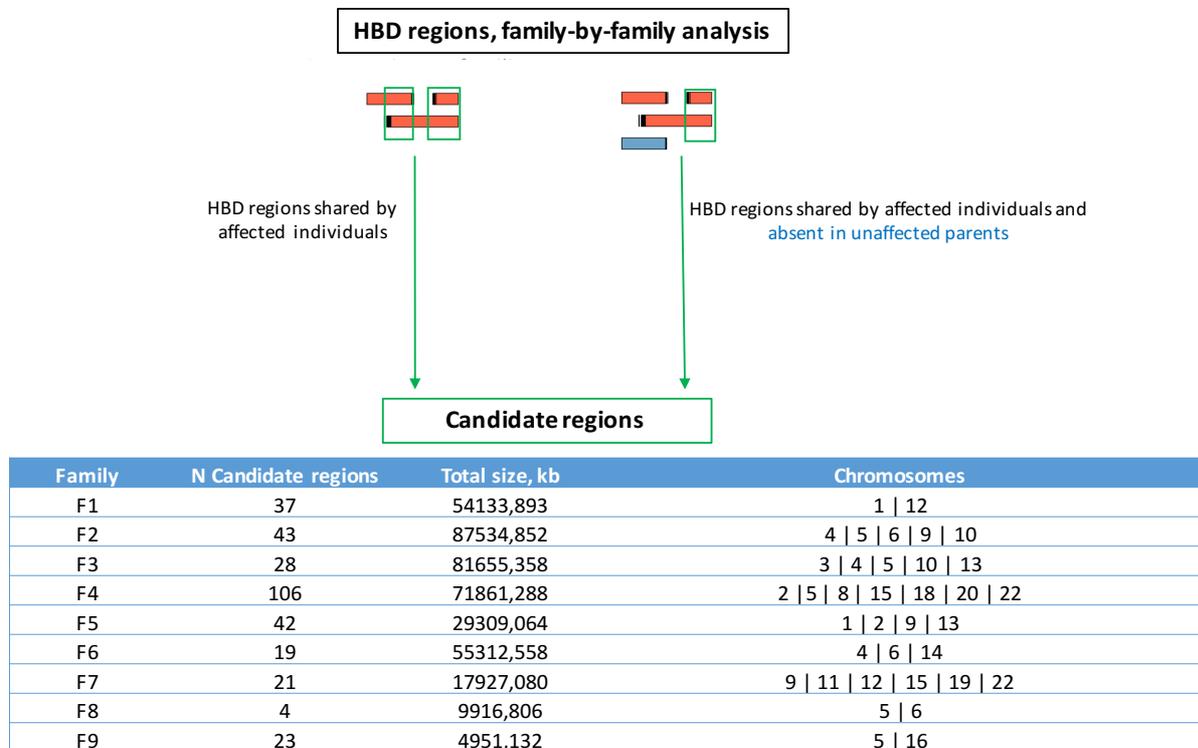


**Figure 34. FSuite estimations of inbreeding coefficient  $f$ .**

## 2. HBD candidate regions

The resulting inbreeding coefficients were used to estimate HBD segments and identify candidate regions likely to contain the pathogenic variant. Candidate regions were defined as HBD segments specific to the affected individual(s) of each family. Specifically, for family F5 presenting non-consanguineous parents, all HBD segments found in the affected subject were taken into account. For all other families, we eliminated HBD segments shared with unaffected consanguineous parents.

Additionally, for families presenting multiple affected subjects (F1, F2, F3, F4), we further restricted the candidate regions by focusing only on HBD segments shared by all the affected individuals (Figure 35).



**Figure 35. Overview of candidate regions identified by homozygosity mapping.**

The total size of the resulting candidate regions differs greatly between the studied families. As mentioned above, the fraction of HBD segments in the genome depends on the inbreeding coefficient. Additionally, restriction of HBD segments to candidate regions can be more or less efficient, depending on various factors such as parents' consanguinity and number of affected children in the family. For example, families F1-F4 present the highest inbreeding coefficients of the affected individuals (ranging from 0.077 to 0.098), resulting in large HBD segments and, consequently, a large size of candidate regions (54 - 87 Mb). On the other hand, family F6 presents lower inbreeding coefficients (0 to 0.062) but contains only one affected child and only one consanguineous parent, which results in a less efficient restriction and a similar size of candidate regions (55 Mb).

Family	Clinical Features	Variant			Frequency (gnomAD)	n HOM gnomAD	Effect	Disease relevance
		Gene	cDNA change	Protein change				
F1	semilobar HPE, hypotelorism, partial agenesis of corpus callosum	<i>STIL</i>	c.2150G>A	p.G717E	0	0	Deleterious missense variant (dbNSFP : 9 predictions) highly conserved nucleotide (GERP)	Previously identified pathogenic variant in known disease gene
		<i>SPEN</i>	c.2015G>A	p.R672Q	0,0000159	0	Deleterious missense variant (dbNSFP : 5 predictions) highly conserved nucleotide (GERP)	Interactor with known disease pathways (Notch)
F2	HPE microform, Single median maxillary incisor, hypotelorism, epicanthus	<i>DLL1</i>	c.2133C>T	p.S711S	0,00399	3	Synonymous variant altering splicing (HSF, creation of an exonic ESE site)	Known disease gene
		<i>IL6ST</i>	c.180A>G	p.V60V	0,0000517	0	Synonymous variant altering splicing (HSF, creation of an exonic ESS site)	Interactor with known disease pathways (SHH,Notch,FGF)
		<i>SKIV2L2</i>	c.2287A>G	.I763V	0,0000519	0	Deleterious missense variant (dbNSFP : 5 predictions)	Associated with disease phenotypes (reduced brain and eye size)
F3	Alobar HPE, Microcephaly, Median cleft lip/palate	<i>FAT1</i>	c.11195A>G	p.N3732S	0,000322	0	Deleterious missense variant (dbNSFP : 1 prediction) highly conserved nucleotide (GERP)	Known disease gene
		<i>PLCH1</i>	c.3211delT	p.C1071fs	0	0	Frameshift variant	Candidate pathway (Phospholipase C) linked to known disease mechanism (SHH)
F4	Semilobar HPE, Oral cleft	<i>KIF23</i>	c.933G>A	.Q311Q	0,0000915	0	Synonymous variant altering splicing (HSF, creation of an exonic cryptic acceptor site) highly conserved nucleotide (GERP)	Interactor with known disease pathways (SHH)
		<i>MYO9A</i>	c.1762G>C	p.D588H	0,0000719	1	Deleterious missense variant (dbNSFP : 13 predictions) highly conserved nucleotide (GERP)	Associated with disease phenotypes (hydrocephalus, ciliopathies)
F6	Alobar HPE, hypotelorism, Microcephaly, Proboscis	<i>SLC30A9</i>	c.197T>A	p.L66X	0	0	Stopgain variant	Candidate pathway (Wnt/PCP) linked to known disease mechanism (SHH, Fgf, Notch)
F8	Alobar HPE, fused thalami, partial agenesis of corpus callosum	<i>IQGAP2</i>	c.2197A>T	p.M733L	0,00000402	0	Deleterious missense variant (dbNSFP : 10 predictions) highly conserved nucleotide (GERP)	Interactor with known disease gene (Cdc42)
F9	Alobar HPE	<i>TBC1D24</i>	c.793G>A	p.265M	0,0000241	0	Deleterious missense variant (dbNSFP : 4 predictions) highly conserved nucleotide (GERP)	Candidate pathway (Wnt/PCP) linked to known disease mechanism (SHH, Fgf, Notch)

**Table 8. Overview of candidate variants identified in this study.**

### 3. Candidate variants

The candidate regions were integrated into our exome analysis pipeline to identify a restricted list of rare homozygous variants for further analysis. In family F1, previously described by Mouden et al. (Mouden et al., 2016), the pathogenic *STIL* variant was found in the candidate regions, indicating the capacity of homozygosity mapping to identify causal mutations underlying consanguineous cases of HPE. We manually explored candidate variants identified by our approach. Our main findings are summarized in Table 8 and the strongest candidate variants are described below.

#### *Variants in known disease genes*

Besides *STIL* variant in family F1, our analysis highlighted 2 additional variants in known disease genes. Initially discarded during molecular diagnosis, these variants were reconsidered due to their HBD status.

In family F2, the patient carried a homozygous synonymous variant in *DLL1*, previously implicated in disease-related processes. *DLL1* acts a ligand of NOTCH signaling pathway and its expression depends on the activity of the FGF pathway (Wahl et al., 2007; Dupé et al., 2011). Pathogenic mutations in this gene have been previously reported in HPE patients (Dupé et al., 2011). The detected variant is observed with a frequency of 0.004 in gnomAD and associated with a potential splicing defect by altering an exonic ESE site, as predicted by Human Splicing Finder. Although observed at homozygous state in 3 gnomAD individuals, these results indicate that this variant could participate in disease pathogenesis.

In family F3, we identified a homozygous missense variant in *FAT1*, a known regulator of Wnt/PCP pathway. Similar variants in this gene have been implicated in HPE under oligogenic model (Kim A et al., 2019; see Part I). The detected variant is extremely rare in general population (gnomAD frequency of 0.0003) and was never found at homozygous state. The variant is predicted deleterious by FATHMM-MKL and the nucleotide position is highly conserved in mammalian species (GERP score > 2), consistent with this variant having functional impact.

These results indicate that restricting the exome analysis to the candidate regions identified by homozygosity mapping enables to pinpoint pertinent homozygous variants in disease-related genes which would have been discarded otherwise.

### *PLCH1*

In family F3, we additionally identified a variant in *PLCH1*. The detected variant was absent in the gnomAD database and leads to a production of a truncated protein. *PLCH1* encodes a member of phosphoinositide-specific phospholipase C (PLC) family of enzymes, which plays an essential role in several signaling pathways (Hwang et al., 2005). Activated by various ligand-receptor systems, PLC enzymes induce the production of inositol-1,4,5-trisphosphate (IP3) and diacylglycerol, which act as signal transducers: IP3 induces an increase in intracellular Ca<sup>2+</sup> while diacylglycerol activates protein kinase C (PKC) (Kim et al., 2000; Berridge et al., 2003). The PLC enzymes and their products have been previously implicated in SHH signaling, the main disease pathway of HPE. In particular, IP3-controlled Ca<sup>2+</sup> activity has been linked to the activation of SHH pathway in developing neurons (Belgacem and Borodinsky, 2011), while PKC has been implicated in non-canonical SHH signaling involving the direct activation of downstream GLI targets bypassing the PTCH1-SMO mechanism (Yang et al., 2010; Gu and Xie, 2015). Although the precise interactions between SHH and PLC pathways are still poorly understood, variants in PLC-related genes could potentially contribute to disease mechanisms implicated in HPE.

### *SLC30A9*

In family F6, we identified a homozygous stopgain variant in *SLC30A9* gene, absent from gnomAD. *SLC30A9* is involved in the transcriptional regulation of canonical WNT pathway, which plays key roles during forebrain and craniofacial development along with SHH, NOTCH and FGF pathways (Chen et al., 2007). Mutations in Wnt genes and their downstream targets have been implicated in human craniofacial abnormalities including cleft lip/palate, ciliopathies and, very recently, HPE under oligogenic model (Vijayan et al., 2018; Waters and Beales, 2011; Kim A et al., 2019). Interestingly,

*SLC30A9* acts as a zinc transporter and zinc deficiency has been previously linked with alterations of SHH activity (Xie et al., 2015).

#### *IQGAP2*

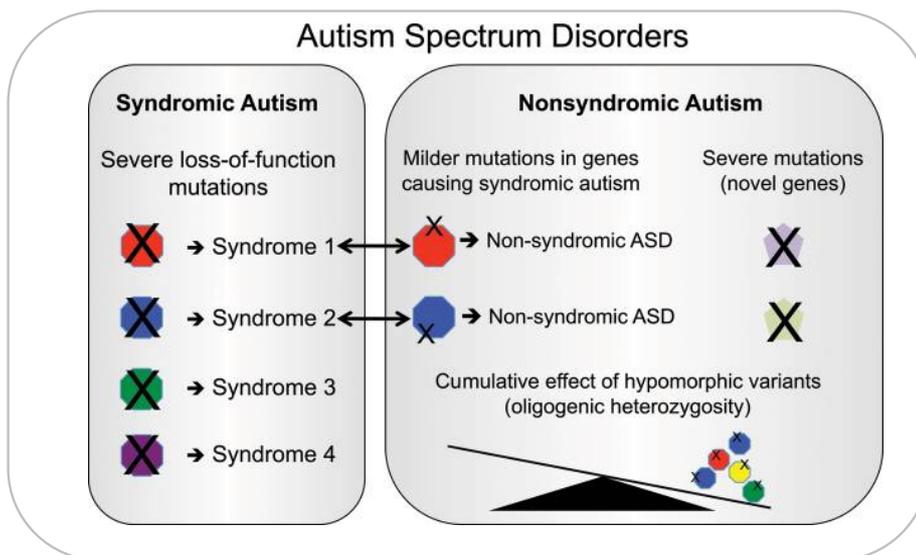
Patient of family F8 is affected by alobar HPE and harbored a homozygous missense variant in *IQGAP2*. The variant was found only once in gnomAD database (overall frequency of 0.00000402) and is predicted deleterious by several algorithms including SIFT, FATHMM-MKL, MutationTaster and MutationAssessor. *IQGAP2*, a scaffolding protein playing central roles in cell organization, binds to Cdc42 which was shown to induce Sonic Hedgehog-independent holoprosencephaly in mouse embryos (Ozdemir et al., 2018; Chen et al., 2006).

# ***DISCUSSION***

## I. Oligogenic inheritance in complex genetic disorders

My findings described in Part I illustrate the clinical relevance of oligogenic inheritance in HPE and concord with the advances in understanding of other complex disorders such as autism, NTDs and ciliopathies.

In non-syndromic autism, the hypothesis of a single locus inheritance has been rejected in favor of a multi-locus genetic model involving interactions between several susceptibility genes to produce the disease phenotype (Folstein et Rosen-Sheidley, 2001; Pickles et al., 1995). The concept termed as ‘oligogenic heterozygosity’ proposes that certain cases of ASD result from a cumulative effect of several hypomorphic variants, either inherited or occurring *de novo*, each conferring a moderate increase in the disease risk (Figure 36) (Schaaf et al., 2011). Similar to HPE, in which parents of affected children can exhibit minor disease signs (slight hypotelorism, epicanthus), parents of autistic children have been shown to manifest subthreshold autistic traits (Constantino and Todd, 2005). This observation is consistent with the oligogenic



**Figure 36. Proposed models of inheritance for syndromic and nonsyndromic autism.**

Multiple hypomorphic variants in genes associated with syndromic autism may have a cumulative effect, resulting in non-syndromic autism (Schaaf et al., 2011).

heterozygosity model and suggests that accumulation of hypomorphic variants up to a certain threshold can lead to a clinical disease manifestation.

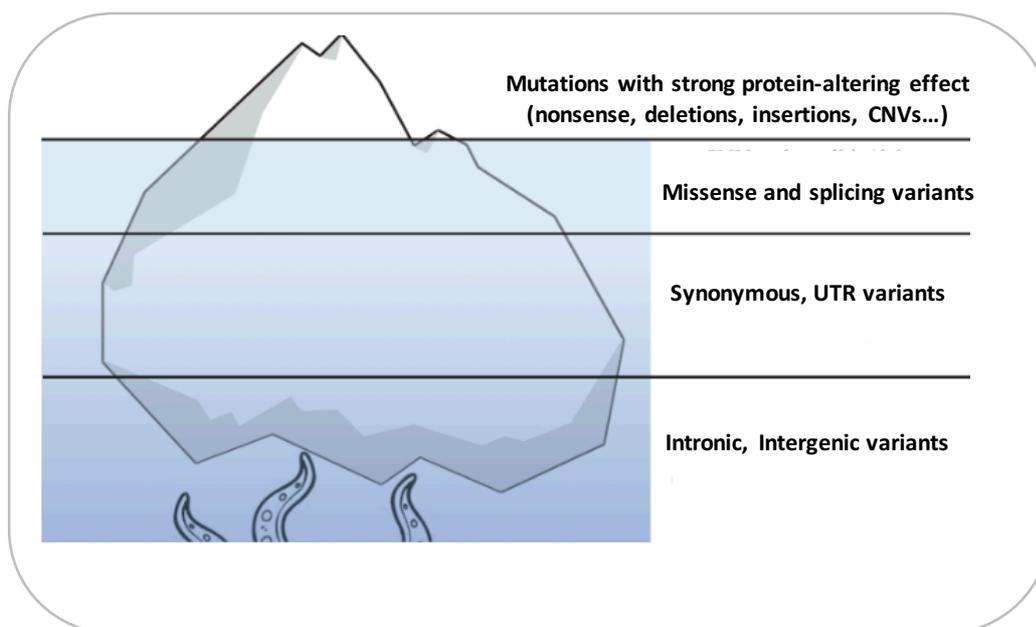
NTDs are also believed to involve oligogenic inheritance to a certain degree. In mouse models, genetic background has been shown to strongly affect the risk of NTD onset, indicating the essential role of genetic modifiers in the disease pathogenesis (Wilde et al., 2014). Extensive mouse studies have led to the identification of more than 300 genes, which, when mutated, cause various NTD phenotypes. Although most reported mouse mutants involve null mutations, some are hypomorphic and cause NTDs in digenic, trigenic or oligogenic combinations (Harris and Jurillof, 2010). Similar to HPE, most of oligogenic combinations in NTDs involve alterations of genes belonging to the same signaling pathway. For example, a simultaneous knockout of 3 genes of the PCP pathway - *Sfrp1*, *Sfrp2* and *Sfrp5* - causes a severe NTD phenotype in mice (Sato et al., 2008). These observations are consistent with findings in human NTD patients. Combined deleterious variants affecting multiple disease genes of the PCP pathway have been identified in human patients, suggesting a multi-locus etiology involving a cumulative effect of several genetic alterations (Beaumont et al., 2020; Wang et al. 2018).

Oligogenic inheritance has also been reported in ciliopathies. In addition to BBS discussed in the Introduction, clinical studies of Joubert and Meckel syndromes often identify rare deleterious variants in  $\geq 2$  known disease genes in a given patient (Phelps et al., 2018). Although the relevance of oligogenicity in ciliopathies remains unclear, digenic inheritance has been reported in a small number of families affected by nephronophthisis (Schaffer, 2013).

Overall, the results of my thesis concord with other studies underlining that oligogenic inheritance of hypomorphic variants contributes, at least in part, to the etiology of complex genetic disorders.

## II. Novel strategies for investigating the clinical impact of hypomorphic variants

While the strategies and criteria for identifying pathogenic variants with strong protein-altering effect are now well established (Richards et al., 2015), such mutations represent only the 'tip of the iceberg' of disease-causing variations of the genome (Figure 37). Establishing the pathogenicity of hypomorphic variants is challenging as they can be rapidly dismissed as benign due to their low deleterious effect and inheritance from asymptomatic family members (Leitch et al., 2008; Chapman et al., 2020; Bahi-Buisson et al., 2013). In my thesis, we attempt to address these challenges by proposing novel strategies to investigate the clinical impact of hypomorphic variants in complex disorders.



**Figure 37. Tip of the iceberg of disease-causing variants.** Novel strategies of clinical variant interpretation are needed to elucidate genetic mechanisms underlying complex genetic disorders.

### *Gene-based prioritization*

The integrative approach described in Part I allows for prioritization of hypomorphic variants by focusing on genes presenting significant association with the disease based on clinical and/or transcriptomic evidence. Indeed, a hypomorphic variant in a disease-related gene can play a greater pathogenic role than a null variant in a gene without

any relation to the disease mechanism. Both clinically-driven and transcriptome-driven prioritization approaches resulted in identification of novel disease genes. Several genes, such as *FAT1* and *NDST1*, were identified by clinical strategy due to their knockout in mice leading to HPE-related phenotypes. Certain genes, such as *SCUBE2* and *TCTN3*, were prioritized by transcriptome analysis due to their co-expression with known disease genes (*SHH*, *SIX3*, *TGIF1*, *ZIC2*) during embryonic development. Further investigation illustrated the relevance of the transcriptomic approach, as the co-expressed genes are also implicated in disease-related pathways (Thomas et al., 2012; Jakobs et al., 2014). Thus, co-expression analysis of relevant transcriptomic data can provide additional insight into disease pathogenesis by establishing the first link between previously unrelated genes.

A future challenge would be to generalize our approach to other complex disorders, but such a task will require incorporating gene-phenotype associations and transcriptomic data specific for each disease. Moreover, our strategy of gene prioritization can be further improved by implementing additional methods. For example, the disease implication of genes can be further investigated based on other network-based approaches, such as PPI networks, or using gene-phenotype associations of other model organisms, such as rat and zebrafish (Erten et al., 2011; Mashimo et al., 2005; Cheng et al., 2011).

#### *Predicting the impact of synonymous variants*

Our study described in Part II provides an analytical framework to investigate the pathogenic impact of sSNVs at both the mRNA and the protein level.

At the mRNA level, splicing-related defect can first be investigated by *in silico* tools, such as SPiCE and HSF, followed by *in vitro* validation of the most promising candidates using the minigene approach. Impact of sSNVs on miRNA-based regulation can be evaluated using *in silico* tools such as RegRNA, which can assess whether a particular variant alters a known or predicted regulatory site. For example, RegRNA has previously been used to investigate the miRNA-related impact of the c.313C>T sSNV in *IRGM*, which was shown to alter a miRNA binding site and influence gene expression

(Brest et al., 2011). Impact of sSNVs on mRNA folding can be explored by comparing the secondary structures of wildtype and mutant sequences predicted by mFold, as it has been done in our study. Additionally, several *in silico* tools, such as SilVA and DDIG-SN, can assess the impact of sSNVs on mRNA folding by automatically computing the variant-induced changes in free energy ( $\Delta G$ ) (Buske et al., 2013; Livingstone et al., 2017). Although the *SHH* variants analyzed in our study were not associated with any of the abovementioned defects, impacts at the mRNA level should always be considered when assessing the pathogenicity of sSNVs.

At the protein level, our study illustrated that deleterious impacts of sSNVs on translation can be accurately predicted using *in silico* estimations of codon usage. In particular, the variations of codon usage measured by a three-codon window analysis ( $\Delta$ MinMax measure) have shown a high correlation with the experimental measures of protein reduction. Although based on a small number of studied variants ( $n = 8$ ), the predictive power of our *in-silico* estimations is promising for future investigations. Additional studies are, however, required to assess what is the best *in silico* method to predict impacts of sSNVs on codon-mediated translational efficiency. Besides RSCU, RSBU and MinMax algorithms used in our study, codon optimality can be estimated by other algorithms such as tRNA adaptation index (tAI) or Codon Adaptation Index (CAI) (Sharp and Li, 1987; Reis et al., 2004). tAI score defines translationally optimal and non-optimal codons based on the estimated abundance of the corresponding tRNAs in the cell. However, the tRNA abundance is rarely measured experimentally and instead predicted *in silico*, using copy numbers of each tRNA gene as proxy (Wei et al., 2019). The CAI measure defines optimal codons as those which are preferentially found in highly expressed genes. As such, using CAI in clinical context can be challenging as it requires appropriate expression dataset for each study. Therefore, CAI and tAI measures are not based solely on codon usage frequencies and rely on additional biological data which, if not appropriate to the studied disease, can introduce a certain bias. Moreover, CAI and tAI scores have been mostly used to study codon bias at the scale of an entire gene, or even an entire organism (Pechmann and Frydman, 2013;

Jansen et al., 2003; Sahoo et al., 2019). Such measures can, therefore, be not sensitive enough to detect important variations of codon usage at a single variant level (Hockenberry et al., 2014).

### III. Autosomal recessive inheritance in HPE

My work described in part III explored the autosomal recessive inheritance in HPE and established a new method of variant prioritization in consanguineous families by homozygosity mapping. The identification of a previously reported pathogenic mutation in *STIL* (family F1) confirmed the ability of this approach to identify causal variants underlying consanguineous cases of HPE. Although our analysis didn't reveal any novel pathogenic variants presenting high evidence of causality, the reported genes, such as *PLCH1* and *SLC30A9*, can provide novel directions for the study of disease mechanisms underlying HPE. We underline the importance of investigating homozygous variants in HPE, especially in a consanguineous context. As illustrated by cases of homozygous *STIL* and *FGF8* mutations described previously, autosomal recessive inheritance may account for a certain part of HPE etiology (Mouden et al., 2015; Kakar et al., 2015; McCabe et al., 2011). Importantly, a recessive origin must also be considered in non-consanguineous families, as illustrated by the case of composite heterozygous mutations in *DISP1* identified in a HPE patient during molecular diagnosis (Dubourg et al., 2016; Mouden et al., 2016).

### IV. Perspectives

#### *Impact of non-coding variants in HPE*

Overall, the interpretation of impact of non-coding variants remains challenging despite numerous efforts. As such, pinpointing specific genomic regions in which non-coding variants can have a functional effect can be of great interest. As mentioned in the Introduction, certain non-coding regions, such as promoters and enhancers, act as crucial regulators of gene expression. In clinical context, non-coding variants affecting such regulatory regions can contribute to the disease pathogenesis by altering the

expression of genes involved in the disease etiology. With the advent and increasing availability of WGS in clinical setting, exploration of non-coding variants can substantially improve the molecular diagnosis of complex disorders.

Considering the central role of *SHH* in forebrain development, variants affecting its regulatory regions can be of particular interest in HPE. To date, seven enhancer regions regulating the *SHH* expression in brain, named SHH Brain Enhancer 1-7 (SBE1-SBE7), have been identified (Jeong et al., 2006; Yao et al., 2016; Sagai 2019; Benabdallah et al., 2016). Additionally, two regions regulating the *SHH* expression in the ventral midline of spinal cord and hindbrain, named *SHH* floorplate enhancer 1 and 2 (SFPE1, SFPE2), were reported (Jeong et al., 2008). Several rearrangements affecting these regions have been reported in HPE patients (Roessler et al., 1997). Additionally, a point mutation in SBE2 region, resulting in a reduced *SHH* expression, has been detected in a patient affected by semilobar HPE, (Jeong et al., 2008). Regulatory regions of other HPE genes, such as *SIX3*, *GLI2*, *FGF8* and *FGFR1* have also been reported and could represent a potential source of pathogenic non-coding variants (Parakati and DiMario, 2002; Marinic et al., 2013; Minhas et al., 2015; Nakanishi et al., 2016).

Moreover, *GLI2* and *GLI3*, the final effectors of the SHH signaling pathway, act as transcription factors (Armas-López et al., 2017). Identifying the downstream transcriptional targets of the SHH-GLI pathway and precise genomic locations of their target sites (*i.e.*, binding sites for *GLI2/GLI3* transcription factors) could potentially pinpoint novel genes and non-coding regions implicated in HPE pathogenesis.

#### *Common etiologies between developmental anomalies and cancers*

It is now well known that early embryonic development presents many similarities with tumorigenesis, at both molecular and genetic levels. With recent advances in genetics, the similarities between developmental and cancer biologies are increasingly being discovered, particularly in terms of epigenetic regulation and gene expression (Ma et al. 2010). The existence of children simultaneously affected by cancer and developmental abnormalities further indicates the relationship between the mechanisms of embryonic development and tumorigenesis (Ruyman et al., 1988).

Moreover, defects in developmental signaling pathways are also associated with various cancers. For example, the major HPE gene *SHH*, is also implicated in pediatric medulloblastoma, indicating its essential role in both forebrain development and tumorigenesis (Kool et al., 2014). Overall, the evidence accumulated to date makes it possible to propose a common etiological model in which genetic alterations would disrupt the expression programs of developmental genes, leading to oncogenesis and developmental anomalies (Vladoiu et al 2019).

In this context, our research team leads a multicentre genetic study EXOCARE (EXOME Childhood Rare Cancers), which aims to identify novel genes underlying both pediatric cancers and developmental defects. This project involves analysis of WES data of children simultaneously affected by cancer and developmental abnormalities. The EXOCARE study includes extensive analysis of the WES data to identify causal genetic variants and functional tests to assess the significance of the most relevant candidates. This project will allow a better characterization of molecular and genetic mechanisms involved in tumorigenesis and developmental defects, thus providing future directions for elucidating their common genetic etiology. In context of HPE, the results of this project will help further elucidate the molecular pathways implicated in both forebrain development and cancer, such as *SHH* (medulloblastoma), Notch (breast cancer, squamous cell carcinoma) and *WNT* (hepatocellular carcinoma, Wilm's tumor) (Aster et al., 2017; Polakis, 2012).

*Organoids: a novel method to explore the functional impact of genetic variants*

In the current state of knowledge, the diagnostic yield of complex disorders remains low and the identification of causal genetic factors is complicated by a large number of VUS, both coding and non-coding, found in every individual's genome. Moreover, as illustrated in this thesis, it is also important to consider oligogenic inheritance and the combined effect of several hypomorphic variants in multiple genes. Even in case of efficient variant prioritization and reliable *in silico* predictors, validating the causality of candidate variants requires exhaustive functional explorations which remain

challenging in the current state of knowledge and clinical means. Existing methods to functionally validate the disease implication of a candidate variant rely on genotype-phenotype correlations observed in animal studies or *in vitro* experiments to illustrate the deleterious impact on gene/protein function. Such approaches are very cost- and time-consuming as they require the establishment of one experimental system for each candidate variant/oligogenic combination.

Using NGS data, it is also possible to investigate the impact of a variant on the overall “omics state” (i.e., transcriptomic, metabolomic, epigenetic consequences), but such studies are often complicated by the lack of relevant biological material. For example, several studies have shown that 35 % to 80 % of mRNAs are present in both brain and blood samples, but their expression levels are poorly correlated between the two tissues (Tylee et al., 2013). In case of HPE, investigating the impact of a candidate variant on the transcriptome is challenging as the relevant tissues (e.g., embryonic brain) are often unavailable for functional studies.

To circumvent these limitations, our research team is now planning to explore a novel diagnostic method based on molecular signatures reflecting the overall biological state associated to the pathology. To do this, we plan to generate induced pluripotent stem cells (iPSCs) from patients' somatic cells and to differentiate them into the disease-relevant tissue (neuroectoderm and/or neural cell) generating the corresponding forebrain organoids (Cederquist et al., 2019). The idea is to define molecular signatures of the disease by comparing patient- and control-derived organoids at transcriptomic, epigenetic and metabolomics levels. Transcriptomic analyses will allow identifying genes presenting aberrant expression in disease cases. Metabolomic analyses should allow identification and quantification of metabolites, thus identifying the disease-associated metabolic consequences. Epigenetic signatures will allow us to investigate alterations of histone and/or chromatin accessibility. Initially, we will focus on determining molecular signatures for disease cases with established molecular etiology of high redundancy (e.g. cases with pathogenic mutations in *SHH*, *ZIC2*, *SIX3*

and *TGIF1*). This first step will allow us to study consequences of known pathogenic mutations at multiple biological levels and establish relevant biomarkers associated with various manifestations of the disease spectrum. In the future, it would be possible to test the disease implication of a candidate variant by genome editing the iPSCs and studying the impact on molecular signatures / biomarkers (*i.e.*, whether the variant-associated molecular signature corresponds to those previously established from the disease cases). This method has the potential to provide a novel functional framework of pathogenic variant interpretation. In particular, it can provide a necessary functional model to investigate the impact of oligogenic variants and further elucidate the genetic basis of complex disorders.

## Conclusion

The main findings of this thesis elucidate the complex genetic architecture of HPE and underline the necessity to investigate unconventional genetic mechanisms of disease pathogenesis, currently discarded by traditional diagnostic procedures. Such mechanisms include oligogenic inheritance of hypomorphic variants, pathogenic impact of sSNVs as well as the overall necessity to consider multiple inheritance patterns in diagnosis of HPE. To explore such mechanisms, this work also provides novel strategies and methods of pathogenic variant interpretation integrating appropriate clinical and biological data with NGS output. Ultimately, these results will help avoid misdiagnosis and greatly improve patient care in unelucidated genetic pathologies.

While significant progress has been made in deciphering the genetics and physiopathology of HPE, the disease arises after a set of complex and incompletely understood molecular events. Progress in understanding the etiology of complex disorders is increasingly driven by analysis of large scale, multi-omics data. Single omics and reductionist methods (*i.e.*, the “needle in a haystack” approach) can achieve only limited success in identifying disease-causing factors. Due to the genetic, clinical and

molecular complexity of HPE, future progress will require more holistic, multi-omics methods integrating heterogeneous biological and clinical data to address the interrelatedness of biological processes underlying forebrain development.

# ***REFERENCES***

Abe Y, Kruszka P, Martinez AF, Roessler E, Shiota K, Yamada S, et al. Clinical and Demographic Evaluation of a Holoprosencephaly Cohort From the Kyoto Collection of Human Embryos. *Anat Rec (Hoboken)* 2018; 301: 973–86.

Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Current Protocols in Human Genetics* 2013; 76: 7.20.1-7.20.41.

Aldhous MC, Satsangi J. The impact of smoking in Crohn's disease: no smoke without fire. *Frontline Gastroenterol* 2010; 1: 156–64.

Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nature Biotechnology* 2015; 33: 831–8.

Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Res* 2015; 43: D789-798.

Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015; 31: 166–9.

Anderson JL, Waller DK, Canfield MA, Shaw GM, Watkins ML, Werler MM. Maternal Obesity, Gestational Diabetes, and Central Nervous System Birth Defects: *Epidemiology* 2005; 16: 87–92.

Antonarakis SE, Beckmann JS. Mendelian disorders deserve more attention. *Nature Reviews Genetics* 2006; 7: 277–82.

Antonarakis SE, Krawczak M, Cooper DN. Disease-causing mutations in the human genome. *European Journal of Pediatrics* 2000; 159: S173–8.

Apweiler R. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Research* 2004; 32: 115D – 119.

Arauz RF, Solomon BD, Pineda-Alvarez DE, Gropman AL, Parsons JA, Roessler E, et al. A Hypomorphic Allele in the *FGF8* Gene Contributes to Holoprosencephaly and Is Allelic to Gonadotropin-Releasing Hormone Deficiency in Humans. *Molecular Syndromology* 2010; 1: 59–66.

Armas-López L, Zúñiga J, Arrieta O, Ávila-Moreno F. The Hedgehog-Gli pathway in embryonic development and cancer: implications for pulmonary oncology therapy. *Oncotarget* 2017; 8: 60684–703.

Armour CM, Smith A, Hartley T, Chardon JW, Sawyer S, Schwartzentruber J, et al. Syndrome disintegration: Exome sequencing reveals that Fitzsimmons syndrome is a co-occurrence of multiple events: Syndrome Disintegration: Exome Sequencing Reveals. *American Journal of Medical Genetics Part A* 2016; 170: 1820–5.

Artavanis-Tsakonas S. Notch Signaling: Cell Fate Control and Signal Integration in Development. *Science* 1999; 284: 770–6.

Aster JC, Pear WS, Blacklow SC. The Varied Roles of Notch in Cancer. *Annu Rev Pathol* 2017; 12: 245–75.

Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline [Internet]. *Current Protocols in Bioinformatics* 2013; 43[cited 2020 May 25] Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/0471250953.bi1110s43>

Aykul S, Martinez-Hackert E. Transforming Growth Factor- $\beta$  Family Ligands Can Function as Antagonists by Competing for Type II Receptor Binding. *Journal of Biological Chemistry* 2016; 291: 10792–804.

Bachiller D, Klingensmith J, Kemp C, Belo JA, Anderson RM, May SR, et al. The organizer factors Chordin and Noggin are required for mouse forebrain development. *Nature* 2000; 403: 658–61.

Bachman H, Clark RD, Salahi W. Holoprosencephaly and polydactyly: a possible expression of the hydrolethalus syndrome. *J Med Genet* 1990; 27: 50–2.

Badano JL, Katsanis N. Beyond Mendel: an evolving view of human genetic disease transmission. *Nature Reviews Genetics* 2002; 3: 779–89.

Badano JL, Mitsuma N, Beales PL, Katsanis N. The Ciliopathies: An Emerging Class of Human Genetic Disorders. *Annual Review of Genomics and Human Genetics* 2006; 7: 125–48.

Bae G-U, Domené S, Roessler E, Schachter K, Kang J-S, Muenke M, et al. Mutations in CDON, Encoding a Hedgehog Receptor, Result in Holoprosencephaly and Defective Interactions with Other Hedgehog Receptors. *The American Journal of Human Genetics* 2011; 89: 231–40.

Bahi-Buisson N, Souville I, Fourniol FJ, Toussaint A, Moores CA, Houdusse A, et al. New insights into genotype-phenotype correlations for the doublecortin-related lissencephaly spectrum. *Brain* 2013; 136: 223–44.

Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics* 2011; 12: 745–55.

Baron M. The search for complex disease genes: fault by linkage or fault by association? *Molecular Psychiatry* 2001; 6: 143–9.

Bartoszewski RA, Jablonsky M, Bartoszezewska S, Stevenson L, Dai Q, Kappes J, et al. A synonymous single nucleotide polymorphism in DeltaF508 CFTR alters the secondary structure of the mRNA and the expression of the mutant protein. *J Biol Chem* 2010; 285: 28741–8.

Baynam G, Pachter N, McKenzie F, Townshend S, Slee J, Kiraly-Borri C, et al. The rare and undiagnosed diseases diagnostic service – application of massively parallel sequencing in a state-wide clinical service [Internet]. *Orphanet Journal of Rare Diseases* 2016; 11[cited 2020 Jun 20] Available from: <http://ojrd.biomedcentral.com/articles/10.1186/s13023-016-0462-7>

Bear KA, Solomon BD, Antonini S, Arnhold IJP, França MM, Gerkes EH, et al. Pathogenic mutations in *GLI2* cause a specific phenotype that is distinct from holoprosencephaly. *Journal of Medical Genetics* 2014; 51: 413–8.

Belgacem YH, Borodinsky LN. Sonic hedgehog signaling is decoded by calcium spike activity in the developing spinal cord. *Proceedings of the National Academy of Sciences* 2011; 108: 4482–7.

Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proceedings of the National Academy of Sciences* 2015; 112: 5473–8.

Belloni E, Muenke M, Roessler E, Traverse G, Siegel-Bartelt J, Frumkin A, et al. Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly. *Nature Genetics* 1996; 14: 353–6.

Benabdallah NS, Gautier P, Hekimoglu-Balkan B, Lettice LA, Bhatia S, Bickmore WA. SBE6: a novel long-range enhancer involved in driving sonic hedgehog expression in neural progenitor cells. *Open Biology* 2016; 6: 160197.

Bendavid C. Multicolour FISH and quantitative PCR can detect submicroscopic deletions in holoprosencephaly patients with a normal karyotype. *Journal of Medical Genetics* 2006; 43: 496–500.

Bendavid C, Dupé V, Rochard L, Gicquel I, Dubourg C, David V. Holoprosencephaly: An update on cytogenetic abnormalities. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 2010; 154C: 86–92.

Bendavid C, Rochard L, Dubourg C, Seguin J, Gicquel I, Pasquier L, et al. Array-CGH analysis indicates a high prevalence of genomic rearrangements in holoprosencephaly: An updated map of candidate loci. *Human Mutation* 2009; 30: 1175–82.

Berridge MJ, Bootman MD, Roderick HL. Calcium signalling: dynamics, homeostasis and remodelling. *Nature Reviews Molecular Cell Biology* 2003; 4: 517–29.

Bertrand N, Dahmane N. Sonic hedgehog signaling in forebrain development and its interactions with pathways that modify its effects. *Trends in Cell Biology* 2006; 16: 597–605.

Bhatia S, Drake DM, Miller L, Wells PG. Oxidative stress and DNA damage in the mechanism of fetal alcohol spectrum disorders. *Birth Defects Res* 2019; 111: 714–48.

Biesecker LG, Abbott M, Allen J, Clericuzio C, Feuillan P, Graham JM, et al. Report from the workshop on Pallister-Hall syndrome and related phenotypes. *American Journal of Medical Genetics* 1996; 65: 76–81.

Blake JA. MGD: the Mouse Genome Database. *Nucleic Acids Research* 2003; 31: 193–5.

Blaxter M. Revealing the Dark Matter of the Genome. *Science* 2010; 330: 1758–9.

Boareto M, Jolly MK, Lu M, Onuchic JN, Clementi C, Ben-Jacob E. Jagged-Delta asymmetry in Notch signaling can give rise to a Sender/Receiver hybrid phenotype. *Proc Natl Acad Sci USA* 2015; 112: E402-409.

Bohlander SK. ABCs of genomics. *Hematology* 2013; 2013: 316–23.

Bond AM, Bhalala OG, Kessler JA. The dynamic role of bone morphogenetic proteins in neural stem cell fate and maturation. *Dev Neurobiol* 2012; 72: 1068–84.

Bonilla-Claudio M, Wang J, Bai Y, Klysik E, Selever J, Martin JF. Bmp signaling regulates a dose-dependent transcriptional program to control facial skeletal development. *Development* 2012; 139: 709–19.

Borrego S, Ruiz-Ferrer M, Fernández RM, Antiñolo G. Hirschsprung's disease as a model of complex genetic etiology. *Histol Histopathol.* 2013;28(9):1117-1136.

Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genetics* 2003; 33: 228–37.

Brancati F, Mingarelli R, Dallapiccola B. Recurrent triploidy of maternal origin. *European Journal of Human Genetics* 2003; 11: 972–4.

Brandão RD, Mensaert K, López-Perolio I, Tserpelis D, Xenakis M, Lattimore V, et al. Targeted RNA-seq successfully identifies normal and pathogenic splicing events in breast/ovarian cancer susceptibility and Lynch syndrome genes: Splicing events in breast/ovarian cancer. *International Journal of Cancer* 2019; 145: 401–14.

Brest P, Lapaquette P, Souidi M, Lebrigand K, Cesaro A, Vouret-Craviari V, et al. A synonymous variant in IRGM alters a binding site for miR-196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. *Nature Genetics* 2011; 43: 242–5.

Briscoe J, Théron PP. The mechanisms of Hedgehog signalling and its roles in development and disease. *Nature Reviews Molecular Cell Biology* 2013; 14: 416–29.

Brooks A, Oostra B, Hofstra R. Studying the genetics of Hirschsprung's disease: unraveling an oligogenic disorder: Genetics of Hirschsprung's disease. *Clinical Genetics* 2004; 67: 6–14.

Brown LY. Holoprosencephaly due to mutations in ZIC2: alanine tract expansion mutations may be caused by parental somatic recombination. *Human Molecular Genetics* 2001; 10: 791–6.

Brown SA, Warburton D, Brown LY, Yu C, Roeder ER, Stengel-Rutkowski S, et al. Holoprosencephaly due to mutations in ZIC2, a homologue of Drosophila odd-paired. *Nature Genetics* 1998; 20: 180–3.

Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research* 2019; 47: W402–7.

Buhr F, Jha S, Thommen M, Mittelstaet J, Kutz F, Schwalbe H, et al. Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. *Molecular Cell* 2016; 61: 341–51.

Bureau A, Parker MM, Ruczinski I, Taub MA, Marazita ML, Murray JC, et al. Whole exome sequencing of distant relatives in multiplex families implicates rare variants in candidate genes for oral clefts. *Genetics* 2014; 197: 1039–44.

Burgess DJ. Deciphering non-coding variation with 3D epigenomics. *Nature Reviews Genetics* 2017; 18: 4–4.

Buske OJ, Manickaraj A, Mital S, Ray PN, Brudno M. Identification of deleterious synonymous variants in human genomes. *Bioinformatics* 2013; 29: 1843–50.

Cabral RM, Kurban M, Wajid M, Shimomura Y, Petukhova L, Christiano AM. Whole-exome sequencing in a single proband reveals a mutation in the CHST8 gene in autosomal recessive peeling skin syndrome. *Genomics* 2012; 99: 202–8.

Camera G, Lituania M, Cohen MM. Holoprosencephaly and primary craniosynostosis: The Genoa syndrome. *American Journal of Medical Genetics* 1993; 47: 1161–5.

Campbell IM, Rao M, Arredondo SD, Lalani SR, Xia Z, Kang S-HL, et al. Fusion of Large-Scale Genomic Knowledge and Frequency Data Computationally Prioritizes Variants in Epilepsy. *PLoS Genetics* 2013; 9: e1003797.

Candia AF, Watabe T, Hawley SH, Onichtchouk D, Zhang Y, Derynck R, et al. Cellular interpretation of multiple TGF-beta signals: intracellular antagonism between activin/BVg1 and BMP-2/4 signaling mediated by Smads. *Development* 1997; 124: 4467–80.

Carrasco H, Olivares GH, Faunes F, Oliva C, Larraín J. Heparan sulfate proteoglycans exert positive and negative effects in Shh activity. *J Cell Biochem* 2005; 96: 831–8.

Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Reviews Genetics* 2002; 3: 285–98.

- Carter H, Douville C, Stenson PD, Cooper DN, Karchin R. Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics* 2013; 14 Suppl 3: S3.
- Cau E, Blader P. Notch activity in the nervous system: to switch or not switch? *Neural Development* 2009; 4: 36.
- Cederquist GY, Ascioia JJ, Tchieu J, Walsh RM, Cornacchia D, Resh MD, et al. Specification of positional identity in forebrain organoids. *Nature Biotechnology* 2019; 37: 436–44.
- Chamary JV, Parmley JL, Hurst LD. Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nature Reviews Genetics* 2006; 7: 98–108.
- Chang DT, López A, von Kessler DP, Chiang C, Simandl BK, Zhao R, et al. Products, genetic linkage and limb patterning activity of a murine hedgehog gene. *Development* 1994; 120: 3339–53.
- Chao H-K, Hsiao K-J, Su T-S. A silent mutation induces exon skipping in the phenylalanine hydroxylase gene in phenylketonuria. *Human Genetics* 2001; 108: 14–9.
- Chapman G, Moreau JLM, I P E, Szot JO, Iyer KR, Shi H, et al. Functional genomics and gene-environment interaction highlight the complexity of congenital heart disease caused by Notch pathway variants. *Human Molecular Genetics* 2020; 29: 566–79.
- Chapman JM, Cooper JD, Todd JA, Clayton DG. Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. *Hum Hered* 2003; 56: 18–31.
- Chen C-P, Huang J-P, Chen Y-Y, Chern S-R, Wu P-S, Su J-W, et al. Chromosome 18p deletion syndrome presenting holoprosencephaly and premaxillary agenesis: Prenatal diagnosis and aCGH characterization using uncultured amniocytes. *Gene* 2013; 527: 636–41.
- Chen L, Liao G, Yang L, Campbell K, Nakafuku M, Kuan C-Y, et al. Cdc42 deficiency causes Sonic hedgehog-independent holoprosencephaly. *Proceedings of the National Academy of Sciences* 2006; 103: 16520–5.
- Chen R, Shi L, Hakenberg J, Naughton B, Sklar P, Zhang J, et al. Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. *Nature Biotechnology* 2016; 34: 531–8.
- Chen VS, Morrison JP, Southwell MF, Foley JF, Bolon B, Elmore SA. *Histology Atlas of the Developing Prenatal and Postnatal Mouse Central Nervous System, with Emphasis on Prenatal Days E7.5 to E18.5*. *Toxicol Pathol* 2017; 45: 705–44.
- Chen X, Tukachinsky H, Huang C-H, Jao C, Chu Y-R, Tang H-Y, et al. Processing and turnover of the Hedgehog protein in the endoplasmic reticulum. *The Journal of Cell Biology* 2011; 192: 825–38.
- Chen X, Zhao K, Sheng X, Li Y, Gao X, Zhang X, et al. Targeted Sequencing of 179 Genes Associated with Hereditary Retinal Dystrophies and 10 Candidate Genes Identifies Novel and Known Mutations in Patients with Various Retinal Diseases. *Investigative Ophthalmology & Visual Science* 2013; 54: 2186.
- Chen Y-H, Yang CK, Xia M, Ou C-Y, Stallcup MR. Role of GAC63 in transcriptional activation mediated by -catenin. *Nucleic Acids Research* 2007; 35: 2084–92.
- Cheng KC, Xin X, Clark DP, La Riviere P. Whole-animal imaging, gene function, and the Zebrafish Phenome Project. *Current Opinion in Genetics & Development* 2011; 21: 620–9.
- Chiang C, Litingtung Y, Lee E, Young KE, Corden JL, Westphal H, et al. Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* 1996; 383: 407–13.
- Chitnis A, Kintner C. Sensitivity of proneural genes to lateral inhibition affects the pattern of primary neurons in *Xenopus* embryos. *Development* 1996; 122: 2295–301.
- Choudhry Z, Rikani AA, Choudhry AM, Tariq S, Zakaria F, Asghar MW, et al. Sonic hedgehog signalling pathway: a complex network [Internet]. *Annals of Neurosciences* 2014; 21[cited 2020 Jun 25] Available from: <http://annalsofneurosciences.org/journal/index.php/annal/article/view/531>
- Chuang P-T. Feedback control of mammalian Hedgehog signaling by the Hedgehog-binding protein, Hip1, modulates Fgf signaling during branching morphogenesis of the lung. *Genes & Development* 2003; 17: 342–7.
- Cirino AL, Lakdawala NK, McDonough B, Conner L, Adler D, Weinfeld M, et al. A Comparison of Whole Genome Sequencing to Multigene Panel Testing in Hypertrophic Cardiomyopathy Patients [Internet]. *Circulation: Cardiovascular Genetics* 2017; 10[cited 2020 May 14] Available from: <https://www.ahajournals.org/doi/10.1161/CIRCGENETICS.117.001768>

Cirulli ET, Lasseigne BN, Petrovski S, Sapp PC, Dion PA, Leblond CS, et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* 2015; 347: 1436–41.

Clarke TF, Clark PL. Rare Codons Cluster. *PLoS ONE* 2008; 3: e3412.

Clarke TF, Clark PL. Increased incidence of rare codon clusters at 5' and 3' gene termini: implications for function. *BMC Genomics* 2010; 11: 118.

Collins FS. The Human Genome Project: Lessons from Large-Scale Biology. *Science* 2003; 300: 286–90.

Comings DE. Polygenic inheritance and micro/minisatellites. *Molecular Psychiatry* 1998; 3: 21–31.

Constantino JN, Todd RD. Intergenerational transmission of subthreshold autistic traits in the general population. *Biological Psychiatry* 2005; 57: 655–60.

Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Human Genetics* 2013; 132: 1077–130.

Cooper GM, Shendure J. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nature Reviews Genetics* 2011; 12: 628–40.

Cornish AJ, David A, Sternberg MJE. PhenoRank: reducing study bias in gene prioritization through simulation. *Bioinformatics* 2018; 34: 2087–95.

Correa A, Gilboa SM, Besser LM, Botto LD, Moore CA, Hobbs CA, et al. Diabetes mellitus and birth defects. *American Journal of Obstetrics and Gynecology* 2008; 199: 237.e1-237.e9.

de la Cruz JM, Bamford RN, Burdine RD, Roessler E, Barkovich JA, Donnai D, et al. A loss-of-function mutation in the CFC domain of TDGF1 is associated with human forebrain defects. *Human Genetics* 2002; 110: 422–8.

Cummings BB, Karczewski KJ, Kosmicki JA, Seaby EG, Watts NA, Singer-Berk M, et al. Transcript expression-aware annotation improves rare variant discovery and interpretation [Internet]. *Genomics*; 2019[cited 2020 May 17] Available from: <http://biorxiv.org/lookup/doi/10.1101/554444>

Cutting GR. Modifier genes in Mendelian disorders: the example of cystic fibrosis. *Ann N Y Acad Sci* 2010; 1214: 57–69.

Davis EE, Katsanis N. The ciliopathies: a transitional model into systems biology of human genetic disease. *Current Opinion in Genetics & Development* 2012; 22: 290–303.

Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. *PLoS Computational Biology* 2010; 6: e1001025.

De Meirleir L, Lissens W, Benelli C, Ponsot G, Desguerre I, Marsac C, et al. Aberrant Splicing of Exon 6 in the Pyruvate Dehydrogenase-E1 $\alpha$  mRNA Linked to a Silent Mutation in a Large Family with Leigh's Encephalomyelopathy. *Pediatric Research* 1994; 36: 707–12.

Dennis JF, Kurosaka H, Iulianella A, Pace J, Thomas N, Beckham S, et al. Mutations in Hedgehog Acyltransferase (Hhat) Perturb Hedgehog Signaling, Resulting in Severe Acrania-Holoprosencephaly-Agnathia Craniofacial Defects. *PLoS Genetics* 2012; 8: e1002927.

Desbordes SC. The glypican Dally-like is required for Hedgehog signalling in the embryonic epidermis of *Drosophila*. *Development* 2003; 130: 6245–55.

Desmet F-O, Hamroun D, Lalande M, Collod-Bérout G, Claustres M, Bérout C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Research* 2009; 37: e67–e67.

Dessaud E, McMahon AP, Briscoe J. Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. *Development* 2008; 135: 2489–503.

Diederichs S, Bartsch L, Berkmann JC, Fröse K, Heitmann J, Hoppe C, et al. The dark matter of the cancer genome: aberrations in regulatory elements, untranslated regions, splice sites, non-coding RNA and synonymous mutations. *EMBO Molecular Medicine* 2016; 8: 442–57.

Dipple KM, McCabe ERB. Phenotypes of Patients with "Simple" Mendelian Disorders Are Complex Traits: Thresholds, Modifiers, and Systems Dynamics. *The American Journal of Human Genetics* 2000; 66: 1729–35.

Drake N. What is the human genome worth? [Internet]. *Nature* 2011[cited 2020 May 12] Available from: <http://www.nature.com/articles/news.2011.281>

Dubourg C, Bendavid C, Pasquier L, Henry C, Odent S, David V. Holoprosencephaly [Internet]. *Orphanet Journal of Rare Diseases* 2007; 2[cited 2020 Jun 21] Available from: <https://ojrd.biomedcentral.com/articles/10.1186/1750-1172-2-8>

Dubourg C, Carré W, Hamdi-Rozé H, Mouden C, Roume J, Abdelmajid B, et al. Mutational Spectrum in Holoprosencephaly Shows That FGF is a New Major Signaling Pathway: HUMAN MUTATION. *Human Mutation* 2016; 37: 1329–39.

Dubourg C, Kim A, Watrin E, de Tayrac M, Odent S, David V, et al. Recent advances in understanding inheritance of holoprosencephaly. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 2018; 178: 258–69.

Dupe V, Rochard L, Mercier S, Le Petillon Y, Gicquel I, Bendavid C, et al. NOTCH, a new signaling pathway implicated in holoprosencephaly. *Human Molecular Genetics* 2011; 20: 1122–31.

Durmaz AA, Karaca E, Demkow U, Toruner G, Schoumans J, Cogulu O. Evolution of Genetic Techniques: Past, Present, and Beyond. *BioMed Research International* 2015; 2015: 1–7.

Echelard Y, Epstein DJ, St-Jacques B, Shen L, Mohler J, McMahon JA, et al. Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell* 1993; 75: 1417–30.

Eilbeck K, Quinlan A, Yandell M. Settling the score: variant prioritization and Mendelian disease. *Nature Reviews Genetics* 2017; 18: 599–612.

El-Jaick KB, Powers SE, Bartholin L, Myers KR, Hahn J, Orioli IM, et al. Functional analysis of mutations in TGIF associated with holoprosencephaly. *Molecular Genetics and Metabolism* 2007; 90: 97–111.

Emison ES, McCallion AS, Kashuk CS, Bush RT, Grice E, Lin S, et al. A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk. *Nature* 2005; 434: 857–63.

Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. [No title found]. *Genome Biol* 2003; 5: R1.

Erten S, Bebek G, Koyutürk M. Vavien: an algorithm for prioritizing candidate disease genes based on topological similarity of proteins in interaction networks. *J Comput Biol* 2011; 18: 1561–74.

Estibeiro JP, Brook FA, Copp AJ. Interaction between splotch (Sp) and curly tail (ct) mouse mutants in the embryonic development of neural tube defects. *Development* 1993; 119: 113–21.

Exome Aggregation Consortium, Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; 536: 285–91.

Faure G, Ogurtsov AY, Shabalina SA, Koonin EV. Role of mRNA structure in the control of protein folding. *Nucleic Acids Res* 2016; 44: 10898–911.

Faye-Petersen O, David E, Rangwala N, Seaman JP, Hua Z, Heller DS. OTOCEPHALY: REPORT OF FIVE NEW CASES AND A LITERATURE REVIEW. *Fetal and Pediatric Pathology* 2006; 25: 277–96.

Feiglin A, Allen BK, Kohane IS, Kong SW. Comprehensive Analysis of Tissue-wide Gene Expression and Phenotype Data Reveals Tissues Affected in Rare Genetic Disorders. *Cell Systems* 2017; 5: 140-148.e2.

Fernandes M, Gutin G, Alcorn H, McConnell SK, Hebert JM. Mutations in the BMP pathway in mice support the existence of two molecular classes of holoprosencephaly. *Development* 2007; 134: 3789–94.

Ferrari R, Kia DA, Tomkins JE, Hardy J, Wood NW, Lovering RC, et al. Stratification of candidate genes for Parkinson's disease using weighted protein-protein interaction network analysis [Internet]. *BMC Genomics* 2018; 19[cited 2020 May 17] Available from: <https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-018-4804-9>

Folstein SE, Rosen-Sheidley B. Genetics of autism: complex aetiology for a heterogeneous disorder. *Nature Reviews Genetics* 2001; 2: 943–55.

Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research* 2012; 41: D808–15.

Franco PG, Paganelli AR, López SL, Carrasco AE. Functional association of retinoic acid and hedgehog signaling in *Xenopus* primary neurogenesis. *Development* 1999; 126: 4257–65.

- Fu Y, Liu Z, Lou S, Bedford J, Mu XJ, Yip KY, et al. FunSeq2: a framework for prioritizing noncoding regulatory variants in cancer [Internet]. *Genome Biology* 2014; 15[cited 2020 May 17] Available from: <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0480-5>
- Furtado MB, Solloway MJ, Jones VJ, Costa MW, Biben C, Wolstein O, et al. BMP/SMAD1 signaling sets a threshold for the left/right pathway in lateral plate mesoderm and limits availability of SMAD4. *Genes & Development* 2008; 22: 3037–49.
- Fürthauer M, Reifers F, Brand M, Thisse B, Thisse C. sprouty4 acts in vivo as a feedback-induced antagonist of FGF signaling in zebrafish. *Development* 2001; 128: 2175–86.
- Furuta Y, Piston DW, Hogan BL. Bone morphogenetic proteins (BMPs) as regulators of dorsal forebrain development. *Development* 1997; 124: 2203–12.
- Gabriel SB, Salomon R, Pelet A, Angrist M, Amiel J, Fornage M, et al. Segregation at three loci explains familial and population risk in Hirschsprung disease. *Nature Genetics* 2002; 31: 89–93.
- Gaither JBS, Lammi GE, Li JL, Gordon DM, Kuck HC, Kelly BJ, et al. Global Analysis of Human mRNA Folding Disruptions in Synonymous Variants Demonstrates Significant Population Constraint [Internet]. *Genomics*; 2019[cited 2020 Jun 2] Available from: <http://biorxiv.org/lookup/doi/10.1101/712679>
- Gammill LS, Bronner-Fraser M. Neural crest specification: migrating into genomics. *Nature Reviews Neuroscience* 2003; 4: 795–805.
- Garber M, Guttman M, Clamp M, Zody MC, Friedman N, Xie X. Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* 2009; 25: i54–62.
- Garcia-Rodriguez E, Garcia-Garcia E, Perez-Sanchez A, Pavon-Delgado A. A NEW OBSERVATION OF 13q DELETION SYNDROME: SEVERE UNDESCRIBED FEATURES. *Genet Couns* 2015; 26: 213–7.
- Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing [Internet]. arXiv:12073907 [q-bio] 2012[cited 2020 May 25] Available from: <http://arxiv.org/abs/1207.3907>
- Gazal S, Génin E, Leutenegger A-L. Relationship inference from the genetic data on parents or offspring: A comparative study. *Theoretical Population Biology* 2016; 107: 31–8.
- Gazal S, Sahbatou M, Babron M-C, Génin E, Leutenegger A-L. FSuite: exploiting inbreeding in dense SNP chip and exome data. *Bioinformatics* 2014; 30: 1940–1.
- Geng X, Oliver G. Pathogenesis of holoprosencephaly. *J Clin Invest* 2009; 119: 1403–13.
- Gibson G. Rare and common variants: twenty arguments. *Nature Reviews Genetics* 2012; 13: 135–45.
- Gilissen C, Hoischen A, Brunner HG, Veltman JA. Unlocking Mendelian disease using exome sequencing. *Genome Biology* 2011; 12: 228.
- Girdea M, Dumitriu S, Fiume M, Bowdin S, Boycott KM, Chénier S, et al. PhenoTips: Patient Phenotyping Software for Clinical and Research Use. *Human Mutation* 2013; 34: 1057–65.
- Glusman G, Caballero J, Mauldin DE, Hood L, Roach JC. Kaviar: an accessible system for testing SNV novelty. *Bioinformatics* 2011; 27: 3216–7.
- Goetz JA. A Highly Conserved Amino-terminal Region of Sonic Hedgehog Is Required for the Formation of Its Freely Diffusible Multimeric Form. *Journal of Biological Chemistry* 2006; 281: 4087–93.
- Goruppi S, Procopio M-G, Jo S, Clocchiatti A, Neel V, Dotto GP. The ULK3 Kinase Is Critical for Convergent Control of Cancer-Associated Fibroblast Activation by CSL and GLI. *Cell Reports* 2017; 20: 2468–79.
- Gouya L, Puy H, Robreau A-M, Bourgeois M, Lamoril J, Da Silva V, et al. The penetrance of dominant erythropoietic protoporphyria is modulated by expression of wildtype FECH. *Nature Genetics* 2002; 30: 27–8.
- Greater Middle East Variome Consortium, Scott EM, Halees A, Itan Y, Spencer EG, He Y, et al. Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nature Genetics* 2016; 48: 1071–6.
- Greene NDE, Copp AJ. Neural Tube Defects. *Annual Review of Neuroscience* 2014; 37: 221–42.

Gripp KW, Wotton D, Edwards MC, Roessler E, Ades L, Meinecke P, et al. Mutations in TGIF cause holoprosencephaly and link NODAL signalling to human neural axis determination. *Nature Genetics* 2000; 25: 205–8.

GROUP Investigators, Rees E, Han J, Morgan J, Carrera N, Escott-Price V, et al. De novo mutations identified by exome sequencing implicate rare missense variants in SLC6A1 in schizophrenia. *Nature Neuroscience* 2020; 23: 179–84.

GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013; 45: 580–5.

Gu D, Xie J. Non-Canonical Hh Signaling in Cancer—Current Understanding and Future Directions. *Cancers* 2015; 7: 1684–98.

Guo MH, Plummer L, Chan Y-M, Hirschhorn JN, Lippincott MF. Burden Testing of Rare Variants Identified through Exome Sequencing via Publicly Available Control Data. *The American Journal of Human Genetics* 2018; 103: 522–34.

Gupta S, Chatterjee S, Mukherjee A, Mutsuddi M. Whole exome sequencing: Uncovering causal genetic variants for ocular diseases. *Experimental Eye Research* 2017; 164: 139–50.

Gurrieri F, Trask BJ, van den Engh G, Krauss CM, Schinzel A, Pettenati MJ, et al. Physical mapping of the holoprosencephaly critical region on chromosome 7q36. *Nature Genetics* 1993; 3: 247–51.

Gusella JF, MacDonald ME, Lee J-M. Genetic modifiers of Huntington's disease: GENETIC MODIFIERS OF HD. *Movement Disorders* 2014; 29: 1359–65.

Gutierrez J, Sepulveda W, Saez R, Carstens E, Sanchez J. Prenatal diagnosis of 13q- syndrome in a fetus with holoprosencephaly and thumb agenesis: 13q- syndrome. *Ultrasound in Obstetrics and Gynecology* 2001; 17: 166–8.

Gutin G. FGF signalling generates ventral telencephalic cells independently of SHH. *Development* 2006; 133: 2937–46.

Hageman GS, Anderson DH, Johnson LV, Hancox LS, Taiber AJ, Hardisty LI, et al. From The Cover: A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proceedings of the National Academy of Sciences* 2005; 102: 7227–32.

Hahn JS, Barnes PD, Clegg NJ, Stashinko EE. Septopreoptic Holoprosencephaly: A Mild Subtype Associated with Midline Craniofacial Anomalies. *American Journal of Neuroradiology* 2010; 31: 1596–601.

Hall ET, Cleverdon ER, Ogden SK. Dispatching Sonic Hedgehog: Molecular Mechanisms Controlling Deployment. *Trends in Cell Biology* 2019; 29: 385–95.

Hall TMT, Porter JA, Young KE, Koonin EV, Beachy PA, Leahy DJ. Crystal Structure of a Hedgehog Autoprocessing Domain: Homology between Hedgehog and Self-Splicing Proteins. *Cell* 1997; 91: 85–97.

Hamdi-Rozé H, Ware M, Guyodo H, Rizzo A, Ratié L, Rupin M, et al. Disrupted hypothalamo-pituitary axis in association with reduced SHH underlies the pathogenesis of NOTCH-deficiency. *J Clin Endocrinol Metab* 2020

Hamosh A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Research* 2002; 30: 52–5.

Harris MJ, Juriloff DM. An update to the list of mouse mutants with neural tube closure defects and advances toward a complete genetic perspective of neural tube closure. *Birth Defects Res Part A Clin Mol Teratol* 2010; 88: 653–69.

Harrison-Uy SJ, Pleasure SJ. Wnt Signaling and Forebrain Development. *Cold Spring Harbor Perspectives in Biology* 2012; 4: a008094–a008094.

Havrilla JM, Pedersen BS, Layer RM, Quinlan AR. A map of constrained coding regions in the human genome. *Nature Genetics* 2019; 51: 88–95.

Hayhurst M, McConnell SK. Mouse models of holoprosencephaly: Current Opinion in Neurology 2003; 16: 135–41.

Hehr U, Pineda-Alvarez DE, Uyanik G, Hu P, Zhou N, Hehr A, et al. Heterozygous mutations in SIX3 and SHH are associated with schizencephaly and further expand the clinical spectrum of holoprosencephaly. *Human Genetics* 2010; 127: 555–61.

- Helms JA. New insights into craniofacial morphogenesis. *Development* 2005; 132: 851–61.
- Henrique D, Hirsinger E, Adam J, Roux IL, Pourquié O, Ish-Horowicz D, et al. Maintenance of neuroepithelial progenitor cells by Delta–Notch signalling in the embryonic chick retina. *Current Biology* 1997; 7: 661–70.
- Heyer J, Escalante-Alcalde D, Lia M, Boettinger E, Edelmann W, Stewart CL, et al. Postgastrulation Smad2-deficient embryos show defects in embryo turning and anterior morphogenesis. *Proceedings of the National Academy of Sciences* 1999; 96: 12595–600.
- van Heyningen V. Mechanisms of non-Mendelian inheritance in genetic disease. *Human Molecular Genetics* 2004; 13: R225–33.
- Hockenberry AJ, Simer MI, Amaral LAN, Jewett MC. Quantifying Position-Dependent Codon Usage Bias. *Molecular Biology and Evolution* 2014; 31: 1880–93.
- Hong M, Krauss RS. Rescue of holoprosencephaly in fetal alcohol-exposed Cdon mutant mice by reduced gene dosage of Ptch1. *PLoS ONE* 2013; 8: e79269.
- Hong M, Krauss RS. Ethanol itself is a holoprosencephaly-inducing teratogen. *PLoS ONE* 2017; 12: e0176440.
- Hong M, Krauss RS. Modeling the complex etiology of holoprosencephaly in mice. *Am J Med Genet C Semin Med Genet* 2018; 178: 140–50.
- Hong M, Srivastava K, Kim S, Allen BL, Leahy DJ, Hu P, et al. *BOC* is a modifier gene in holoprosencephaly. *Human Mutation* 2017; 38: 1464–70.
- Hoodless PA, Pye M, Chazaud C, Labbé E, Attisano L, Rossant J, et al. FoxH1 (Fast) functions to specify the anterior primitive streak in the mouse. *Genes Dev* 2001; 15: 1257–71.
- Hu D. A zone of frontonasal ectoderm regulates patterning and growth in the face. *Development* 2003; 130: 1749–58.
- Hu D, Marcucio RS. A SHH-responsive signaling center in the forebrain regulates craniofacial morphogenesis via the facial ectoderm. *Development* 2009; 136: 107–16.
- Hu H, Huff CD, Moore B, Flygare S, Reese MG, Yandell M. VAAST 2.0: improved variant classification and disease-gene identification using a conservation-controlled amino acid substitution matrix. *Genet Epidemiol* 2013; 37: 622–34.
- Hu T, Kruszka P, Martinez AF, Ming JE, Shabason EK, Raam MS, et al. Cytogenetics and holoprosencephaly: A chromosomal microarray study of 222 individuals with holoprosencephaly. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 2018; 178: 175–86.
- Huang H-Y, Chien C-H, Jen K-H, Huang H-D. RegRNA: an integrated web server for identifying regulatory RNA motifs and elements. *Nucleic Acids Research* 2006; 34: W429–34.
- Hughes JJ, Alkhunaizi E, Kruszka P, Pyle LC, Grange DK, Berger SI, et al. Loss-of-Function Variants in PPP1R12A: From Isolated Sex Reversal to Holoprosencephaly Spectrum and Urogenital Malformations. *The American Journal of Human Genetics* 2020; 106: 121–8.
- Hunt RC, Simhadri VL, Iandoli M, Sauna ZE, Kimchi-Sarfaty C. Exposing synonymous mutations. *Trends in Genetics* 2014; 30: 308–21.
- Hwang J-I, Oh Y-S, Shin K-J, Kim H, Ryu SH, Suh P-G. Molecular cloning and characterization of a novel phospholipase C, PLC-eta. *Biochem J* 2005; 389: 181–6.
- Ideker T, Sharan R. Protein networks in disease. *Genome Research* 2008; 18: 644–52.
- Inatani M. Mammalian Brain Morphogenesis and Midline Axon Guidance Require Heparan Sulfate. *Science* 2003; 302: 1044–6.
- Izzi L, Lévesque M, Morin S, Laniel D, Wilkes BC, Mille F, et al. Boc and Gas1 Each Form Distinct Shh Receptor Complexes with Ptch1 and Are Required for Shh-Mediated Cell Proliferation. *Developmental Cell* 2011; 20: 788–801.
- Jackson M, Marks L, May GHW, Wilson JB. The genetic basis of disease. *Essays in Biochemistry* 2018; 62: 643–723.
- Jacob J, Briscoe J. Gli proteins and the control of spinal-cord patterning. *EMBO reports* 2003; 4: 761–5.

- Jakobs P, Schulz P, Ortmann C, Schürmann S, Exner S, Rebolledo-Rios R, et al. Bridging the gap: heparan sulfate and Scube2 assemble Sonic hedgehog release complexes at the surface of producing cells. *Sci Rep* 2016; 6: 26435.
- Jansen R, Bussemaker HJ, Gerstein M. Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. *Nucleic Acids Res* 2003; 31: 2242–51.
- Jeong Y. A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. *Development* 2006; 133: 761–72.
- Jeong Y, Leskow FC, El-Jaick K, Roessler E, Muenke M, Yocum A, et al. Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nature Genetics* 2008; 40: 1348–53.
- Jeong Y, Leskow FC, El-Jaick K, Roessler E, Muenke M, Yocum A, et al. Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nature Genetics* 2008; 40: 1348–53.
- Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Research* 2014; 42: 13534–44.
- Jiang Y, Wu C, Zhang Y, Zhang S, Yu S, Lei P, et al. GTX.Digest.VCF: an online NGS data interpretation system based on intelligent gene ranking and large-scale text mining [Internet]. *BMC Medical Genomics* 2019; 12[cited 2020 May 17] Available from: <https://bmcmedgenomics.biomedcentral.com/articles/10.1186/s12920-019-0637-x>
- Jinks RN, Puffenberger EG, Baple E, Harding B, Crino P, Fogo AB, et al. Recessive nephrocerebellar syndrome on the Galloway-Mowat syndrome spectrum is caused by homozygous protein-truncating mutations of *WDR73*. *Brain* 2015; 138: 2173–90.
- Johnson CY, Rasmussen SA. Non-genetic risk factors for holoprosencephaly. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 2010; 154C: 73–85.
- Jones CM, Kuehn MR, Hogan BL, Smith JC, Wright CV. Nodal-related signals induce axial mesoderm and dorsalize mesoderm during gastrulation. *Development* 1995; 121: 3651–62.
- Jordan DM, Frangakis SG, Golzio C, Cassa CA, Kurtzberg J, Task Force for Neonatal Genomics, et al. Identification of cis-suppression of human disease mutations by comparative genomics. *Nature* 2015; 524: 225–9.
- Júnior EA, Filho HAG, Pires CR, Filho SMZ. Prenatal diagnosis of the 13q-syndrome through three-dimensional ultrasonography: a case report. *Archives of Gynecology and Obstetrics* 2006; 274: 243–5.
- Kagan KO, Staboulidou I, Syngelaki A, Cruz J, Nicolaidis KH. The 11-13-week scan: diagnosis and outcome of holoprosencephaly, exomphalos and megacystis. *Ultrasound in Obstetrics and Gynecology* 2010; 36: 10–4.
- Kaiser J. Affordable ‘Exomes’ Fill Gaps in a Catalog of Rare Diseases. *Science* 2010; 330: 903–903.
- Kakar N, Ahmad J, Morris-Rosendahl DJ, Altmüller J, Friedrich K, Barbi G, et al. STIL mutation causes autosomal recessive microcephalic lobar holoprosencephaly. *Human Genetics* 2015; 134: 45–51.
- Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans [Internet]. *Genomics*; 2019[cited 2020 May 15] Available from: <http://biorxiv.org/lookup/doi/10.1101/531210>
- Karki R, Pandya D, Elston RC, Ferlini C. Defining ‘mutation’ and ‘polymorphism’ in the era of personal genomics. *BMC Med Genomics* 2015; 8: 37.
- Katsanis N. Triallelic Inheritance in Bardet-Biedl Syndrome, a Mendelian Recessive Disorder. *Science* 2001; 293: 2256–9.
- Katsanis N. The continuum of causality in human genetic disorders [Internet]. *Genome Biology* 2016; 17[cited 2020 May 6] Available from: <http://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-1107-9>
- Kauvar EF, Muenke M. Holoprosencephaly: recommendations for diagnosis and management: *Current Opinion in Pediatrics* 2010; 22: 687–95.
- Keller MC, Visscher PM, Goddard ME. Quantification of Inbreeding Due to Distant Ancestors and Its Detection Using Dense Single Nucleotide Polymorphism Data. *Genetics* 2011; 189: 237–49.

Kennedy MA. Mendelian Genetic Disorders [Internet]. In: John Wiley & Sons, Ltd, editor(s). Encyclopedia of Life Sciences. Chichester, UK: John Wiley & Sons, Ltd; 2005[cited 2020 May 8] Available from: <http://doi.wiley.com/10.1038/npg.els.0003934>

Kim MJ, Kim E, Ryu SH, Suh P-G. The mechanism of phospholipase C- $\gamma$ 1 regulation. *Experimental & Molecular Medicine* 2000; 32: 101–9.

Kimchi-Sarfaty C, Oh JM, Kim I-W, Sauna ZE, Calcagno AM, Ambudkar SV, et al. A ‘Silent’ Polymorphism in the MDR1 Gene Changes Substrate Specificity. *Science* 2007; 315: 525–8.

Kimura M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* 1977; 267: 275–6.

Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics* 2014; 46: 310–5.

Knight HM, Maclean A, Irfan M, Naeem F, Cass S, Pickard BS, et al. Homozygosity mapping in a family presenting with schizophrenia, epilepsy and hearing impairment. *European Journal of Human Genetics* 2008; 16: 750–8.

Kobayashi Y, Yang S, Nykamp K, Garcia J, Lincoln SE, Topper SE. Pathogenic variant burden in the ExAC database: an empirical approach to evaluating population data for clinical variant interpretation [Internet]. *Genome Medicine* 2017; 9[cited 2020 May 16] Available from: <http://genomemedicine.biomedcentral.com/articles/10.1186/s13073-017-0403-7>

Köhler S, Schulz MH, Krawitz P, Bauer S, Dölken S, Ott CE, et al. Clinical Diagnostics in Human Genetics with Semantic Similarity Searches in Ontologies. *The American Journal of Human Genetics* 2009; 85: 457–64.

Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, et al. The Human Phenotype Ontology in 2017. *Nucleic Acids Res* 2017; 45: D865–76.

Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, et al. The Human Phenotype Ontology in 2017. *Nucleic Acids Research* 2017; 45: D865–76.

Kong JH, Yang L, Dessaud E, Chuang K, Moore DM, Rohatgi R, et al. Notch Activity Modulates the Responsiveness of Neural Progenitors to Sonic Hedgehog Signaling. *Developmental Cell* 2015; 33: 373–87.

Kool M, Jones DTW, Jäger N, Northcott PA, Pugh TJ, Hovestadt V, et al. Genome Sequencing of SHH Medulloblastoma Predicts Genotype-Related Response to Smoothed Inhibition. *Cancer Cell* 2014; 25: 393–405.

Kousi M, Katsanis N. Genetic Modifiers and Oligogenic Inheritance. *Cold Spring Harbor Perspectives in Medicine* 2015; 5: a017145–a017145.

Krauss RS. Holoprosencephaly: new models, new insights. *Expert Reviews in Molecular Medicine* 2007; 9: 1–17.

Krauss S, Concordet J-P, Ingham PW. A functionally conserved homolog of the Drosophila segment polarity gene hh is expressed in tissues with polarizing activity in zebrafish embryos. *Cell* 1993; 75: 1431–44.

Kruszka P, Berger SI, Casa V, Dekker MR, Gaesser J, Weiss K, et al. Cohesin complex-associated holoprosencephaly. *Brain* 2019; 142: 2631–43.

Kruszka P, Martinez AF, Muenke M. Molecular testing in holoprosencephaly. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 2018; 178: 187–93.

Kruszka P, Muenke M. Syndromes associated with holoprosencephaly. *Am J Med Genet C Semin Med Genet* 2018; 178: 229–37.

Kryukov GV, Pennacchio LA, Sunyaev SR. Most Rare Missense Alleles Are Deleterious in Humans: Implications for Complex Disease and Association Studies. *The American Journal of Human Genetics* 2007; 80: 727–39.

Kubota Y, Ito K. Chemotactic migration of mesencephalic neural crest cells in the mouse. *Developmental Dynamics* 2000; 217: 170–9.

Kulski JK. Next-Generation Sequencing — An Overview of the History, Tools, and “Omic” Applications [Internet]. In: Kulski JK, editor(s). Next Generation Sequencing - Advances, Applications and Challenges. InTech; 2016[cited 2020 Jun 18] Available from: <http://www.intechopen.com/books/next-generation-sequencing-advances-applications-and-challenges/next-generation-sequencing-an-overview-of-the-history-tools-and-omic-applications>

- Kumar D. Disorders of the genome architecture: a review. *Genomic Medicine* 2008; 2: 69–76.
- Lacbawan F, Solomon BD, Roessler E, El-Jaick K, Domené S, Vélez JI, et al. Clinical spectrum of SIX3-associated mutations in holoprosencephaly: correlation between genotype, phenotype and function. *J Med Genet* 2009; 46: 389–98.
- Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Research* 2016; 44: D862–8.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis [Internet]. *BMC Bioinformatics* 2008; 9[cited 2020 May 25] Available from: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559>
- Lappé M, Lau L, Dudovitz RN, Nelson BB, Karp EA, Kuo AA. The Diagnostic Odyssey of Autism Spectrum Disorder. *Pediatrics* 2018; 141: S272–9.
- Lee J, Ekker S, von Kessler D, Porter J, Sun B, Beachy P. Autoproteolysis in hedgehog protein biogenesis. *Science* 1994; 266: 1528–37.
- Leitch CC, Zaghloul NA, Davis EE, Stoetzel C, Diaz-Font A, Rix S, et al. Hypomorphic mutations in syndromic encephalocele genes are associated with Bardet-Biedl syndrome. *Nature Genetics* 2008; 40: 443–8.
- Leman R, Gaildrat P, Le Gac G, Ka C, Fichou Y, Audrezet M-P, et al. Novel diagnostic tool for prediction of variant spliceogenicity derived from a set of 395 combined in silico/in vitro studies: an international collaborative effort. *Nucleic Acids Research* 2018; 46: 7913–23.
- Lemmers RJLF, Tawil R, Petek LM, Balog J, Block GJ, Santen GWE, et al. Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2. *Nature Genetics* 2012; 44: 1370–4.
- Leoncini E, Baranello G, Orioli IM, Annerén G, Bakker M, Bianchi F, et al. Frequency of holoprosencephaly in the International Clearinghouse Birth Defects Surveillance Systems: Searching for population variations. *Birth Defects Research Part A: Clinical and Molecular Teratology* 2008; 82: 585–91.
- Leonenko G, Richards AL, Walters JT, Pocklington A, Chambert K, Al Eissa MM, et al. Mutation intolerant genes and targets of FMRP are enriched for nonsynonymous alleles in schizophrenia. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 2017; 174: 724–31.
- Leutenegger A-L, Prum B, Génin E, Verny C, Lemainque A, Clerget-Darpoux F, et al. Estimation of the Inbreeding Coefficient through Use of Genomic Data. *The American Journal of Human Genetics* 2003; 73: 516–23.
- Levey EB, Stashinko E, Clegg NJ, Delgado MR. Management of children with holoprosencephaly. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 2010; 154C: 183–90.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; 25: 1754–60.
- Li Y, Lai-Han Leung E, Pan H, Yao X, Huang Q, Wu M, et al. Identification of potential genetic causal variants for rheumatoid arthritis by whole-exome sequencing. *Oncotarget* 2017; 8: 111119–29.
- Li Z, Ivanov AA, Su R, Gonzalez-Pecchi V, Qi Q, Liu S, et al. The OncoPPi network of cancer-focused protein–protein interactions to inform biological insights and therapeutic strategies [Internet]. *Nature Communications* 2017; 8[cited 2020 May 17] Available from: <http://www.nature.com/articles/ncomms14356>
- Lightbody G, Haberland V, Browne F, Taggart L, Zheng H, Parkes E, et al. Review of applications of high-throughput sequencing in personalized medicine: barriers and facilitators of future progress in research and clinical application. *Briefings in Bioinformatics* 2019; 20: 1795–811.
- Lin AE, Siebert JR, Graham JM. Central nervous system malformations in the CHARGE association. *American Journal of Medical Genetics* 1990; 37: 304–10.
- Lin C, Yao E, Wang K, Nozawa Y, Shimizu H, Johnson JR, et al. Regulation of Sufu activity by p66 $\beta$  and Mycbbp provides new insight into vertebrate Hedgehog signaling. *Genes & Development* 2014; 28: 2547–63.
- Lin X, Buff EM, Perrimon N, Michelson AM. Heparan sulfate proteoglycans are essential for FGF receptor signaling during *Drosophila* embryonic development. *Development* 1999; 126: 3715–23.
- Lindsay SJ, Xu Y, Lisgo SN, Harkin LF, Copp AJ, Gerrelli D, et al. HDBR Expression: A Unique Resource for Global and Individual Gene Expression Studies during Early Human Brain Development [Internet]. *Frontiers in*

Neuroanatomy 2016; 10[cited 2020 May 25] Available from:  
<http://journal.frontiersin.org/article/10.3389/fnana.2016.00086/full>

Lipinski RJ, Godin EA, O'leary-Moore SK, Parnell SE, Sulik KK. Genesis of teratogen-induced holoprosencephaly in mice. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 2010; 154C: 29–42.

Liu W, Qian C, Francke U. Silent mutation induces exon skipping of fibrillin-1 gene in Marfan syndrome. *Nature Genetics* 1997; 16: 328–9.

Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Human Mutation* 2016; 37: 235–41.

Livingstone M, Folkman L, Yang Y, Zhang P, Mort M, Cooper DN, et al. Investigating DNA-, RNA-, and protein-based features as a means to discriminate pathogenic synonymous variants: LIVINGSTONE et al. *Human Mutation* 2017; 38: 1336–47.

Lochovska K, Peterkova R, Pavlikova Z, Hovorakova M. Sprouty gene dosage influences temporal-spatial dynamics of primary enamel knot formation [Internet]. *BMC Developmental Biology* 2015; 15[cited 2020 May 26] Available from: <http://www.biomedcentral.com/1471-213X/15/21>

Lorenz R, Bernhart SH, Höner zu Siederdisen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0 [Internet]. *Algorithms for Molecular Biology* 2011; 6[cited 2020 May 30] Available from:  
<https://almob.biomedcentral.com/articles/10.1186/1748-7188-6-26>

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2 [Internet]. *Genome Biology* 2014; 15[cited 2020 May 25] Available from:  
<http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8>

Lu CC, Robertson EJ. Multiple roles for Nodal in the epiblast of the mouse embryo in the establishment of anterior-posterior patterning. *Developmental Biology* 2004; 273: 149–59.

Lubbe SJ, Escott-Price V, Gibbs JR, Nalls MA, Bras J, Price TR, et al. Additional rare variant analysis in Parkinson's disease cases with and without known pathogenic mutations: evidence for oligogenic inheritance. *Human Molecular Genetics* 2016: ddw348.

Ma Y, Zhang P, Wang F, Yang J, Yang Z, Qin H. The relationship between early embryo development and tumorigenesis. *Journal of Cellular and Molecular Medicine* 2010; 14: 2697–701.

MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature* 2014; 508: 469–76.

Maller J, George S, Purcell S, Fagerness J, Altshuler D, Daly MJ, et al. Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nature Genetics* 2006; 38: 1055–9.

Mann RK, Beachy PA. Novel lipid modifications of secreted protein signals. *Annu Rev Biochem* 2004; 73: 891–923.

Marcucio RS, Cordero DR, Hu D, Helms JA. Molecular interactions coordinating the development of the forebrain and face. *Developmental Biology* 2005; 284: 48–61.

Marcucio RS, Young NM, Hu D, Hallgrímsson B. Mechanisms that underlie co-variation of the brain and face. *genesis* 2011; 49: 177–89.

Martin AR, Williams E, Foulger RE, Leigh S, Daugherty LC, Niblock O, et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nature Genetics* 2019; 51: 1560–5.

Martinez AF, Kruszka PS, Muenke M. Extracerebral manifestations of nonchromosomal, nonsyndromic holoprosencephaly. *Am J Med Genet C Semin Med Genet* 2018; 178: 246–57.

Martnez-Fras ML, Bermejo E, Rodriguez-Pinilla E, Prieto L, Fras JL. Epidemiological analysis of outcomes of pregnancy in gestational diabetic mothers. *American Journal of Medical Genetics* 1998; 78: 140–5.

Mashimo T, Voigt B, Kuramoto T, Serikawa T. Rat Phenome Project: The untapped potential of existing rat strains. *Journal of Applied Physiology* 2005; 98: 371–9.

Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology* 1999; 288: 911–40.

- Mathews DH, Turner DH. Dynalign: an algorithm for finding the secondary structure common to two RNA sequences. *Journal of Molecular Biology* 2002; 317: 191–203.
- Matsunaga E, Shiota K. Holoprosencephaly in human embryos: Epidemiologic studies of 150 cases. *Teratology* 1977; 16: 261–72.
- Maxam AM, Gilbert W. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences* 1977; 74: 560–4.
- McCabe MJ, Gaston-Massuet C, Tziaferi V, Gregory LC, Alatzoglou KS, Signore M, et al. Novel *FGF8* Mutations Associated with Recessive Holoprosencephaly, Craniofacial Defects, and Hypothalamo-Pituitary Dysfunction. *The Journal of Clinical Endocrinology & Metabolism* 2011; 96: E1709–18.
- McCarroll SA, Huett A, Kuballa P, Chilewski SD, Landry A, Goyette P, et al. Deletion polymorphism upstream of *IRGM* associated with altered *IRGM* expression and Crohn's disease. *Nature Genetics* 2008; 40: 1107–12.
- McCarthy MI, Smedley D, Hide W. [No title found]. *Genome Biol* 2003; 4: 119.
- McDonald JH, Kreitman M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 1991; 351: 652–4.
- McKusick VA. Mendelian Inheritance in Man and Its Online Version, OMIM. *The American Journal of Human Genetics* 2007; 80: 588–604.
- Melbourne Genomics Health Alliance, Stark Z, Dashnow H, Lunke S, Tan TY, Yeung A, et al. A clinically driven variant prioritization framework outperforms purely computational approaches for the diagnostic analysis of singleton WES data. *European Journal of Human Genetics* 2017; 25: 1268–72.
- Mercier S, Dubourg C, Garcelon N, Campillo-Gimenez B, Gicquel I, Belleguic M, et al. New findings for phenotype-genotype correlations in a large European series of holoprosencephaly cases. *Journal of Medical Genetics* 2011; 48: 752–60.
- Meyers EN, Lewandoski M, Martin GR. An *Fgf8* mutant allelic series generated by Cre- and Flp-mediated recombination. *Nature Genetics* 1998; 18: 136–41.
- Miller EA, Rasmussen SA, Siega-Riz AM, Frías JL, Honein MA, the National Birth Defects Prevention Study. Risk factors for non-syndromic holoprosencephaly in the National Birth Defects Prevention Study. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 2010; 154C: 62–72.
- Ming JE, Kaupas ME, Roessler E, Brunner HG, Golabi M, Tekin M, et al. Mutations in *PATCHED-1*, the receptor for *SONIC HEDGEHOG*, are associated with holoprosencephaly. *Human Genetics* 2002; 110: 297–301.
- Miyake K, Hirasawa T, Koide T, Kubota T. Epigenetics in Autism and Other Neurodevelopmental Diseases. In: Ahmad SI, editor(s). *Neurodegenerative Diseases*. New York, NY: Springer US; 2012. p. 91–8
- Mouden C, Dubourg C, Carré W, Rose S, Quelin C, Akloul L, et al. Complex mode of inheritance in holoprosencephaly revealed by whole exome sequencing: Complex mode of inheritance in holoprosencephaly. *Clinical Genetics* 2016; 89: 659–68.
- Mouden C, de Tayrac M, Dubourg C, Rose S, Carré W, Hamdi-Rozé H, et al. Homozygous *STIL* mutation causes holoprosencephaly and microcephaly in two siblings. *PLoS ONE* 2015; 10: e0117418.
- Muenke M. Genetics of ventral forebrain development and holoprosencephaly. *Current Opinion in Genetics & Development* 2000; 10: 262–9.
- Muenke M, Gurrieri F, Bay C, Yi DH, Collins AL, Johnson VP, et al. Linkage of a human brain malformation, familial holoprosencephaly, to chromosome 7 and evidence for genetic heterogeneity. *Proceedings of the National Academy of Sciences* 1994; 91: 8102–6.
- Muir P, Li S, Lou S, Wang D, Spakowicz DJ, Salichos L, et al. The real cost of sequencing: scaling computation to keep pace with data generation. *Genome Biol* 2016; 17: 53.
- Müller F, Albert S, Blader P, Fischer N, Hallonet M, Strähle U. Direct action of the nodal-related signal cyclops in induction of sonic hedgehog in the ventral midline of the CNS. *Development* 2000; 127: 3889–97.
- Murdoch JN, Copp AJ. The relationship between sonic Hedgehog signaling, cilia, and neural tube defects. *Birth Defects Research Part A: Clinical and Molecular Teratology* 2010; 88: 633–52.
- Murone M, Rosenthal A, de Sauvage FJ. Hedgehog Signal Transduction: From Flies to Vertebrates. *Experimental Cell Research* 1999; 253: 25–33.

Nadeau JH. Modifier genes in mice and humans. *Nature Reviews Genetics* 2001; 2: 165–74.

Nagai T, Aruga J, Minowa O, Sugimoto T, Ohno Y, Noda T, et al. *Zic2* regulates the kinetics of neurulation. *Proceedings of the National Academy of Sciences* 2000; 97: 1618–23.

Nakamura Y. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Research* 2000; 28: 292–292.

Nanni L, Ming JE, Bocian M, Steinhaus K, Bianchi DW, de Die-Smulders C, et al. The Mutational Spectrum of the Sonic Hedgehog Gene in Holoprosencephaly: SHH Mutations Cause a Significant Proportion of Autosomal Dominant Holoprosencephaly. *Human Molecular Genetics* 1999; 8: 2479–88.

Nava C, Rupp J, Boissel J-P, Mignot C, Rastetter A, Amiet C, et al. Hypomorphic variants of cationic amino acid transporter 3 in males with autism spectrum disorders. *Amino Acids* 2015; 47: 2647–58.

Ng PC, Henikoff S. Predicting Deleterious Amino Acid Substitutions. *Genome Research* 2001; 11: 863–74.

Nomura M, Li E. *Smad2* role in mesoderm formation, left–right patterning and craniofacial development. *Nature* 1998; 393: 786–90.

Nüsslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in *Drosophila*. *Nature* 1980; 287: 795–801.

Odent S, Le Marec B, Munnich A, Le Merrer M, Bonati-Pelli C. Segregation analysis in nonsyndromic holoprosencephaly. *American Journal of Medical Genetics* 1998; 77: 139–43.

O’Leary NA, Wright MW, Brister JR, Ciuffo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* 2016; 44: D733–45.

Oliver G, Mailhos A, Wehr R, Copeland NG, Jenkins NA, Gruss P. *Six3*, a murine homologue of the *sine oculis* gene, demarcates the most anterior border of the developing neural plate and is expressed during eye development. *Development* 1995; 121: 4045–55.

Olsen CL, Hughes JP, Youngblood LG, Sharpe-Stimac M. Epidemiology of holoprosencephaly and phenotypic characteristics of affected children: New York State, 1984-1989. *Am J Med Genet* 1997; 73: 217–26.

Ornitz DM, Itoh N. The Fibroblast Growth Factor signaling pathway. *Wiley Interdisciplinary Reviews: Developmental Biology* 2015; 4: 215–66.

O’Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, et al. Multiplex Targeted Sequencing Identifies Recurrently Mutated Genes in Autism Spectrum Disorders. *Science* 2012; 338: 1619–22.

Oti M. Predicting disease genes using protein-protein interactions. *Journal of Medical Genetics* 2006; 43: 691–8.

Ozdemir ES, Jang H, Gursoy A, Keskin O, Li Z, Sacks DB, et al. Unraveling the molecular mechanism of interactions of the Rho GTPases *Cdc42* and *Rac1* with the scaffolding protein *IQGAP2*. *Journal of Biological Chemistry* 2018; 293: 3685–99.

Padmanabhan R, Shafiullah M. Effect of maternal diabetes and ethanol interactions on embryo development in the mouse. *Molecular and Cellular Biochemistry* 2004; 261: 43–56.

Pagel KA, Kim R, Moad K, Busby B, Zheng L, Tokheim C, et al. Integrated Informatics Analysis of Cancer-Related Variants. *JCO Clinical Cancer Informatics* 2020: 310–7.

Paine-Saunders S, Viviano BL, Zupicich J, Skarnes WC, Saunders S. *glypican-3* Controls Cellular Responses to *Bmp4* in Limb Patterning and Skeletal Development. *Developmental Biology* 2000; 225: 179–87.

Pak E, Segal RA. Hedgehog Signal Transduction: Key Players, Oncogenic Drivers, and Cancer Therapy. *Developmental Cell* 2016; 38: 333–44.

Palazzo AF, Gregory TR. The Case for Junk DNA. *PLoS Genetics* 2014; 10: e1004351.

Pan Y, Woodbury A, Esko JD, Grobe K, Zhang X. Heparan sulfate biosynthetic gene *Ndst1* is required for FGF signaling in early lens development. *Development* 2006; 133: 4933–44.

Papadimitriou S, Gazzo A, Versbraegen N, Nachtegael C, Aerts J, Moreau Y, et al. Predicting disease-causing variant combinations. *Proceedings of the National Academy of Sciences* 2019: 201815601.

Parikshak NN, Swarup V, Belgard TG, Irimia M, Ramaswami G, Gandal MJ, et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* 2016; 540: 423–7.

Pechmann S, Frydman J. Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding. *Nature Structural & Molecular Biology* 2013; 20: 237–43.

Peebles DM. Holoprosencephaly. *Prenat Diagn* 1998; 18: 477–80.

Perea-Gomez A, Vella FDJ, Shawlot W, Oulad-Abdelghani M, Chazaud C, Meno C, et al. Nodal Antagonists in the Anterior Visceral Endoderm Prevent the Formation of Multiple Primitive Streaks. *Developmental Cell* 2002; 3: 745–56.

Perlea M. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Research* 2001; 29: 1185–90.

Petracchi F, Crespo L, Michia C, Igarzabal L, Gadow E. Holoprosencephaly at prenatal diagnosis: analysis of 28 cases regarding etiopathogenic diagnoses. *Prenatal Diagnosis* 2011: n/a-n/a.

Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic Intolerance to Functional Variation and the Interpretation of Personal Genomes. *PLoS Genetics* 2013; 9: e1003709.

Petryk A, Graf D, Marcucio R. Holoprosencephaly: signaling interactions between the brain and the face, the environment and the genes, and the phenotypic variability in animal models and humans: Pathogenesis of holoprosencephaly. *Wiley Interdisciplinary Reviews: Developmental Biology* 2015; 4: 17–32.

Phelps IG, Dempsey JC, Grout ME, Isabella CR, Tully HM, Doherty D, et al. Interpreting the clinical significance of combined variants in multiple recessive disease genes: systematic investigation of Joubert syndrome yields little support for oligogenicity. *Genetics in Medicine* 2018; 20: 223–33.

Pickles A, Bolton P, Macdonald H, Bailey A, Le Couteur A, Sim CH, et al. Latent-class analysis of recurrence risks for complex phenotypes with selection and measurement error: a twin and family history study of autism. *Am J Hum Genet* 1995; 57: 717–26.

Pineda-Alvarez DE, Roessler E, Hu P, Srivastava K, Solomon BD, Siple CE, et al. Missense substitutions in the GAS1 protein present in holoprosencephaly patients reduce the affinity for its ligand, SHH. *Human Genetics* 2012; 131: 301–10.

Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics* 2011; 12: 32–42.

Polakis P. Wnt signaling in cancer. *Cold Spring Harb Perspect Biol* 2012; 4

Porter FD. Smith–Lemli–Opitz syndrome: pathogenesis, diagnosis and management. *European Journal of Human Genetics* 2008; 16: 535–41.

Portnoi M-F, Gruchy N, Marlin S, Finkel L, Denoyelle F, Dubourg C, et al. Midline defects in deletion 18p syndrome: clinical and molecular characterization of three patients: *Clinical Dysmorphology* 2007; 16: 247–52.

Powers SE, Taniguchi K, Yen W, Melhuish TA, Shen J, Walsh CA, et al. Tgif1 and Tgif2 regulate Nodal signaling and are required for gastrulation. *Development* 2010; 137: 249–59.

Pradhan M, Agarwal S, Prasun P. One gene, many phenotypes. *Journal of Postgraduate Medicine* 2007; 53: 257.

Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics* 2015; 31: 761–3.

Quax TEF, Claassens NJ, Söll D, van der Oost J. Codon Bias as a Means to Fine-Tune Gene Expression. *Molecular Cell* 2015; 59: 149–61.

Quélin C, Bendavid C, Dubourg C, de la Rochebrochard C, Lucas J, Henry C, et al. Twelve new patients with 13q deletion syndrome: Genotype–phenotype analyses in progress. *European Journal of Medical Genetics* 2009; 52: 41–6.

Raina SK, Sharma S, Bhardwaj A, Singh M, Chaudhary S, Kashyap V. Malnutrition as a cause of mental retardation: A population-based study from Sub-Himalayan India. *J Neurosci Rural Pract* 2016; 7: 341–5.

Ramsbottom SA, Pownall ME. Regulation of Hedgehog Signalling Inside and Outside the Cell. *J Dev Biol* 2016; 4: 23.

Reis M d. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Research* 2004; 32: 5036–44.

Renaux A, Papadimitriou S, Versbraegen N, Nachtegaele C, Boutry S, Nowé A, et al. ORVAL: a novel platform for the prediction and exploration of disease-causing oligogenic variant combinations. *Nucleic Acids Research* 2019; 47: W93–8.

Ribeiro LA, Murray JC, Richieri-Costa A. PTCH mutations in four Brazilian patients with holoprosencephaly and in one with holoprosencephaly-like features and normal MRI. *American Journal of Medical Genetics Part A* 2006; 140A: 2584–6.

Richards S, Aziz N, Bale S, Bick D, Das S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine* 2015; 17: 405–23.

Rimkus T, Carpenter R, Qasem S, Chan M, Lo H-W. Targeting the Sonic Hedgehog Signaling Pathway: Review of Smoothed and GLI Inhibitors. *Cancers* 2016; 8: 22.

Riobo NA. Cholesterol and its derivatives in Sonic Hedgehog signaling and cancer. *Current Opinion in Pharmacology* 2012; 12: 736–41.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 2015; 43: e47–e47.

Robertson EJ. Dose-dependent Nodal/Smad signals pattern the early mouse embryo. *Seminars in Cell & Developmental Biology* 2014; 32: 73–9.

Robinson PN, Kohler S, Oellrich A, Sanger Mouse Genetics Project, Wang K, Mungall CJ, et al. Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Research* 2014; 24: 340–8.

Rodnina MV. The ribosome in action: Tuning of translational efficiency and protein folding: The Ribosome in Action. *Protein Science* 2016; 25: 1390–406.

Roessler E, Belloni E, Gaudenz K, Jay P, Berta P, Scherer SW, et al. Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. *Nature Genetics* 1996; 14: 357–60.

Roessler E, Belloni E, Gaudenz K, Vargas F, Scherer SW, Tsui L-C, et al. Mutations in the C-Terminal Domain of Sonic Hedgehog Cause Holoprosencephaly. *Human Molecular Genetics* 1997; 6: 1847–53.

Roessler E, El-Jaick KB, Dubourg C, Vélez JI, Solomon BD, Pineda-Álvarez DE, et al. The mutational spectrum of holoprosencephaly-associated changes within the *SHH* gene in humans predicts loss-of-function through either key structural alterations of the ligand or its altered synthesis. *Human Mutation* 2009; 30: E921–35.

Roessler E, Ma Y, Ouspenskaia MV, Lacbawan F, Bendavid C, Dubourg C, et al. Truncating loss-of-function mutations of *DISP1* contribute to holoprosencephaly-like microform features in humans. *Human Genetics* 2009; 125: 393–400.

Roessler E, Ouspenskaia MV, Karkera JD, Vélez JI, Kantipong A, Lacbawan F, et al. Reduced NODAL Signaling Strength via Mutation of Several Pathway Members Including *FOXH1* Is Linked to Human Heart Defects and Holoprosencephaly. *The American Journal of Human Genetics* 2008; 83: 18–29.

Rogozin IB, Gertz EM, Baranov PV, Poliakov E, Schaffer AA. Genome-Wide Changes in Protein Translation Efficiency Are Associated with Autism. *Genome Biology and Evolution* 2018; 10: 1902–19.

Romero-Barrios N, Legascue MF, Benhamed M, Ariel F, Crespi M. Splicing regulation by long noncoding RNAs. *Nucleic Acids Research* 2018; 46: 2169–84.

Roopra A, Dingledine R, Hsieh J. Epigenetics and epilepsy. *Epilepsia* 2012; 53 Suppl 9: 2–10.

Rosa RFM, Correia EPE, Bastos CS, da Silva GS, Correia JD, da Rosa EB, et al. Trisomy 18 and holoprosencephaly. *American Journal of Medical Genetics Part A* 2017; 173: 1985–7.

Rossant J, Tam PPL. Emerging Asymmetry and Embryonic Patterning in Early Mouse Development. *Developmental Cell* 2004; 7: 155–64.

Rubenstein JL, Beachy PA. Patterning of the embryonic forebrain. *Curr Opin Neurobiol* 1998; 8: 18–26.

Ruymann FB, Maddux HR, Ragab A, Soule EH, Palmer N, Beltangady M, et al. Congenital anomalies associated with rhabdomyosarcoma: An autopsy study of 115 cases. A report from the intergroup rhabdomyosarcoma study committee (representing the children's cancer study group, the pediatric oncology group, the United Kingdom children's cancer study group, and the pediatric intergroup statistical center). *Medical and Pediatric Oncology* 1988; 16: 33–9.

Rybicki BA, Elston RC. The Relationship between the Sibling Recurrence-Risk Ratio and Genotype Relative Risk. *The American Journal of Human Genetics* 2000; 66: 593–604.

Sagai T, Amano T, Maeno A, Ajima R, Shiroishi T. SHH signaling mediated by a prechordal and brain enhancer controls forebrain organization. *Proceedings of the National Academy of Sciences* 2019; 116: 23636–42.

Sahoo S, Das SS, Rakshit R. Codon usage pattern and predicted gene expression in *Arabidopsis thaliana*. *Gene: X* 2019; 2: 100012.

Sampath K, Rubinstein AL, Cheng AMS, Liang JO, Fekany K, Solnica-Krezel L, et al. Induction of the zebrafish ventral brain and floorplate requires cyclops/nodal signalling. *Nature* 1998; 395: 185–9.

Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* 1977; 74: 5463–7.

Santos-Cortez RLP, Khan V, Khan FS, Mughal Z-N, Chakchouk I, Lee K, et al. Novel candidate genes and variants underlying autosomal recessive neurodevelopmental disorders with intellectual disability. *Human Genetics* 2018; 137: 735–52.

Satoh W, Matsuyama M, Takemura H, Aizawa S, Shimono A. Sfrp1, Sfrp2, and Sfrp5 regulate the Wnt/ $\beta$ -catenin and the planar cell polarity pathways during early trunk formation in mouse. *genesis* 2008; 46: 92–103.

Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. *Nature Reviews Genetics* 2011; 12: 683–91.

Savastano CP, El-Jaick KB, Costa-Lima MA, Abath CMB, Bianca S, Cavalcanti DP, et al. Molecular analysis of holoprosencephaly in South America. *Genetics and Molecular Biology* 2014; 37: 250–62.

Schaaf CP, Sabo A, Sakai Y, Crosby J, Muzny D, Hawes A, et al. Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. *Hum Mol Genet* 2011; 20: 3366–75.

Schaefer MH, Lopes TJS, Mah N, Shoemaker JE, Matsuoka Y, Fontaine J-F, et al. Adding Protein Context to the Human Protein-Protein Interaction Network to Reveal Meaningful Interactions. *PLoS Computational Biology* 2013; 9: e1002860.

Schäffer AA. Digenic inheritance in medical genetics. *Journal of Medical Genetics* 2013; 50: 641–52.

Schell U, Wienberg J, Köhler A, Bray-Ward P, Ward DE, Wilson WG, et al. Molecular Characterization of Breakpoints in Patients with Holoprosencephaly and Definition of the HPE2 Critical Region 2p21. *Human Molecular Genetics* 1996; 5: 223–9.

Schneider RA, Hu D, Rubenstein JL, Maden M, Helms JA. Local retinoid signaling coordinates forebrain and facial morphogenesis by maintaining FGF8 and SHH. *Development* 2001; 128: 2755–67.

Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature Methods* 2014; 11: 361–2.

Seo H-C, Drivenes Ø, Ellingsen S, Fjose A. Expression of two zebrafish homologues of the murine Six3 gene demarcates the initial eye primordia. *Mechanisms of Development* 1998; 73: 45–57.

Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Research* 1987; 15: 7155–74.

Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. *Carcinogenesis* 2010; 31: 27–36.

Sharp PM, Li W-H. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research* 1987; 15: 1281–95.

Shastri VM, Schmidt KH. Cellular defects caused by hypomorphic variants of the Bloom syndrome helicase gene BLM. *Mol Genet Genomic Med* 2016; 4: 106–19.

Shen MM. Nodal signaling: developmental roles and regulation. *Development* 2007; 134: 1023–34.

Sherry ST. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research* 2001; 29: 308–11.

Shihab HA, Rogers MF, Gough J, Mort M, Cooper DN, Day INM, et al. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. *Bioinformatics* 2015; 31: 1536–43.

Shyr D, Liu Q. Next generation sequencing in cancer research and clinical application [Internet]. *Biological Procedures Online* 2013; 15[cited 2020 Jun 25] Available from: <https://biologicalproceduresonline.biomedcentral.com/articles/10.1186/1480-9222-15-4>

Simon EM, Hevner RF, Pinter JD, Clegg NJ, Delgado M, Kinsman SL, et al. The middle interhemispheric variant of holoprosencephaly. *AJNR Am J Neuroradiol* 2002; 23: 151–6.

Simonis N, Migeotte I, Lambert N, Perazzolo C, de Silva DC, Dimitrov B, et al. *FGFR1* mutations cause Hartsfield syndrome, the unique association of holoprosencephaly and ectrodactyly. *Journal of Medical Genetics* 2013; 50: 585–92.

Singh T, Walters JTR, Johnstone M, Curtis D, et al. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nature Genetics* 2017; 49: 1167–73.

Slatkin M. Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics* 2008; 9: 477–85.

Slavotinek A. Genetic modifiers in human development and malformation syndromes, including chaperone proteins. *Human Molecular Genetics* 2003; 12: 45R – 50.

Slifer SH. PLINK: Key Functions for Data Analysis: PLINK: Key Functions for Data Analysis. *Current Protocols in Human Genetics* 2018; 97: e59.

Smedley D, Jacobsen JOB, Jäger M, Köhler S, Holtgrewe M, Schubach M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nature Protocols* 2015; 10: 2004–15.

Smedley D, Robinson PN. Phenotype-driven strategies for exome prioritization of human Mendelian disease genes [Internet]. *Genome Medicine* 2015; 7[cited 2020 May 17] Available from: <https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-015-0199-2>

Solomon BD, Bear KA, Wyllie A, Keaton AA, Dubourg C, David V, et al. Genotypic and phenotypic analysis of 396 individuals with mutations in *Sonic Hedgehog*. *Journal of Medical Genetics* 2012; 49: 473–9.

Solomon BD, Kruszka P, Muenke M. Holoprosencephaly flashcards: An updated summary for the clinician. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 2018; 178: 117–21.

Solomon BD, Lacbawan F, Jain M, Domené S, Roessler E, Moore C, et al. A novel *SIX3* mutation segregates with holoprosencephaly in a large family. *Am J Med Genet A* 2009; 149A: 919–25.

Solomon BD, Lacbawan F, Mercier S, Clegg NJ, Delgado MR, Rosenbaum K, et al. Mutations in *ZIC2* in human holoprosencephaly: description of a Novel *ZIC2* specific phenotype and comprehensive analysis of 157 individuals. *Journal of Medical Genetics* 2010; 47: 513–24.

Solomon BD, Rosenbaum KN, Meck JM, Muenke M. Holoprosencephaly due to numeric chromosome abnormalities. *Am J Med Genet C Semin Med Genet* 2010; 154C: 146–8.

Song HH, Shi W, Xiang Y-Y, Filmus J. The Loss of Glypican-3 Induces Alterations in Wnt Signaling. *Journal of Biological Chemistry* 2005; 280: 2116–25.

Song J, Oh SP, Schrewe H, Nomura M, Lei H, Okano M, et al. The Type II Activin Receptors Are Essential for Egg Cylinder Growth, Gastrulation, and Rostral Head Development in Mice. *Developmental Biology* 1999; 213: 157–69.

Soukariéh O, Gaildrat P, Hamieh M, Drouet A, Baert-Desurmont S, Frébourg T, et al. Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. *PLOS Genetics* 2016; 12: e1005756.

Spataro N, Rodríguez JA, Navarro A, Bosch E. Properties of human disease genes and the role of genes linked to Mendelian disorders in complex disease aetiology. *Human Molecular Genetics* 2017: ddw405.

Stamatakis D. A gradient of Gli activity mediates graded Sonic Hedgehog signaling in the neural tube. *Genes & Development* 2005; 19: 626–41.

Stasiulewicz M, Gray SD, Mastromina I, Silva JC, Bjorklund M, Seymour PA, et al. A conserved role for Notch signaling in priming the cellular response to Shh through ciliary localisation of the key Shh transducer Smo. *Development* 2015; 142: 2291–303.

Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Human Genetics* 2017; 136: 665–77.

Stevens CA. Steinfeld syndrome: Further delineation. *American Journal of Medical Genetics Part A* 2010; 152A: 1789–92.

Stiles J, Jernigan TL. The basics of brain development. *Neuropsychol Rev* 2010; 20: 327–48.

Stokes B, Berger SI, Hall BA, Weiss K, Martinez AF, Hadley DW, et al. *SIX3* deletions and incomplete penetrance in families affected by holoprosencephaly: *SIX3* deletions and holoprosencephaly. *Congenital Anomalies* 2018; 58: 29–32.

Stolfi A, Wagner E, Taliaferro JM, Chou S, Levine M. Neural tube patterning by Ephrin, FGF and Notch signaling relays. *Development* 2011; 138: 5429–39.

Storm EE. Dose-dependent functions of *Fgf8* in regulating telencephalic patterning centers. *Development* 2006; 133: 1831–44.

Summers AD, Reefhuis J, Taliano J, Rasmussen SA. Nongenetic risk factors for holoprosencephaly: An updated review of the epidemiologic literature. *Am J Med Genet C Semin Med Genet* 2018; 178: 151–64.

Taniguchi K, Ayada T, Ichiyama K, Kohno R, Yonemitsu Y, Minami Y, et al. *Sprouty2* and *Sprouty4* are essential for embryonic morphogenesis and regulation of FGF signaling. *Biochemical and Biophysical Research Communications* 2007; 352: 896–902.

Teare MD, Santibanez Koref MF. Linkage analysis and the study of Mendelian disease in the era of whole exome and genome sequencing. *Briefings in Functional Genomics* 2014; 13: 378–83.

Tekendo-Ngongang C, Kruszka P, Martinez AF, Muenke M. Novel heterozygous variants in *KMT2D* associated with holoprosencephaly. *Clinical Genetics* 2019; 96: 266–70.

Tekendo-Ngongang C, Muenke M, Kruszka P. Holoprosencephaly Overview [Internet]. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJ, Stephens K, et al., editor(s). *GeneReviews*®. Seattle (WA): University of Washington, Seattle; 1993[cited 2020 Jun 21] Available from: <http://www.ncbi.nlm.nih.gov/books/NBK1530/>

Tenzen T, Allen BL, Cole F, Kang J-S, Krauss RS, McMahon AP. The Cell Surface Membrane Proteins *Cdo* and *Boc* Are Components and Targets of the Hedgehog Signaling Pathway and Feedback Network in Mice. *Developmental Cell* 2006; 10: 647–56.

The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015; 526: 68–74.

The ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 2007; 447: 799–816.

The FANTOM Consortium. The Transcriptional Landscape of the Mammalian Genome. *Science* 2005; 309: 1559–63.

Thuriot F, Buote C, Gravel E, Chénier S, Désilets V, Maranda B, et al. Clinical validity of phenotype-driven analysis software PhenoVar as a diagnostic aid for clinical geneticists in the interpretation of whole-exome sequencing data. *Genetics in Medicine* 2018; 20: 942–9.

Torrey EF, Miller J, Rawlings R, Yolken RH. Seasonality of births in schizophrenia and bipolar disorder: a review of the literature. *Schizophrenia Research* 1997; 28: 1–38.

Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009; 25: 1105–11.

Tukachinsky H, Lopez LV, Salic A. A mechanism for vertebrate Hedgehog signaling: recruitment to cilia and dissociation of SuFu–Gli protein complexes. *The Journal of Cell Biology* 2010; 191: 415–28.

- Tylee DS, Kawaguchi DM, Glatt SJ. On the outside, looking in: A review and evaluation of the comparability of blood and brain “-omes”. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 2013; 162: 595–603.
- Varlet I, Collignon J, Norris DP, Robertson EJ. Nodal signaling and axis formation in the mouse. *Cold Spring Harb Symp Quant Biol* 1997; 62: 105–13.
- Verloes A, Dodinval P, Beco L, Bonnivert J, Lambotte C. Lambotte syndrome: Microcephaly, holoprosencephaly, intrauterine growth retardation, facial anomalies, and early lethality—A new sublethal multiple congenital anomaly/mental retardation syndrome in four sibs. *American Journal of Medical Genetics* 1990; 37: 119–23.
- Versbraegen N, Fouché A, Nachtegaele C, Papadimitriou S, Gazzo A, Smits G, et al. Using game theory and decision decomposition to effectively discern and characterise bi-locus diseases. *Artif Intell Med* 2019; 99: 101690.
- Vijayan V, Ummer R, Weber R, Silva R, Letra A. Association of WNT Pathway Genes With Nonsyndromic Cleft Lip With or Without Cleft Palate. *The Cleft Palate-Craniofacial Journal* 2018; 55: 335–41.
- Visscher PM, Brown MA, McCarthy MI, Yang J. Five Years of GWAS Discovery. *The American Journal of Human Genetics* 2012; 90: 7–24.
- Vladoiu MC, El-Hamamy I, Donovan LK, Farooq H, Holgado BL, Sundaravadanam Y, et al. Childhood cerebellar tumours mirror conserved fetal transcriptional programs. *Nature* 2019; 572: 67–73.
- Voelkerding KV, Dames S, Durtschi JD. Next generation sequencing for clinical diagnostics-principles and application to targeted resequencing for hypertrophic cardiomyopathy: a paper from the 2009 William Beaumont Hospital Symposium on Molecular Pathology. *J Mol Diagn* 2010; 12: 539–51.
- Voelkerding KV, Dames SA, Durtschi JD. Next-Generation Sequencing: From Basic Research to Diagnostics. *Clinical Chemistry* 2009; 55: 641–58.
- Votintseva AA, Bradley P, Pankhurst L, del Ojo Elias C, Loose M, Nilgiriwala K, et al. Same-Day Diagnostic and Surveillance Data for Tuberculosis via Whole-Genome Sequencing of Direct Respiratory Samples. *Journal of Clinical Microbiology* 2017; 55: 1285–98.
- Vu TN, Deng W, Trac QT, Calza S, Hwang W, Pawitan Y. A fast detection of fusion genes from paired-end RNA-seq data [Internet]. *BMC Genomics* 2018; 19[cited 2020 May 15] Available from: <https://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-018-5156-1>
- Wahl MB, Deng C, Lewandoski M, Pourquie O. FGF signaling acts upstream of the NOTCH and WNT signaling pathways to control segmentation clock oscillations in mouse somitogenesis. *Development* 2007; 134: 4033–41.
- Wai LT, Chandran S. Cyclopia: isolated and with agnathia–otocephaly complex. *BMJ Case Reports* 2017: bcr-2017-220159.
- Wallis DE, Roessler E, Hehr U, Nanni L, Wiltshire T, Richieri-Costa A, et al. Mutations in the homeodomain of the human SIX3 gene cause holoprosencephaly. *Nature Genetics* 1999; 22: 196–8.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* 2010; 38: e164–e164.
- Wang RN, Green J, Wang Z, Deng Y, Qiao M, Peabody M, et al. Bone Morphogenetic Protein (BMP) signaling in development and human diseases. *Genes & Diseases* 2014; 1: 87–105.
- Wang T, Zhou T, Saadat L, Garcia JG. A MYLK variant regulates asthmatic inflammation via alterations in mRNA secondary structure. *European Journal of Human Genetics* 2015; 23: 874–6.
- Wang W, Corominas R, Lin GN. De novo Mutations From Whole Exome Sequencing in Neurodevelopmental and Psychiatric Disorders: From Discovery to Application. *Front Genet* 2019; 10: 258.
- Wang X, Song P, Huang C, Yuan N, Zhao X, Xu C. Weighted gene co-expression network analysis for identifying hub genes in association with prognosis in Wilms tumor [Internet]. *Molecular Medicine Reports* 2019[cited 2020 May 17] Available from: <http://www.spandidos-publications.com/10.3892/mmr.2019.9881>
- Wang Y, Song L, Zhou CJ. The canonical Wnt/ $\beta$ -catenin signaling pathway regulates Fgf signaling for early facial development. *Developmental Biology* 2011; 349: 250–60.
- Wang Y-P, Wang D-J, Niu Z-B, Cui W-T. Chromosome 13q deletion syndrome involving 13q31-qter: A case report. *Molecular Medicine Reports* 2017; 15: 3658–64.

- Warr N, Powles-Glover N, Chappell A, Robson J, Norris D, Arkell RM. Zic2 -associated holoprosencephaly is caused by a transient defect in the organizer region during gastrulation. *Human Molecular Genetics* 2008; 17: 2986–96.
- Waters AM, Beales PL. Ciliopathies: an expanding disease spectrum. *Pediatric Nephrology* 2011; 26: 1039–56.
- Wei Y, Silke JR, Xia X. An improved estimation of tRNA expression to better elucidate the coevolution between tRNA abundance and codon usage in bacteria [Internet]. *Scientific Reports* 2019; 9[cited 2020 Jun 14] Available from: <http://www.nature.com/articles/s41598-019-39369-x>
- Weiling F. Historical study: Johann Gregor Mendel 1822-1884. *American Journal of Medical Genetics* 1991; 40: 1–25.
- Weinreich SS, Mangon R, Sikkens JJ, Teeuw ME en, Cornel MC. [Orphanet: a European database for rare diseases]. *Ned Tijdschr Geneesk* 2008; 152: 518–9.
- Weiss KM, Clark AG. Linkage disequilibrium and the mapping of complex human traits. *Trends in Genetics* 2002; 18: 19–24.
- Whiffin N, Minikel E, Walsh R, O'Donnell-Luria AH, Karczewski K, Ing AY, et al. Using high-resolution variant frequencies to empower clinical genome interpretation. *Genetics in Medicine* 2017; 19: 1151–8.
- Wilde JJ, Petersen JR, Niswander L. Genetic, epigenetic, and environmental contributions to neural tube closure. *Annu Rev Genet* 2014; 48: 583–611.
- Wilson SW, Houart C. Early steps in the development of the forebrain. *Dev Cell* 2004; 6: 167–81.
- Winnier G, Blessing M, Labosky PA, Hogan BL. Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes & Development* 1995; 9: 2105–16.
- Wozniak RH, Leezenbaum NB, Northrup JB, West KL, Iverson JM. The development of autism spectrum disorders: variability and causal complexity. *Wiley Interdisciplinary Reviews: Cognitive Science* 2017; 8: e1426.
- Xie J, Owen T, Xia K, Singh AV, Tou E, Li L, et al. Zinc Inhibits Hedgehog Autoprocessing: LINKING ZINC DEFICIENCY WITH HEDGEHOG ACTIVATION. *Journal of Biological Chemistry* 2015; 290: 11591–600.
- Yadegari H, Biswas A, Akhter MS, Driesen J, Ivaskевичius V, Marquardt N, et al. Intron retention resulting from a silent mutation in the VWF gene that structurally influences the 5' splice site. *Blood* 2016; 128: 2144–52.
- Yamakawa K, Mitchell S, Hubert R, Chen X-N, Colbern S, Huo Y-K, et al. Isolation and characterization of a candidate gene for progressive myoclonus epilepsy on 21q22.3. *Human Molecular Genetics* 1995; 4: 709–16.
- Yamamoto M, Beppu H, Takaoka K, Meno C, Li E, Miyazono K, et al. Antagonism between Smad1 and Smad2 signaling determines the site of distal visceral endoderm formation in the mouse embryo. *Journal of Cell Biology* 2009; 184: 323–34.
- Yang L, Xie G, Fan Q, Xie J. Activation of the hedgehog-signaling pathway in human cancer and the clinical implications. *Oncogene* 2010; 29: 469–81.
- Yang Y-P, Anderson RM, Klingensmith J. BMP antagonism protects Nodal signaling in the gastrula to promote the tissue interactions underlying mammalian forebrain and craniofacial patterning. *Human Molecular Genetics* 2010; 19: 3030–42.
- Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* 1998; 15: 568–73.
- Yao Y, Minor PJ, Zhao Y-T, Jeong Y, Pani AM, King AN, et al. Cis-regulatory architecture of a brain signaling center predates the origin of chordates. *Nature Genetics* 2016; 48: 575–80.
- Yayon A, Klagsbrun M, Esko JD, Leder P, Ornitz DM. Cell surface, heparin-like molecules are required for binding of basic fibroblast growth factor to its high affinity receptor. *Cell* 1991; 64: 841–8.
- Yeo G, Burge CB. Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *Journal of Computational Biology* 2004; 11: 377–94.
- Yi Z, Yingjun X, Yongzhen C, Liangying Z, Meijiao S, Baojiang C. Prenatal diagnosis of pure partial monosomy 18p associated with holoprosencephaly and congenital heart defects. *Gene* 2014; 533: 565–9.

Young NM, Chong HJ, Hu D, Hallgrímsson B, Marcucio RS. Quantitative analyses link modulation of sonic hedgehog signaling to continuous variation in facial growth and shape. *Development* 2010; 137: 3405–9.

Yu C-H, Dang Y, Zhou Z, Wu C, Zhao F, Sachs MS, et al. Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Molecular Cell* 2015; 59: 744–54.

Zakin L. Inactivation of mouse Twisted gastrulation reveals its role in promoting Bmp4 activity during forebrain development. *Development* 2003; 131: 413–24.

Zentner GE, Layman WS, Martin DM, Scacheri PC. Molecular and phenotypic aspects of *CHD7* mutation in CHARGE syndrome. *American Journal of Medical Genetics Part A* 2010; 152A: 674–86.

Zhang H, Bradley A. Mice deficient for BMP2 are nonviable and have defects in amnion/chorion and cardiac development. *Development* 1996; 122: 2977–86.

Zhang W, Kang J-S, Cole F, Yi M-J, Krauss RS. Cdo Functions at Multiple Points in the Sonic Hedgehog Pathway, and Cdo-Deficient Mice Accurately Model Human Holoprosencephaly. *Developmental Cell* 2006; 10: 657–65.

Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning–based sequence model. *Nature Methods* 2015; 12: 931–4.

Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003; 31: 3406–15.

***ACKNOWLEDGEMENTS  
AND PERSONAL NOTES***

*Merci à tous*

*Thank you everyone*

*Благодарю вас всех*

---

*SUCCESS IS NOT FINAL, FAILURE IS NOT FATAL: IT IS THE  
COURAGE TO CONTINUE THAT COUNTS.*

---

*-Winston Churchill*

---

*I THINK THERE ARE PEOPLE THAT HELP YOU BECOME  
THE PERSON YOU END UP BEING,  
AND YOU CAN BE GRATEFUL FOR THEM,  
EVEN IF THEY WERE NEVER MEANT TO BE IN YOUR LIFE  
FOREVER.*

*I'M GLAD I KNEW YOU TOO.*

---

*-Diane*

---

*À LA FIN, TOUT IRA BIEN.  
SI ÇA NE VA PAS, C'EST QUE CE N'EST PAS LA FIN.*

---



**Exploration des bases génétiques de l'Holoprosencéphalie :  
apport de la bio-informatique pour l'étude d'un modèle de maladie génétique  
complexe**

Par Artem KIM

Institut de Génétique et Développement de Rennes (IGDR), UMR6290 CNRS

Université Rennes 1

Les maladies génétiques sont reconnues comme l'une des principales catégories de maladies humaines et résultent d'un dysfonctionnement cellulaire causé par une ou plusieurs variations de la séquence du génome, appelées mutations ou variants génétiques. Lorsque les anomalies en cause touchent les cellules germinales de l'individu (génome constitutionnel), les maladies génétiques sont transmises à la descendance et sont dites héréditaires. Les mutations sont à la fois une source de diversité génétique et un mécanisme majeur de l'évolution. Cependant ces variants génétiques peuvent également induire un mauvais fonctionnement des cellules et donc être à l'origine d'une pathologie. Aujourd'hui plus de 6,000 maladies génétiques sont répertoriées, beaucoup d'entre elles sont mortelles ou provoquent des affections extrêmement graves. L'un des enjeux les plus importants dans l'étude de ces maladies est d'identifier le ou les facteurs génétiques responsables. De par la connaissance des gènes impliqués dans les maladies humaines, le conseil génétique permet alors d'évaluer le risque de survenue d'une maladie héréditaire chez un individu. La mise au point de tests diagnostiques étendus permet aujourd'hui d'affiner le diagnostic et d'établir une prise en charge thérapeutique adaptée. Cependant certaines maladies héréditaires restent difficiles à diagnostiquer et de nombreux gènes restent à découvrir.

C'est le cas des maladies génétiques complexes, également appelées les maladies non-Mendéliennes. Contrairement aux maladies monogéniques ou Mendéliennes, causées par une seule mutation à effet pathogène fort, les maladies complexes résultent d'une accumulation de plusieurs mutations à effet faible, également appelées les variants hypomorphes. La présence d'un seul variant hypomorphe n'a pas de conséquences cliniques, ces variants peuvent donc être retrouvés dans le génome des individus asymptomatiques. Cependant, l'effet cumulé de plusieurs variants hypomorphes dans plusieurs gènes peut conduire à la pathogénèse et l'apparition d'une maladie génétique. De part de leur effet délétère faible et transmission par des individus non-atteints, les variants hypomorphes responsables de maladies génétiques sont extrêmement difficiles à identifier dans le cadre clinique. Ainsi, les maladies génétiques complexes présentent une étiologie impliquant plusieurs facteurs génétiques encore mal compris.

L'identification de ces facteurs nécessite une meilleure interprétation du rôle des variants génétiques identifiés par le séquençage haut-débit dans le cadre clinique. En raison d'un grand nombre de variants identifiés par séquençage, les stratégies classiques d'identification de mutations pathogènes se basent sur les approches de filtration afin d'exclure les variants non-pathogènes (polymorphismes) et identifier un nombre restreint de variants candidats présentant des arguments en faveur de leur implication dans la maladie. Les stratégies classiques de filtration sont basées sur la fréquence des variants en population générale et leur effet délétère prédit par des outils bio-informatiques. Cependant, ces stratégies sont insuffisantes pour identifier les variants hypomorphes impliqués dans les maladies génétiques complexes. Des approches complémentaires de priorisation de variants, comme le phénotypage profond et intégration de données d'expression des gènes, sont nécessaires afin d'améliorer le diagnostic des maladies complexes.

Mes travaux de thèse s'inscrivent dans le cadre des recherches menées dans l'équipe Génétique des pathologies liées au développement (IGDR UMR6290 Rennes), engagée dans l'étude des mécanismes de la voie Sonic Hedgehog (SHH) impliquée dans le développement du cerveau antérieur. Une dérégulation de cette voie conduit à une pathologie extrêmement sévère, caractérisée par des anomalies cranio-faciales, l'holoprosencéphalie (HPE). Cette maladie résulte d'un défaut de clivage du prosencéphale lors du développement cérébral précoce, conduisant à différentes manifestations phénotypiques au degré de sévérité variable. HPE est associée à un rendement diagnostique très faible due à sa complexité génétique. Initialement considérée comme une maladie à transmission autosomique dominante, l'HPE a été redéfinie très récemment comme une maladie complexe nécessitant plusieurs évènements mutationnels touchant des gènes impliqués dans les voies de signalisation du développement cérébral, notamment la voie SHH. L'existence de rares cas de familles consanguines atteintes d'HPE suggère que le mode de transmission autosomique récessif est également envisageable.

L'objectif de ma thèse a été de préciser les mécanismes génétiques impliqués dans l'HPE et de proposer des nouvelles approches bio-informatiques pour améliorer le diagnostic des maladies génétiques complexes.

Dans un premier temps, j'ai exploré et décrit pour la première fois des cas d'HPE causés par un effet cumulatif de plusieurs variants rares agissant de manière hypomorphe. Dans cette étude, j'ai développé une nouvelle stratégie de priorisation de variants en intégrant les associations gène-phénotype et l'étude des profils d'expression des gènes au cours du développement cérébral. Appliquée aux données de séquençage de 26 familles HPE, cette approche a permis d'identifier des combinaisons particulières de variants ayant une association forte avec la maladie et touchant des gènes associés à des voies clés du développement cérébral, notamment la voie SHH. Ces combinaisons de variants étaient retrouvées exclusivement chez les

individus atteints de chaque famille. J'ai pu valider l'implication de ces combinaisons dans la maladie en montrant leur enrichissement significatif dans l'HPE comparé à la population générale (cohortes contrôles GoNL, FREX). Ainsi mes travaux ont montré l'importance de considérer l'effet combiné de plusieurs variants lors du diagnostic des patients.

Par la suite, j'ai démontré l'impact pathogène de certaines variations synonymes du gène *SHH* sur la traduction, conduisant à un défaut de repliement de la protéine et apparition de l'HPE. L'analyse rétrospective de données de séquençage de 931 patients a permis d'identifier huit variants synonymes dans le gène *SHH*, significativement enrichis dans l'HPE comparé aux populations contrôles (gnomAD, FREX, GoNL), suggérant ainsi leur rôle dans la maladie. Les analyses statistiques d'usage de codons ont indiqué un effet de cinq de ces variants sur la traduction du gène *SHH* : par introduction de codons synonymes possédant les propriétés biochimiques différentes, ces variants modifieraient la vitesse de la traduction et la capacité de la protéine à se replier correctement. Les expériences *in vitro* ont démontré que ces cinq variants synonymes sont en effet associés à une réduction significative de la quantité de la protéine produite, réduction allant de 5 % à 23 %. En inhibant le protéasome, nous avons pu restaurer les quantités protéiques pour quatre de ces cinq variants synonymes, confirmant ainsi leur impact sur la stabilité et le repliement de la protéine SHH. De plus, j'ai pu montrer une corrélation significative entre les valeurs de réduction protéique évaluées expérimentalement et les mesures computationnelles d'usage de codons ( $R^2=0.83$ ;  $P=0.0016$ ), soulignant ainsi la pertinence de certains modèles permettant de prédire l'impact des variants synonymes sur la traduction. Les résultats de cette étude indiquent que les variations synonymes peuvent avoir un rôle majeur dans les pathologies génétiques encore mal comprises.

Les travaux de ma thèse contribuent à la compréhension de l'architecture génétique complexe de l'HPE et proposent de nouvelles méthodes d'analyse des mécanismes génétiques des maladies complexes. A terme, ces résultats devraient permettre de réduire l'errance diagnostique et d'améliorer la prise en charge des patients atteints des maladies génétiques complexes.

**Titre :** Exploration de l'interaction entre variants rares et communs dans la susceptibilité génétique à holoprosencéphalie

**Mots clés :** Holoprosencéphalie, maladies complexes, variants hypomorphes, oligogénisme, développement cérébral, bioinformatique, séquençage à haut débit

**Résumé :** Les maladies génétiques complexes présentent une étiologie impliquant plusieurs facteurs génétiques encore mal compris. L'identification de ces facteurs nécessite une meilleure interprétation du rôle des variants génétiques identifiés par le séquençage haut-débit dans le cadre clinique. Mes travaux de thèse concernent l'étude d'holoprosencéphalie (HPE) - une pathologie cérébrale extrêmement sévère, associée à un rendement diagnostique très faible due à sa complexité génétique. L'objectif de ma thèse est de préciser les mécanismes génétiques impliqués dans l'HPE et de proposer des nouvelles approches bio-informatiques pour améliorer le diagnostic des maladies génétiques complexes.

Au cours de ma thèse, j'ai tout d'abord identifié et décrit pour la première fois des cas d'HPE oligogénique, causés par un effet cumulatif de plusieurs variants rares agissant de manière hypomorphe dans les gènes liés à la voie Sonic Hedgehog (SHH). Ces travaux ont démontré

l'importance de considérer l'effet combiné de plusieurs mutations lors du diagnostic des patients. Par la suite, j'ai démontré l'impact pathogène des mutations synonymes du gène *SHH* sur la traduction : par introduction des codons possédant les propriétés biochimiques différentes, ces variants modifient la capacité de la protéine à se replier correctement. Ces résultats indiquent que les mutations synonymes peuvent avoir un rôle majeur dans l'étiologie des maladies génétiques.

Les travaux de ma thèse contribuent à la compréhension de l'architecture génétique complexe de l'HPE et proposent de nouvelles méthodes d'analyse des mécanismes génétiques des maladies complexes, notamment l'hérédité oligogénique et l'impact des variants synonymes. À terme, ces résultats devraient permettre de réduire l'errance diagnostique et d'améliorer la prise en charge des patients atteints des pathologies génétiques encore mal comprises.

**Title:** Exploration of interaction between common and rare variants in genetic susceptibility to holoprosencephaly

**Keywords:** Holoprosencephaly, complex disorders, hypomorphic variants, oligogenism, cerebral development, bioinformatics, next generation sequencing

**Abstract:** The etiology of complex genetic disorders involves multiple genetic factors that are still poorly understood. Identification of these factors requires a better clinical interpretation of genetic variants identified by high-throughput sequencing. My thesis concerns the study of holoprosencephaly (HPE) - an extremely severe cerebral pathology associated with a very low diagnostic yield due to its genetic complexity. The objective of my thesis is to elucidate the genetic mechanisms underlying HPE and to propose novel bioinformatic strategies of pathogenic variant interpretation in complex genetic disorders.

During my thesis, I first identified and described cases of oligogenic HPE resulting from a joint effect of several rare hypomorphic variants in genes related to the Sonic Hedgehog (SHH) pathway. This work has shown the importance to consider the

combined effect of several mutations in molecular diagnosis of complex genetic disorders. Second, I demonstrated the pathogenic impact of synonymous mutations in the *SHH* gene on translation: by introducing codons with different biochemical properties, these variants modify the capacity of the protein to fold correctly. These results indicate that synonymous mutations may play a major role in etiology of genetic disorders.

Overall, my thesis contributes to the understanding of complex genetic architecture of HPE and proposes novel analytical methods to investigate the genetic mechanisms underlying complex disorders, such as oligogenic inheritance and synonymous variants. Ultimately, these results should help avoid misdiagnosis and improve genetic counseling in human disorders that remain to be resolved.