Supplementary Appendix

Integrated Clinical and Omics Approach to Rare Diseases : Novel Genes and Support for Oligogenic Inheritance in Holoprosencephaly

Artem Kim Ph.D., Clara Savary Ph.D., Christèle Dubourg Pharm.D., Ph.D., Wilfrid Carré Ph.D., Charlotte Mouden Ph.D., Houda Hamdi-Rozé M.D, Ph.D., Hélène Guyodo, Jerome Le Douce M.S., FREX Consortium, Laurent Pasquier M.D., Elisabeth Flori M.D., Marie Gonzales M.D., Claire Bénéteau, M.D., Odile Boute M.D., Tania Attié-Bitach M.D. Ph.D., Joelle Roume M.D., Louise Goujon M.S., Linda Akloul M.D., Sylvie Odent M.D., Ph.D., Erwan Watrin Ph.D., Valérie Dupé Ph.D., Marie de Tayrac Ph.D., Véronique David Pharm.D., Ph.D.

Table of Contents

Pipeline description

Whole Exome Sequencing

For data alignment and variant calling, a pipeline using Burrows-Wheeler Aligner (BWA, v0.7.12), Genome Analysis toolkit (GATK 3.x)[1] and Freebayes[2] (v1.1.0) was applied to all patients following standard procedures.

For each patient, the sequenced reads were aligned to the human reference genome (hg19), followed by removal of PCR duplicates ('MarkDuplicates', Picard Tools v1.114) and recalibration of alignment quality scores using GATK Base Quality Score Recalibration (BQSR). Variant calling was performed using 3 different algorithms: *HaplotypeCaller* and *UnifiedGenotyper* from GATK and, additionally, Freebayes. Low-quality variant calls were excluded using the default GATK filters (*Table S1*). A consensus VCF (variant calling format) file was created using the GATK CombineVariants program containing the union of all variants detected by at least one caller, without any given priority (*genotypemergeoption=UNSORTED*, *filteredrecordsmergetype=KEEP_IF_ANY_UNFILTERED*). Genotypes of the resulting variants were then recalculated (custom scripts) using the alternate allele frequencies (AF) as follows:

- Homozygous reference: AF < 0.01
- Heterozygous: AF $\in$ [0.20;0.80]
- Homozygous alternate: AF > 0.99
- Ambiguous: AF $\in$ [0.01;0.20] $\cup$ [0.80;0.99]

The resulting variants were annotated with ANNOVAR (February 2016 build)[3]. A gene-based annotation was performed using RefGene[4] to identify the variant functional classes (missense/intronic/UTR/splicing/frameshift…), the genes and the amino acid changes. For variant frequency annotation, we used allelic population frequency information from 7 control databases: dbSNP[5] (build 147), 1000 Genomes[6] (August 2015), Kaviar[7] (September 2015), Exome Sequencing Project[8] (ESP, March 2015), Greater Middle East Variome[9] (GME), Exome Aggregation Consortium[10] (ExAC) and its new version - Genome Aggregation Database[10] (gnomAD). To identify recurrent false positives, the annotation step was completed by adding our in-house exome data (variants frequency during previous runs and annotations conducted in our lab).
Potential deleterious effects were assessed using a total of 21 prediction algorithms (*Table S2*) provided by dbNSFP[11] (functional predictions of non-synonymous variants, version 3.3a) and dbscSNV[12] (functional predictions of splice variants). Complementary annotations were performed using Alamut Visual v2.8.1 (Interactive Biosoftware, Rouen, France) and Human Splicing Finder[13] (HSF).
A filtering step was performed on a family-by-family basis. All possible Mendelian inheritance models were considered for this study. In order to exclude common polymorphisms, non-

pathogenic variants and the remaining sequencing errors, the following parameters were 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)
(6) Exlude nonframeshift INDELS outside of known HPE genes
(7) Exclude variants with ambiguous allelic ratio (intervals [0.01;0.20] and [0.80;0.99])

Finally, we used Integrative Genomics Viewer[14] (IGV) for visual validation of candidate variants.

To identify disease-related variants among a large number of non-pathogenic polymorphisms, we developed an integrative variant prioritization approach that automatically reduces the exome search space by restricting the variant analysis to candidate disease-relevant genes. It combines classical WES variant filtering based on population frequency and pathogenicity predictions with two complementary gene mapping approaches: clinically-driven and transcriptome-driven strategies.

Clinically-driven strategy

To identify genes associated to clinical phenotypes overlapping with HPE, we established two clinician-generated lists of relevant phenotypes reminiscent of HPE in human and mouse models respectively *(Table S3)*. To select genes associated to any of clinical phenotypes of interest, we used the data of public clinical resources and associated ontologies (*Figure S1*). Human gene-phenotype associations were extracted from Clinvar[15], Orphanet[16], Uniprot[17], Online Mendelian Inheritance In Man(OMIM)[18] and Human Phenotype Ontology (HPO)[19] databases using R package *VarFromPDB* (*https://github.com/cran/VarfromPDB*). The same approach was performed for mouse phenotype database Mouse Genome Informatics (MGI)[20] but using a homemade workflow (R). After combining the results and removing redundancy, we established a list of 659 clinically-driven candidate genes (Table S4).

Transcriptome-driven strategy

The objective of the transcriptome-driven strategy was to identify genes sharing highly similar expression patterns with major HPE genes during cerebral development. Based on the current level of evidence (described in the main text), we considered *SHH, ZIC2, SIX3* and *TGIF1* as major HPE genes, excluding *GLI2* because of its ubiquitous expression pattern in the forebrain during embryonic development.[21] We also explored the expression profiles of minor HPE

genes - *PTCH1*, *FGF8*, *DISP1*, *TDGF1*, *FOXH1*, *CDON*, *NODAL*, *DLL1*, *GAS1*, *STIL* and *FGFR1.* Recent reports also suggest the implication of *SUFU* and *BOC* in HPE, however there was only one case with pathogenic variants in *SUFU* reported so far, while *BOC* was reported to act as a silent modifier in HPE.[51] Due to the lack of sufficient evidence, *SUFU* and *BOC* were not included in the HPE genes list for the transcriptomic analysis.

For transcriptome analysis, we used RNA-Seq data from Human Developmental Biology Resource[22] (HDBR). Data corresponding to a total of 136 samples of pre-natal human brain (*Figure S9*) were selected for this study. Each sample was characterized by a specific developmental stage (4-10 post-conception weeks/pcw) and structure (forebrain, cerebral cortex, diencephalon, telencephalon, temporal lobe). Samples were analyzed with the iRAP pipeline (version 0.8.1d8)[23]. The raw data, results, experimental procedures and detailed analysis methods are available at https://www.ebi.ac.uk/gxa/experiments/E-MTAB-4840. Briefly, reads below minimum quality threshold were discarded; mapping against the reference genome (Ensembl release: 79) was performed by tophat[24] (version 2.1.0) and gene expression quantification was done by htseq2[25] (version 0.6.1p1).

We used R package Weighted Gene Co-expression Network Analysis[26] (WGCNA) to construct co-expression networks and to identify modules of co-expressed genes. The initial dataset contained information on 65217 transcripts including noncoding genes, pseudogenes and non-coding RNAs. Following the recommendations of WGCNA developers, we removed 42385 transcripts with consistently low expression in our sample set (less than 10 reads in 90% of samples). Additionally, genes with relatively low expression levels (last quantile, <22 reads) in samples corresponding to HPE susceptibility period (forebrain, 4 pcw) were removed. As a final step for data-cleaning, we kept only protein-coding genes for further analysis. We then used *varianceStabilizingTransformation* function of DESeq2 package[27] to normalize the expression values and performed samples clustering by WGCNA. Two samples (*ERR1473335* and *ERR1473304*) were removed from further analysis as they were classified as outliers. The final dataset contained normalized expression values of 14459 protein-coding genes across 134 samples. Construction of the Topological Overlap Matrix (TOM), the corresponding coexpression network and module detection was performed by WGCNA (*TOMsimilarityfromExpr* and *blockwiseModules* functions) with the following major parameters: *maximum module size = 8000, networkType = signed, power = 4, mergeCutHeight = 0.25, corType = Pearson*. Co-expression modules were identified using *Dynamic Hybrid tree cut* method implemented in WGCNA package.

Topological Overlap Matrix (TOM) calculates connectivity for each pair of genes which reflects the distance between them in the coexpression network. The connectivity between genes is based on correlation between their expression profiles and their topological similarity. The TOM matrix was used to identify candidate genes sharing highly similar expression profiles with 4 HPE genes (*SHH, ZIC2, SIX3, TGIF1*). More precisely, for each HPE gene, we defined its

top 25% most connected partners (top 25% TOM values) as genes sharing highly similar expression profiles.

Pipeline Results

Whole Exome Sequencing overview

The final analysis included Whole Exome Sequencing data of 80 individuals representing 26 HPE families. Using a filtering approach, we selected a total of 3920 different rare likely damaging variants (non-synonymous, splicing, nonsense or frameshift mutations), with an average of 150 mutational events per family (*Figure S2*). All pre-screened mutations identified by targeted sequencing were validated by WES.

Transcriptome-driven strategy

Our initial hypothesis was that major genes implicated in HPE are highly co-expressed during forebrain development. Therefore, genes sharing highly similar expression patterns with major HPE genes (which, in our case, are *SHH*, *ZIC2*, *SIX3* and *TGIF1*) might be implicated in the same developmental processes and, therefore, be good candidates for HPE. To establish gene transcriptomic signatures, we focused on gene expression in cerebral structures during 4^{th} -10^{th} postconception week (pcw), as this period includes a part of HPE susceptibility stage (4 pcw) and critical processes involved in forebrain development.

Using HDBR transcriptomic data collected by RNA-Seq, we first studied expression profiles of HPE genes in terms of developmental stage and structure. Two HPE genes (*NODAL* and *FOXH1*) could not be analysed during this study as they were excluded during primary filtration (low expression). The similarity between expression profiles of the remaining 14 HPE genes was investigated using hierarchical clustering (*Figure S3*). Overall, the expression of HPE genes is relatively homogeneous during early developmental stages (Carnegie Stage 13-18, corresponding approximately to 4-6 pcw period) but no common transcriptomic signature was identified. However, certain HPE genes shared similar expression patterns which was consistent with biological knowledge. For example, *SHH* was regrouped within the same cluster as *PTCH1* and *SIX3*, reflecting high resemblance of their expression profiles. It has been shown that *PTCH1,* member of Patched family, is involved in intracellular trafficking of *SHH* and acts as its primary receptor[28], while *SIX3* is a transcription factor directly regulating *SHH* during forebrain development.[29]

We then performed a Weighted Co-expression Network Analysis (WGCNA) and identified 14 co-expression modules, labelled by different colours as shown in *Figure S4*. Module sizes ranged from 178 to 3744 (*Figure S4B*). Each module represents a group of co-expressed genes, except for the grey module (*MEgrey*) which contains unassigned genes (*n = 499*). Known HPE genes were regrouped within 6 modules (labelled *red*, *black, pink, green, turquoise* and *brown*). As expected, the distribution of HPE genes among the modules was consistent with the similarity between their expression patterns illustrated in *Figure S3*.
As it was done in previous studies,[30] we investigated the transcriptomic signature of each module by calculating the module eigengenes (MEs, the first principal component of each module) which reflect modules 'average' expression profile (*Figure S4A*). Globally, the 6 modules containing HPE genes are expressed in forebrain during early development (Carnegie Stage 13-18), but no common expression pattern could be detected. It is important to note, however, that studying the expression of a particular gene within a module using its ME profile may lead to inaccurate conclusions due to differences in module sizes and varying correlation between gene expression and MEs (*Figure S4C*).

The final step was to identify candidate genes sharing highly similar expression profiles with major HPE genes. 3 of the 4 major HPE genes (*SHH*, *ZIC2*, *SIX3)* were regrouped within the *red* module (*n = 885*) along with *PTCH1*, while *TGIF1* was regrouped with *GLI2* and *FGFR1* within the *green* module (*n = 1057*). To define candidate genes, we used Topological Overlap Measure (TOM) which reflects the connectivity between each pair of genes within the co-expression network. Using TOM values of *SHH, ZIC2, SIX3* and *TGIF1*, we selected their top 25% most connected partners as candidate genes for HPE. After combining the results and removing redundancy, we established the final list of 547 candidate HPE genes sharing highly similar expression profile with either *SHH, SIX3, ZIC2* or *TGIF1* (Table S5*)*.

Integrated Workflow

The final integrated workflow combining all strategies described above is detailed in *Figure 1* in the main text. Rare damaging variants identified by WES approach (*n=3920*) were restricted to candidate genes obtained by either transcriptomic (*n=547*) or clinical approaches (*n=659*). By applying the integrated workflow to the studied HPE cohort, we identified a total of 232 rare candidate variants across 26 families, detailed in *Table S6*. Among the candidate variants, 153 were in candidate genes identified by clinically-driven strategy (association to a clinical phenotype overlapping with HPE), while 102 were in candidate genes identified by trascriptome-driven strategy (highly similar expression patterns with major HPE genes during cerebral development). 23 variants were identified by both approaches, among which there were mutations in known HPE genes previously revealed by targeted sequencing (*Materials and Methods, Main text*). A functional enrichment analysis using g:profiler[31] was performed to determine pathways significantly enriched in the whole list of resulting candidate variants (*Table S7*).

On average, each family presented ~9 candidate variants. Most of candidate variants were inherited from asymptomatic parents (heterozygous state), suggesting an oligogenic inheritance pattern requiring the association of multiple mutational events. In an attempt to uncover possible disease-causing combinations of mutations underlying HPE, multiple manual analyses of candidate variants were performed on a family-by-family basis. By exploring family pedigree, clinical information, expression studies, variant characteristics, recurrence of mutated genes in our cohort and available biological knowledge, we tried to identify combinations of strong candidate variants, inherited from each parent, in genes connected in a biologically meaningful way and presenting a significant link with holoprosencephaly. The strongest oligogenic combinations are presented in the main findings (*Main text*).

Genes whose knockout (KO) mice mutants exhibit HPE-like phenotype were selected according to Mouse Genome Informatics (MGI) data,[20] and further referred to as KO HPE genes). A total of 32 putative pathogenic variants were found in these KO HPE genes, among which there were variants in known HPE genes *SHH, ZIC2, SIX3, GLI2, TGIF1* and *PTCH1* (*Table S8*).

Case reports

Case reports (1): *FAT1, NDST1, COL2A1*

Besides known HPE genes, *FAT1* was the most represented KO HPE gene, with 4 candidate variants in families F3, F16, F23, F26. Previously associated with various types of cancer[32], *FAT1* is a giant protocadherin (member of the cadherin superfamily), implicated in cell proliferation and migration. Despite its length (more than 139kb) and a large number of reported polymorphisms, rare variants in *FAT1* were recently implicated in several genetic disorders including facioscapulohumeral dystrophy-like disease and glomerulotubular nephropathy.[33-34] Although *FAT1* was identified by our clinically-driven strategy, it also presents an expression pattern that is compatible with the onset of HPE. We performed a study of *FAT1* expression in chicken embryo. It revealed that *FAT1* is expressed at stage 7, in the forming neural plate and future forebrain and that its expression is restricted to the forebrain at stage HH8 (*Figure S5*). Finally, previous studies showed that loss of *FAT1* may lead to HPE or cranial Neural Tube Defects (NTDs) in mice with variable penetrance[35], suggesting that the effect of *FAT1* disruption depends on the genetic background.

The four identified *FAT1* variants were located in exons 2, 10, 18 and 19 without any specific mutational hotspot (*Figure S6*). To further analyze the potential implication of *FAT1* in HPE, all candidate variants in the corresponding families were carefully analyzed by taking into account patients phenotype features, segregation analysis and available biological knowledge of functional relationships between candidate genes.

Family F3

Family F3 comprised 2 parents presenting minor HPE signs (epicanthus and slight hypotelorism), 2 affected fetuses (semilobar HPE) and 1 unaffected child. Additional phenotypic features observed in the affected subjects included proboscis, median cleft lip, abnormal nose morphology, eye defects, big ears and mandibular anomalies. WES was performed on both parents and 1 affected foetus of this family.

A total of 8 candidate variants were identified in the foetus. The strongest candidate was a novel variant in *SHH* (major HPE gene), previously identified by targeted sequencing. SHH signaling is believed to be the primary pathway of HPE[36]. The *SHH* missense variant was located in exon 2 (*c.511G>C*, RefSeq NM_000193) and caused amino acid substitution (*p.D171H*) in Hedgehog N-terminal domain (*Shh-N*), the active part of the protein directly implicated in the hedgehog signaling activity[37]. The *SHH* variant was predicted as damaging by all bioinformatics algorithms. No other mutations were found in known HPE genes in the foetus . *In-silico* results suggest that the *SHH* mutation is very likely implicated in the onset of HPE in this family. However, the variant was inherited from the mother presenting only a minor sign of midline anomaly (epicanthus), suggesting a modifier effect.

Considering the hypothesis of oligogenic inheritance, we suggested that some paternal variants also contribute to HPE phenotype in combination with the maternal *SHH* mutation. Paternal variants in *FAT1* and *NDST1* stood out from the analysis as these genes were also

associated with knockout phenotypes resulting in HPE. *NDST1* is an N-deacetylase sulfotransferase implicated in post-traductional modifications of Heparan Sulfates[38]. Previous studies illustrated that *NDST1* knockout mice mutants exhibited holoprosencephaly with variable penetrance, similarly to *FAT1*$^{-/-}$ mice mutants[39]. Interestingly, *FAT1*$^{-/-}$ and *NDST1*$^{-/-}$ mice exhibited some phenotypic features that are also observed in the foetus (detailed in the *Main text*). There also seems to be a functional relationship between *SHH* and *NDST1*, as *NDST1* knockout mutants exhibit reduced *Shh* signalling in the forebrain and some *Ndst1*$^{+/-}$*; Shh*$^{+/-}$ compound heterozygous mice exhibit craniofacial defects similar to the HPE phenotype[39].

The two variants were predicted deleterious (18 deleterious predictions for *FAT1* and 11 for *NDST1*) but were present in public databases, with maximum population frequency of 0.0001 for *FAT1* (ExAC) and 0.0025 for *NDST1* (ESP). The presence of the variants in public databases may indicate their polymorphic nature but the combined effect of *SHH*, *FAT1* and *NDST1* variants remains unknown.

Complementary Sanger sequencing of the three primary candidates (*SHH*, *FAT1* and *NDST1*) in other members of this family revealed that the second affected foetus also presented the three variants, while the unaffected child inherited only the *FAT1* variant. These observations were consistent with the oligogenic inheritance hypothesis, implicating the association of hypomorphic variants in *SHH*, *FAT1*, *NDST1* in the onset of HPE in this family.

Family F16

Another *FAT1*/*NDST1* combination was found in family F16, comprising 1 child presenting semilobar HPE, 1 child affected with HPE microform and two unaffected parents. Additional phenotypic features included microcephaly, retrognathia, flat nose and absent olfactory bulb, agenesis of corpus callosum. No mutations in known HPE genes were identified in this family.

A total of 14 candidate variants were identified for family F16. Variants in *FAT1*, *NDST1* and *COL2A1* were considered as primary candidates due to their KO HPE phenotype in mice.

COL2A1 codes for procollagen IIA, a cartilage extracellular matrix protein. Expression studies showed that *COL2A1* is a marker of the prechordal plate, the primary structure in the forebrain development[40]. *COL2A1*-null mice display a partially penetrant phenotype including holoprosencephaly and loss of mid-facial structures. Similarly to *NDST1*, *COL2A1* knock out was also associated with reduced *Shh* signalling (prechordal plate). FAT1/*COL2A1* combination also showed recurrence in the cohort (family F26).

Among the remaining candidates in family F16, variants in *CNTN4* (Contactin 4, found in a duplicated region in one case of alobar HPE[41]) and *SHROOM3* (associated to NTD phenotypes similar to those found in case of loss of *FAT1*[42]) retained our attention, but no direct link to HPE was found for these genes. No other relevant information was found for the remaining candidates.

Among the primary candidates, variants in *FAT1* and *NDST1* were inherited from the father, while the *COL2A1* variant was inherited from the mother. Interestingly, foetus carrying all three primary candidates (*FAT1*, *NDST1* and *COL2A1)* was affected by semilobar HPE, while

the child affected with HPE microform lacked the *FAT1* variant. Similarly to the previously described family F3, the foetus presented phenotypic overlap with *NDST1*- and *COL2A1*-null mutant mice (HPE, mandibular anomalies, absent olfactory bulb, abnormal nose morphology). Taken together, these observations illustrate a possible oligogenic inheritance pattern for HPE, requiring several mutational events (*FAT1*, *NDST1, COL2A1*) for the onset of the disease.

Family F23

Family F23 comprised two unaffected parents and one foetus affected with alobar HPE. No mutations in major HPE genes were identified by Targeted/WES sequencing. Among the candidates, variants in *FAT1* and *LRP2* were inherited from 2 different parents and were considered as primary candidates. *LRP2* is an auxiliary receptor of *SHH,* which controls cellular trafficking of Shh/Ptch1 comlpexes and has a critical role in SHH signaling during forebrain development[43]. Disruption of *LRP2* leads to HPE phenotype in mouse models[43]. Another candidate *LRP2* variant was also identified in family *F11* (See Case report 3). Secondary candidates included a novel variant in *TBX15* (transcription factor associated with various craniofacial features[44]) and a variant in *COL4A1* (Alpha 1 chain of collagen IV, implicated in Schizencephaly[45], which may be a part of the larger phenotypic spectrum of HPE).

Family F26

Last FAT1 variant was found in family F26 (2 probands diagnosed with alobar HPE). This variant was classified as novel (not found in any public database) and was inherited from a mother presenting a minor sign of HPE (slightly narrow palate). Additional candidates inherited from the unaffected father included a private variant in *PTCH1* (known HPE gene) and a subpolymorphic variant in previously mentioned *COL2A1*.

Overall these observations further highlight the possible oligogenic inheritance pattern of HPE requiring the association of several mutational events for the onset of the disease.
Taken together, our results propose new candidate genes for holoprosencephaly, among which the recurrently mutated *FAT1*, *NDST1*, *COL2A1* and *LRP2* stand out as strong candidates with clinical and functional evidence.

Case reports (2): *SCUBE2/BOC*

In family F22, we identified an interesting combination of candidate variants in *HIC1*, *BOC,* and *SCUBE2*. The three disrupted genes present a functional link with Sonic Hedgehog (SHH) signaling. *HIC1* codes for Hypermethylated in cancer 1, a sequence-specific transcriptional repressor, involved both in development and tumor growth[46]. Studies of medulloblastoma in mice demonstrated that *HIC1* cooperates with *PTCH1*, the primary receptor of *SHH*[47]. Mice deficient in *HIC1* also exhibit multiple craniofacial defects including HPE[48]. *SCUBE2* and *BOC* are directly implicated in the Sonic Hedgehog signaling mechanism. The vertebrate-specific

SCUBE2 is implicated in the release of SHH from the secreting cell surface while BOC (*Brother of CDON*) acts as its co-receptor, promoting HH signaling.[37] *BOC* was shown to act as a silent HPE modifier gene in human[51], and a possible implication of Scube2 in HPE was previously discussed[52]. Two variants were inherited from two different parents. The association of private variants (frequency < 0.00005) in these genes suggests a possible defect in Shh signaling that could lead to HPE. The intriguing fact is that the *SCUBE2* variant results in a premature stop codon at position 525, thus producing a truncated protein. It has been shown that *SHH*-releasing activity of Scube2 strictly depended on the CUB domain[53], located after the position 525. Therefore, the identified candidate variant can directly affect the activity of *SCUBE2* in the SHH signaling pathway. The BOC variant was located at position 311, corresponding to Ig domains which are believed to be dispensable to HH pathway activity[54], however the exact impact of this variant remains unknown. Oligogenic combination *SCUBE2*/*BOC* showed recurrence as another combination of candidate variants in these genes was identified in family F4. The family F4 also presented a mutation in *SHH* and a private variant in *WNT4*, a member of Wnt pathway, implicated in the regulation of Shh signal transduction.[55]

Overall, these results suggest that the accumulation of rare variants in genes involved in Shh signaling (*SCUBE2, BOC*) may lead to the onset of HPE in these families.

Case reports (3): *Ciliary genes and HPE*

Ciliary proteins were shown to play an important role in several developmental pathways, including the transduction of SHH signal downstream of PTCH1 during forebrain development.[56] Mutations in many ciliary proteins lead to various craniofacial defects including HPE[59,66,67,75]. Moreover, cilia-related disorders, such as Meckel-Gruber syndrome (MKS) or *Joubert syndrome (JS)*, present a wide clinical spectrum of developmental anomalies which include HPE and other anomalies of the prosencephalon encountered in HPE subjects[57]. The functional implication of the primary cilium in the SHH pathway and clinical resemblances between ciliopathies and HPE suggest a possible genetic overlap between these disorders. Therefore, rare variants in ciliary genes observed in HPE subjects might partly explain the onset of the disease.

Further analyses revealed several families presenting candidate variants in ciliary genes. Among these genes, *B9D1* (family F21), *IFT172* (family F18) and *MKS1* (family F11) have a mouse knockout phenotype resulting in HPE.[59,67,73]

Family F21 (alobar HPE) presented a private candidate variant in *B9D1*. Previously implicated in ciliopathies (MKS and JS[58-59]), *B9D1* encodes a B9 domain-containing protein implicated in cilia function. Disruption of this gene is associated with severe developmental anomalies concomitant with compromised ciliogenesis and Hh signal transduction. *B9D1* mice mutants also display HPE.[59] The B9D1 variant was inherited from the mother presenting a minor sign of HPE (incomplete iris), reminiscent of coloboma, frequently encountered in MKS and JS patients. The second candidate was a private variant in *CELSR1* (paternal inheritance), a core gene of Planar Cell Polarity (PCP) pathway. Recent studies have implicated Sonic Hedgehog

signaling and PCP pathways in primary cilia as involved in the patterning of neural tube. Disruptions of core PCP genes (*CELSR1*, *VANGL1, VANGL2, DVL2, FZD6*) lead to adverse effect on neurulation and were associated with severe NTDs (craniorachischisis)[60], also observed in different ciliopathies. Although the investigations on the role of cilia in the PCP pathway are conflicting[61,62], recent evidence suggests their common functions during brain development[63]. Interestingly, the previously discussed *FAT1* was also shown to regulate the PCP pathway in Drosophila[32], and its loss may also lead to cranial NTDs such as exencephaly.[64]

Family F18 comprised 2 unaffected parents and one foetus affected with alobar HPE. Additional clinical features included hypotelorism, eye defects (microphtalmia), cleft palate and polydactyly. The affected foetus presented a private variant (paternal inheritance) in *IFT172*, coding for a core component of intraflagellar transport (IFT) complex. The IFT protein complexes, IFT-A and IFT-B, are required for the construction of primary cilium and, therefore, play integral roles in the regulation of Shh signal transduction[65, 66]. Moreover, *IFT172* mutant mice exhibit severe craniofacial malformations including HPE associated with reduced expression of Shh in the ventral forebrain[67]. These results highlight the crucial role of the primary cilium and, more specifically, *IFT172*, in the forebrain patterning. Secondary candidates included a variant in *PRICKLE1*, which is believed to be a core PCP protein and essential for limb development[68]. Recent studies have linked *PRICKLE1* with defects in primary cilia[69] and the onset of human cleft palate[70]. *PRICKLE1* also showed recurrence (family F8).
IFT172 and *PRICKLE1* were shown to interact genetically in the kidney cyst formation[71], but no such interactions were reported in the forebrain. However, the affected foetus presented cleft palate (associated to PRICKLE1 knockout) and some phenotypic features also found in *IFT172* mutant mice (polydactyly, eye defects). Additionally, a candidate variant in *ABCB6* (ATP-binding cassette transporter associated with ocular coloboma[72]) may partly explain the microphtalmia observed in the affected foetus. Given the absence of mutations in known HPE genes and the presented functional evidence, the association of these variants may explain the onset of HPE in this family.
Family F11 comprised unaffected mother, father presenting a minor form of HPE (solitary median incisor) and a child affected by a HPE microform (microphtalmia and cleft). The father and the affected child presented a rare deleterious mutation in *SHH* identified by targeted sequencing. The affected child presented a combination of rare variants in previously mentioined *SHH, LRP2, CELSR1* and *MKS1*. *MKS1* (implicated in ciliogenesis) was previously associated with Meckel syndrome and *MKS1*$^{-/-}$ mice display holoprosencephaly[73]. Additionally, a secondary candidate variant was identified in *CCP110* (centrosomal protein 110), implicated in ciliogenesis and Shh signaling[74]. It is important to note that other members of this family (not included in this study) also harbored the SHH mutation but presented different HPE-related phenotypes which strongly suggests the implication of other genetic factors that modulate the phenotypic consequences of the SHH mutation.
The affected individual of family F20 presented a heterozygous deletion of SIX2 and *SIX3* inherited from the unaffected father, which was identified by CGH array. Identified candidates

included private variants in ciliary genes *TCTN3* and *TULP3. TCTN3* is a member of transition zone complex (along with previously mentioned *MKS1* and *B9D1)*, implicated in the regulation of ciliary assembly and trafficking[75]. Mutations in *TCTN3* were found in patients affected by Joubert Syndrome and orofaciodigital syndrome (OFD4). Moreover, *TCTN3* was shown to be necessary for the transduction of SHH signal[76]. This is also the case of *TULP3* (tubby-like protein 3), which was shown to be a critical repressor of hedgehog signaling in mice[77].
Tulp3$^{-/-}$ embryos exhibited increased *SHH* signaling in the caudal neural tube and various craniofacial defects such as exencephaly, spina bifida and facial clefting.

Overall, these results present candidate variants found in ciliary genes in HPE subjects. Several genes (*MKS1, B9D1, IFT172*) were already associated to HPE phenotype in mice, while others (*CELSR1*, *PRICKLE1, TCTN3…*) present a functional link with Shh signaling or associated to phenotypic features encountered in HPE subjects. Taken together, these observations illustrate that rare deleterious variants in ciliary genes may explain a substantial part of genetic cases of HPE.

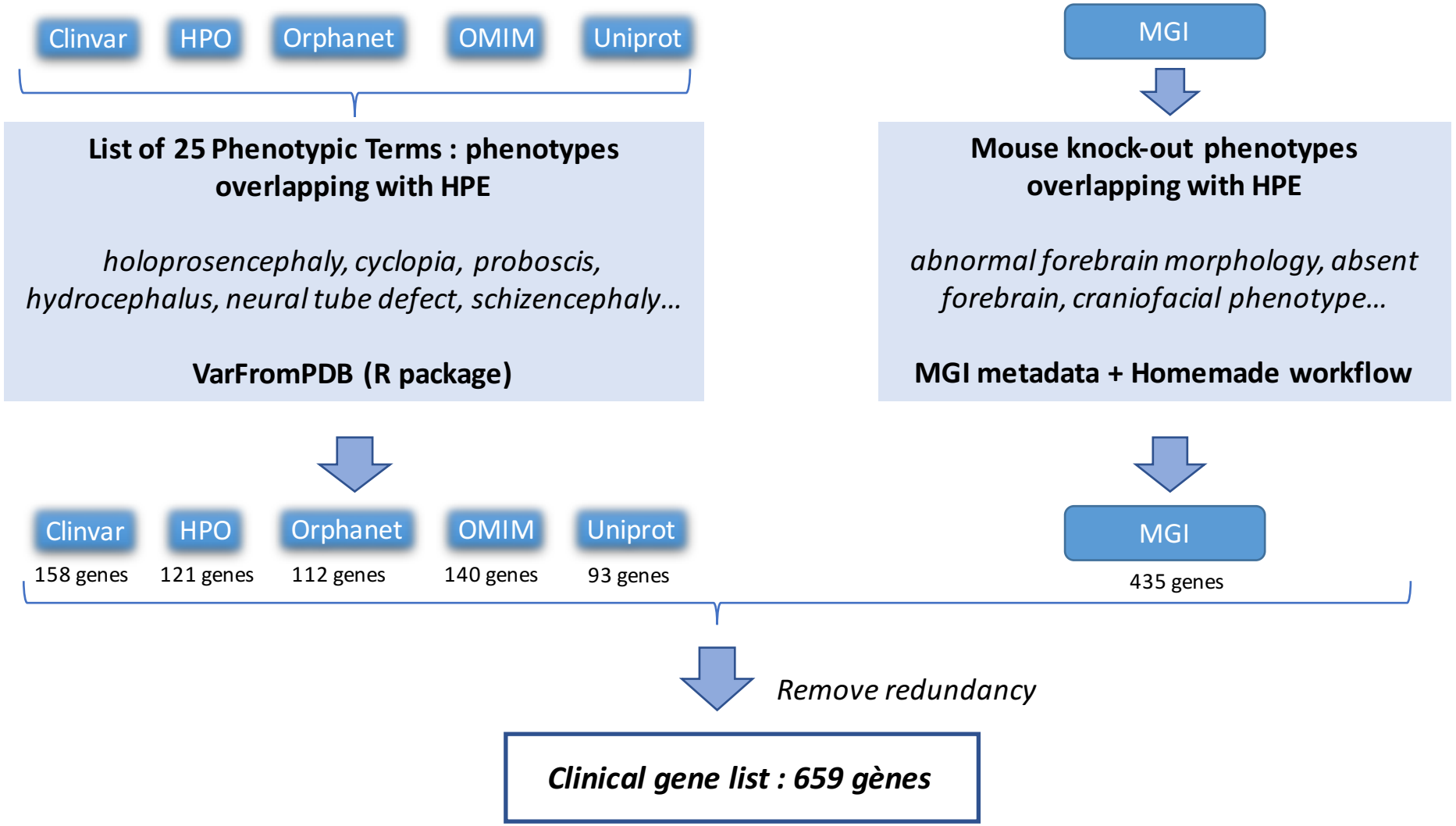


Figure S1. Schematic representation of the clinically-driven strategy.

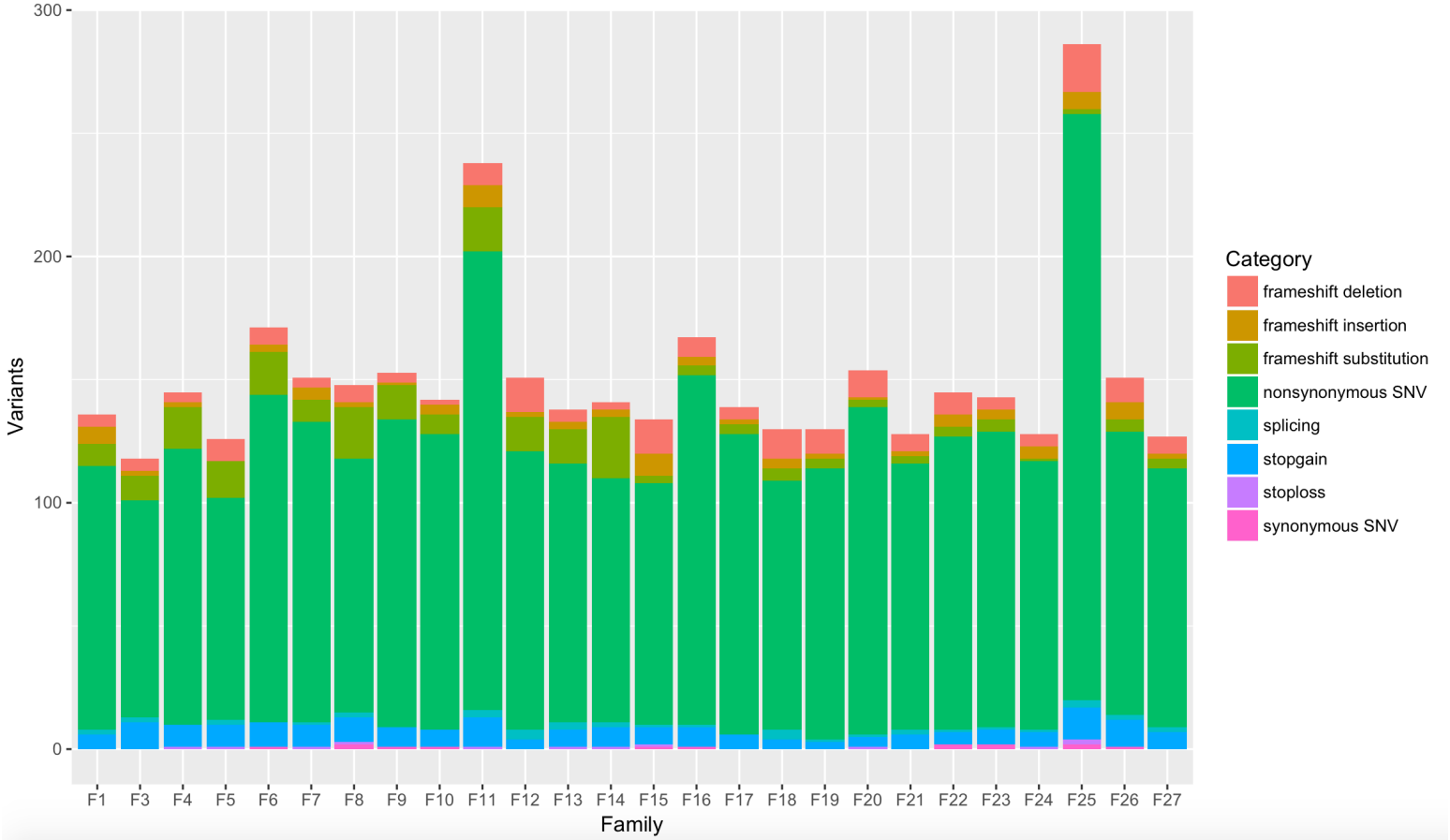


Figure S2. Distribution of the variants among different functional categories.

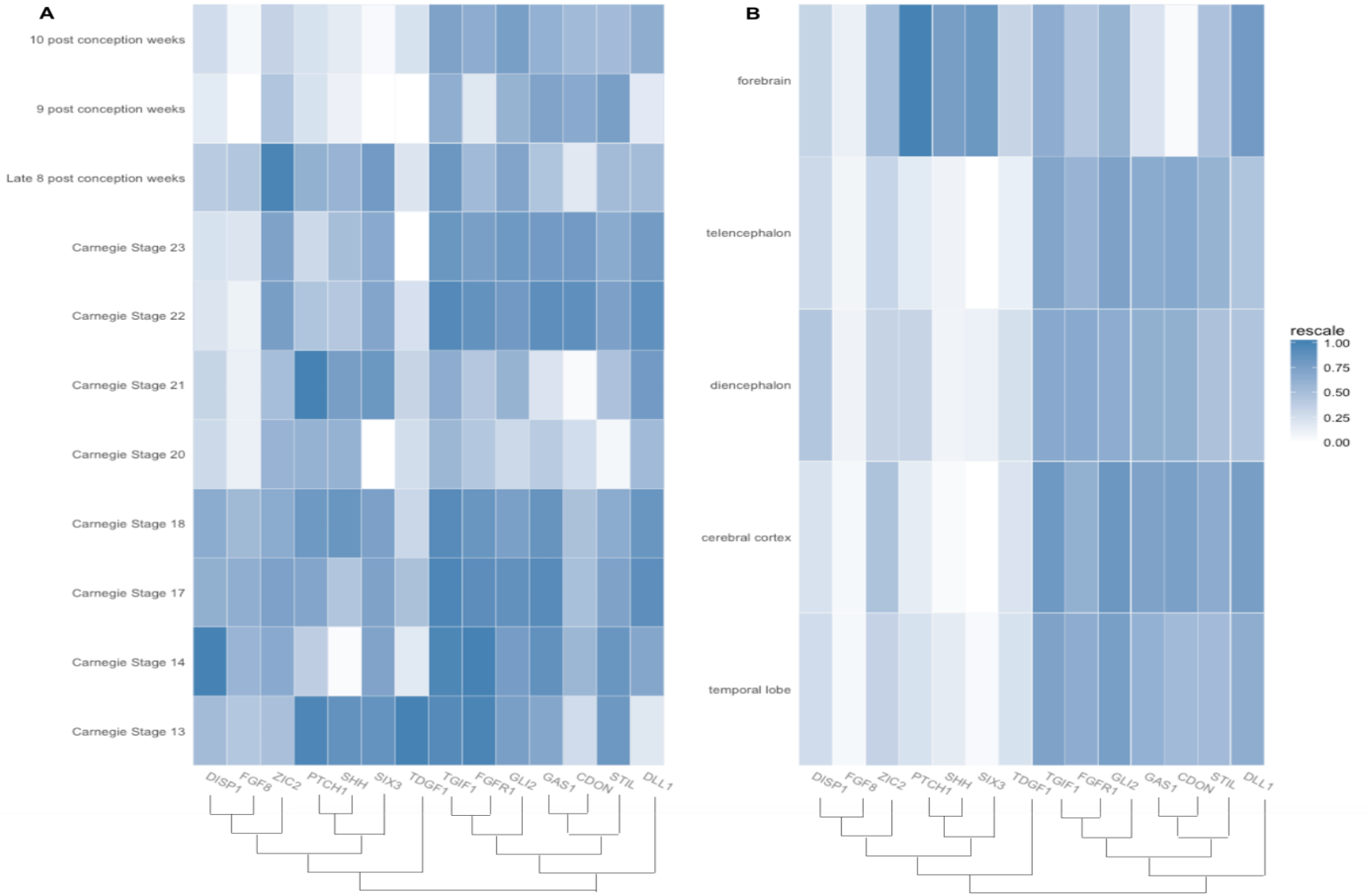


Figure S3. Hierarchical clustering and expression patterns of known HPE genes.

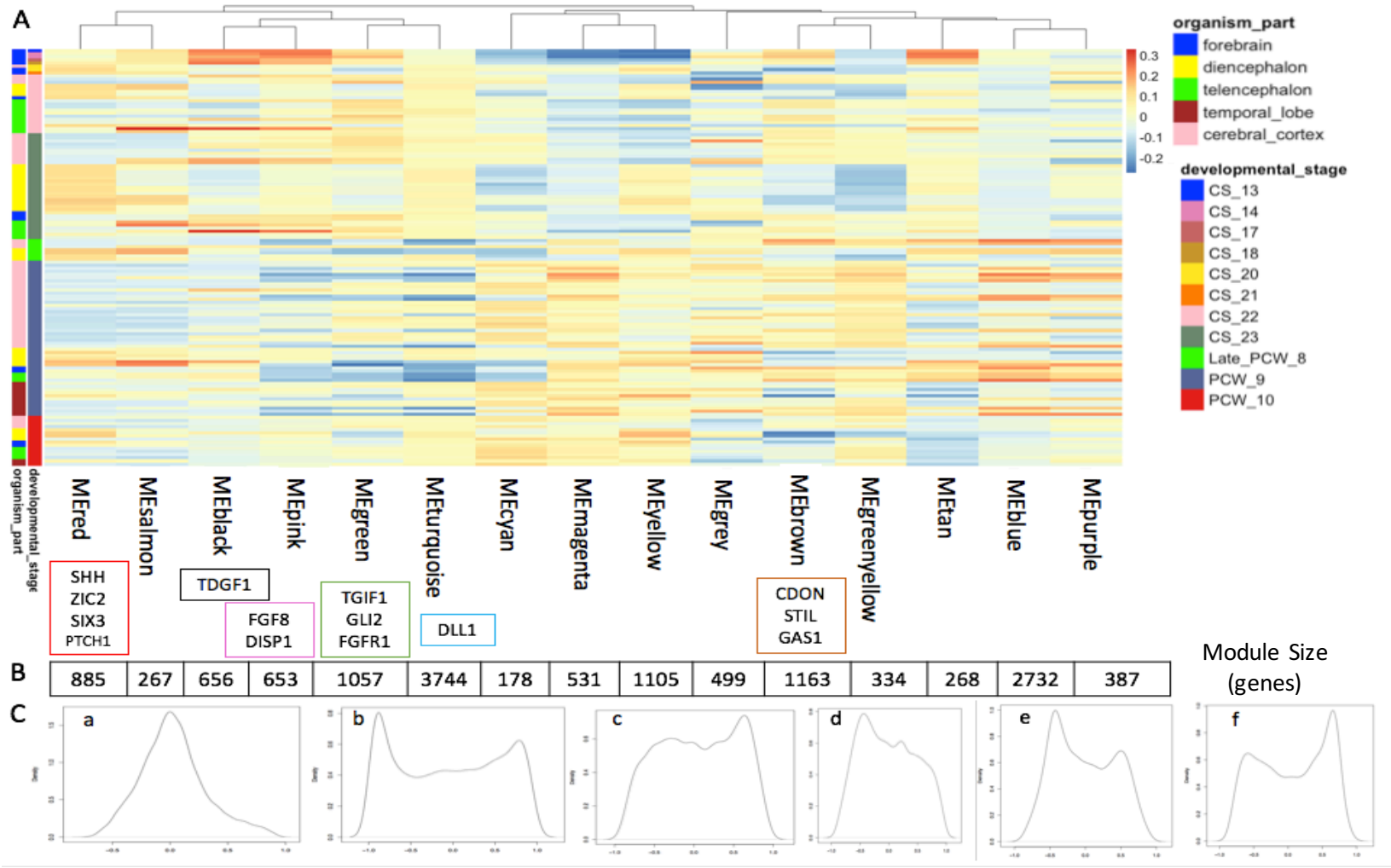
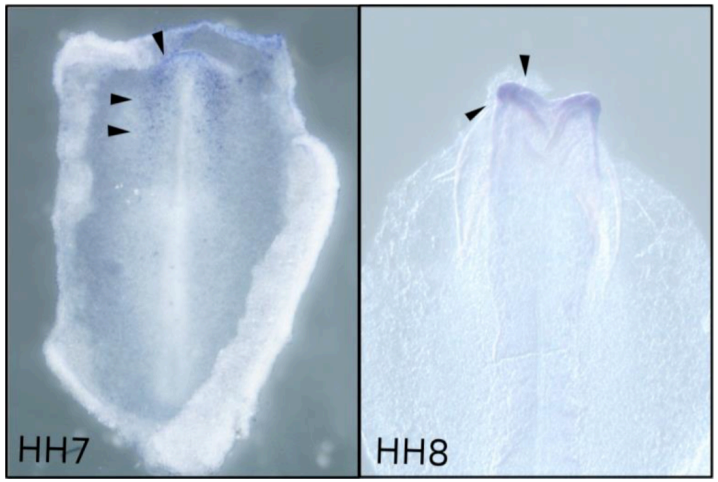


Figure S4. Co-expression modules identified by WGCNA analysis

Figure S5. Expression pattern of *FAT1* in chick embryo (*Gallus gallus*)



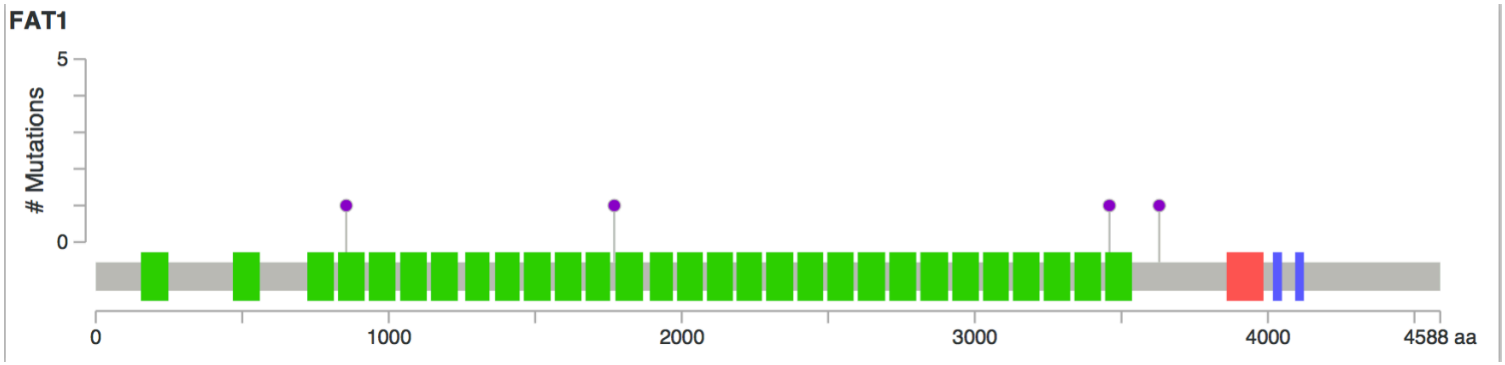


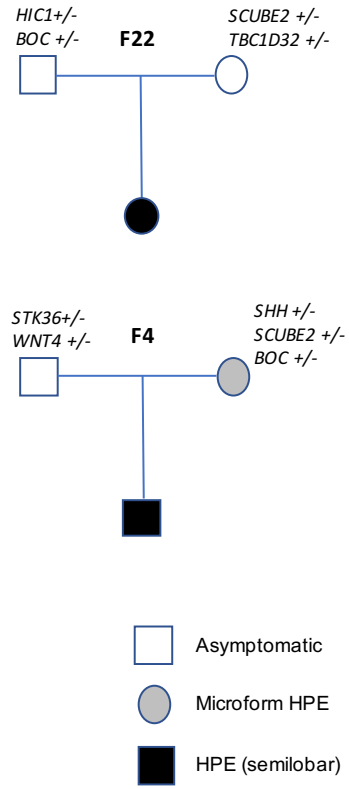
Figure S6. 4 candidate variants identified in *FAT1*

The illustration was made with Mutationmapper online software (http://www.cbioportal.org/mutation_mapper.jsp)

A

Family	Gene	Mutation	Inheritance	Frequency	Candidate	Nature, deleteriousness
F22	*HIC1*	p.Trp511Cys	Paternal	*NA*	Clinical	missense, D:12
	BOC	p.Ala311Val	Paternal	0,00001897	HPE gene	missense, D:10
	SCUBE2	p.Arg525*	Maternal	0,00004061	Transcriptome	stopgain
F4	*SHH*	p.Phe241Val	Maternal	NA	Major HPE gene	missense, D:15
	SCUBE2	p.Thr285Met	Maternal	0,00815	Transcriptome	missense, D:15
	BOC	p.Val861Ile	Maternal	0,0002	HPE gene	missense, D:13
	STK36	p.Arg497Gly	Paternal	0,0025	Clinical	missense, D:11
	WNT4	p.Val204Met	Paternal	0,000004148	Transcriptome	missense, D:11

B



C

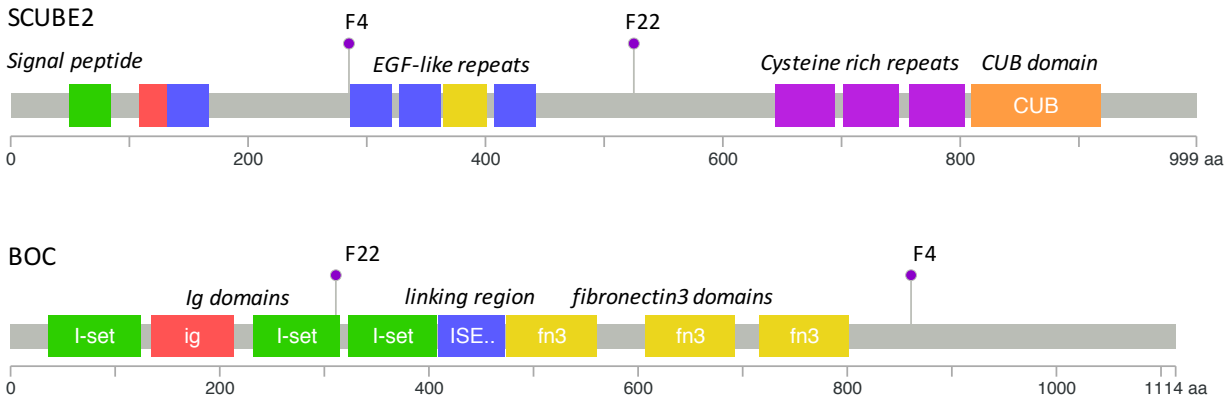
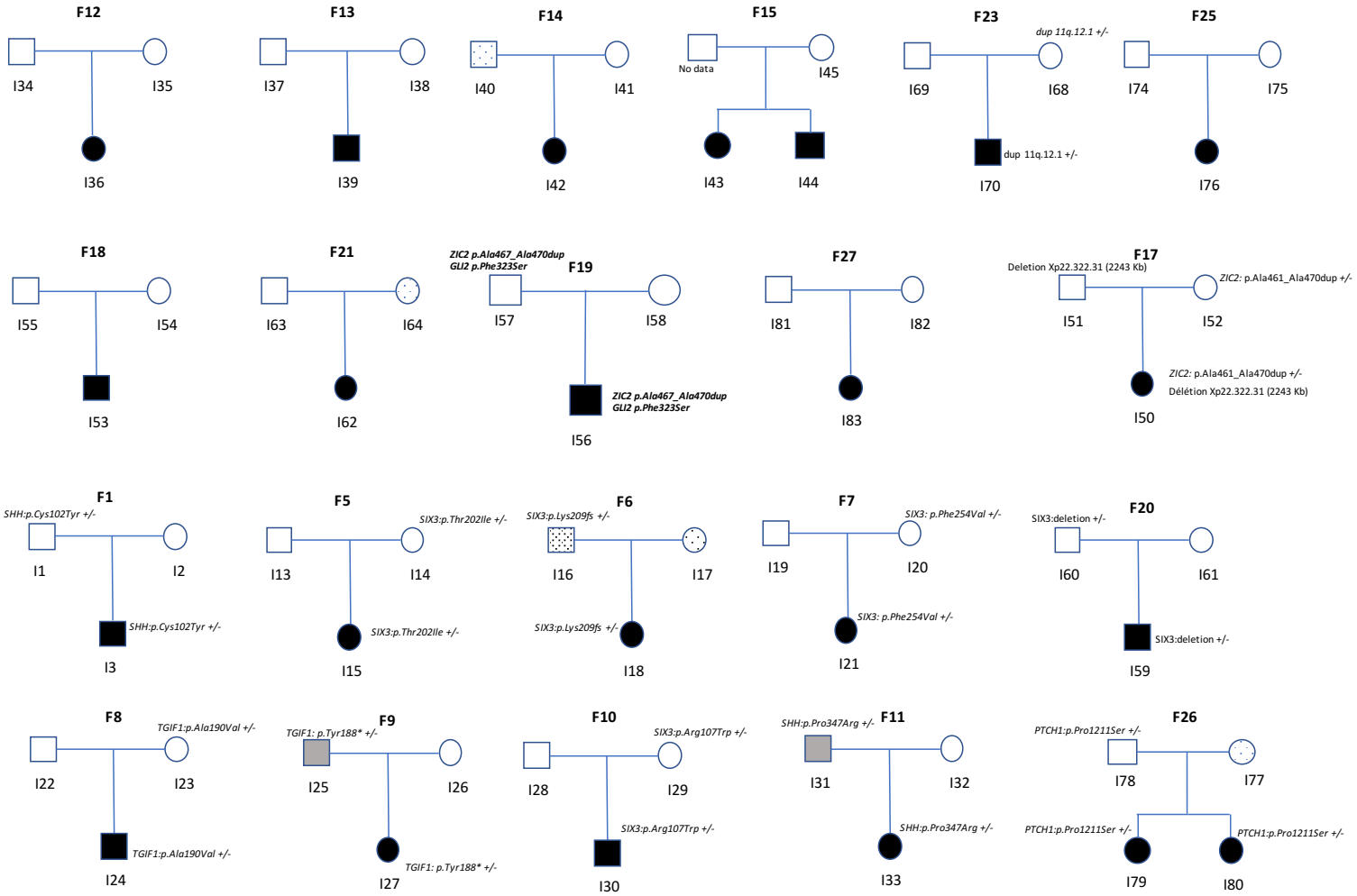


Figure S7. Families with SCUBE2/BOC variants

Figure S8. Additional pedigrees of the studied families.



Black symbols indicate individuals affected with HPE. Grey symbols indicate a parent/sibling affected with microfrom HPE. Dotted symbols indicate a parent presenting a minor sign of HPE. All alterations detected by targeted sequencing/CGH array are marked.

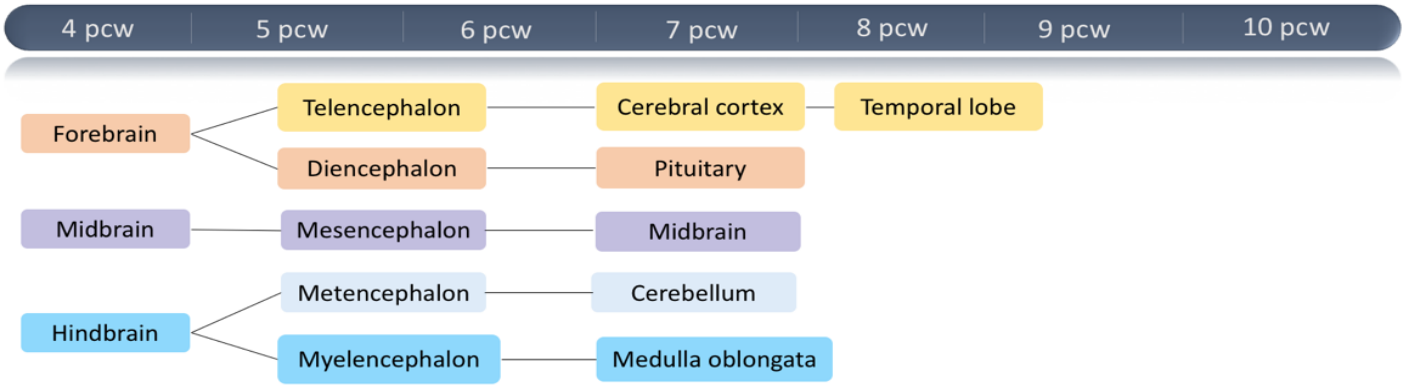


Figure S9. Sample selection from Human Developmental Biology Resource (HDBR).

A total of 136 brain samples from HDBR project were selected for this study.

All samples and detailes protocols are available at https://www.ebi.ac.uk/gxa/experiments/E-MTAB-4840

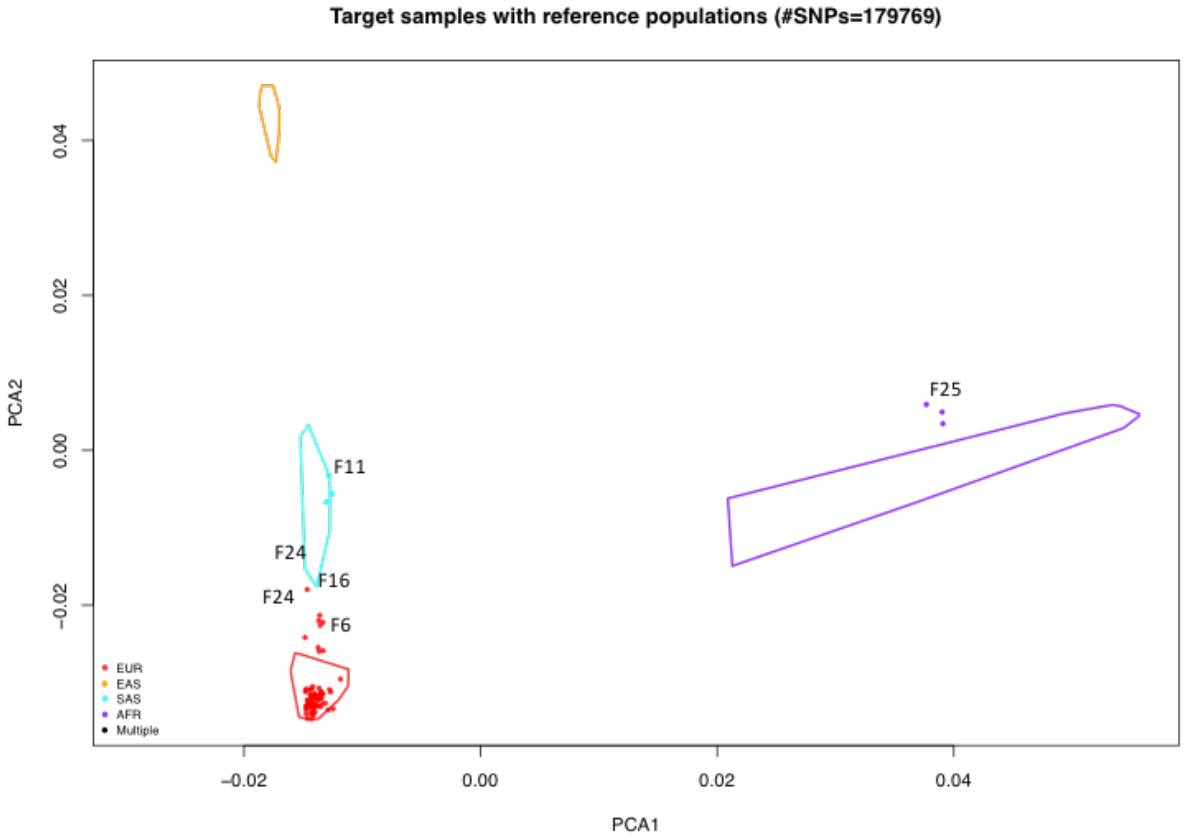


Figure S10. Ethnicity annotation of HPE families

The Principal Component Analysis (PCA) was performed by EthSeq.[78]

Supplementary References

1. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43:11.10.1-33.

2. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv:12073907 [q-bio]*. July 2012.

3. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.

4. O'Leary NA, Wright MW, Brister JR, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44(Database issue):D733-D745.

5. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29(1):308-311.

6. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.

7. Glusman G, Caballero J, Mauldin DE, Hood L, Roach JC. Kaviar: an accessible system for testing SNV novelty. *Bioinformatics*. 2011;27(22):3216-3217.

8. Exome Variant Server. http://evs.gs.washington.edu/EVS/ (2011).

9. Scott EM, Halees A, Itan Y, et al. Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat Genet*. 2016;48(9):1071-1076.

10. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-291.

11. 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. *Hum Mutat*. 2016;37(3):235-241.

12. Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res*. 2014;42(22):13534-13544.

13. Desmet F-O, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 2009;37(9):e67.

14. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinformatics*. 2013;14(2):178-192.

15. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 2016;44(D1):D862-868.

16. Weinreich SS, Mangon R, Sikkens JJ, Teeuw ME en, Cornel MC. [Orphanet: a European database for rare diseases]. Ned Tijdschr Geneeskd. 2008;152(9):518-519.

17. Apweiler R, Bairoch A, Wu CH, et al. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res*. 2004;32(Database issue):D115-119.

18. 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(Database issue):D789-798.

19. Köhler S, Vasilevsky NA, Engelstad M, et al. The Human Phenotype Ontology in 2017. *Nucleic Acids Res*. 2017;45(D1):D865-D876.

20. Blake JA, Richardson JE, Bult CJ, Kadin JA, Eppig JT. MGD: the Mouse Genome Database. *Nucleic Acids Res*. 2003;31(1):193-195.

21. Karlstrom RO, Talbot WS, Schier AF. Comparative synteny cloning of zebrafish you-too: mutations in the Hedgehog target gli2 affect ventral forebrain patterning. *Genes Dev*. 1999;13(4):388-393.

22. Lindsay SJ, Xu Y, Lisgo SN, et al. HDBR Expression: A Unique Resource for Global and Individual Gene Expression Studies during Early Human Brain Development. *Front Neuroanat*. 2016;10:86.

23. Fonseca NA, Petryszak R, Marioni J, Brazma A. iRAP - an integrated RNA-seq Analysis Pipeline. *bioRxiv*. June 2014:005991.

24. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;25(9):1105-1111.

25. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166-169.

26. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559.

27. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014;15:550.

28. Cohen M, Kicheva A, Ribeiro A, et al. Ptch1 and Gli regulate Shh signalling dynamics via multiple mechanisms. *Nature Communications*. 2015;6:ncomms7709.

29. Jeong Y, Leskow FC, El-Jaick K, et al. Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nat Genet*. 2008;40(11):1348-1353.

30. Parikshak NN, Luo R, Zhang A, et al. Integrative Functional Genomic Analyses Implicate Specific Molecular Pathways and Circuits in Autism. *Cell*. 2013;155(5):1008-1021.

31. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. g:Profiler--a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res*. 2007;35(Web Server issue):W193-200.

32. Katoh M. Function and cancer genomics of FAT family genes (review). *Int J Oncol*. 2012;41(6):1913-1918.

33. Gee HY, Sadowski CE, Aggarwal PK, et al. FAT1 mutations cause a glomerulotubular nephropathy. *Nat Commun*. 2016;7.

34. Puppo F, Dionnet E, Gaillard M-C, et al. Identification of Variants in the 4q35 Gene FAT1 in Patients with a Facioscapulohumeral Dystrophy-Like Phenotype. *Human Mutation*. 2015;36(4):443-453.

35. Ciani L, Patel A, Allen ND, ffrench-Constant C. Mice lacking the giant protocadherin mFAT1 exhibit renal slit junction abnormalities and a partially penetrant cyclopia and anophthalmia phenotype. *Mol Cell Biol*. 2003;23(10):3575-3582.

36. Kauvar EF, Muenke M. Holoprosencephaly: recommendations for diagnosis and management. *Curr Opin Pediatr*. 2010;22(6):687-695.

37. Briscoe J, Thérond PP. The mechanisms of Hedgehog signalling and its roles in development and disease. *Nature Reviews Molecular Cell Biology*. 2013;14(7):nrm3598.

38. Pallerla SR, Pan Y, Zhang X, Esko JD, Grobe K. Heparan sulfate Ndst1 gene function variably regulates multiple signaling pathways during mouse development. *Dev Dyn*. 2007;236(2):556-563.

39. 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(16):3777-3786.

40. 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(9):2319-2329.

41. Chen C-P, Huang M-C, Chern S-R, et al. Distal 3p duplication and terminal 7q deletion associated with nuchal edema and cyclopia in a fetus and a review of the literature. *Taiwan J Obstet Gynecol*. 2015;54(3):297-302.

42. Hildebrand JD. Shroom regulates epithelial cell shape via the apical positioning of an actomyosin network. *J Cell Sci*. 2005;118(Pt 22):5191-5203.

43. Christ A, Christa A, Kur E, et al. LRP2 is an auxiliary SHH receptor required to condition the forebrain ventral midline for inductive signals. *Dev Cell*. 2012;22(2):268-278.

44. Lausch E, Hermanns P, Farin H, et al. TBX15 Mutations Cause Craniofacial Dysmorphism, Hypoplasia of Scapula and Pelvis, and Short Stature in Cousin Syndrome. *American journal of human genetics*. 2008;83:649-55.

45. Yoneda Y, Haginoya K, Kato M, et al. Phenotypic spectrum of COL4A1 mutations: porencephaly to schizencephaly. *Ann Neurol*. 2013;73(1):48-57.

46. Zhang W, Zeng X, Briggs KJ, et al. A potential tumor suppressor role for Hic1 in breast cancer through transcriptional repression of ephrin-A1. *Oncogene*. 2010;29(17):2467-2476.

47. Briggs KJ, Corcoran-Schwartz IM, Zhang W, et al. Cooperation between the Hic1 and Ptch1 tumor suppressors in medulloblastoma. *Genes Dev*. 2008;22(6):770-785.

48. Carter MG, Johns MA, Zeng X, et al. 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(3):413-419.

49. Ko HW, Norman RX, Tran J, Fuller KP, Fukuda M, Eggenschwiler JT. Broad-minded links cell cycle-related kinase to cilia assembly and hedgehog signal transduction. *Dev Cell*. 2010;18(2):237-247.

50. Adly N, Alhashem A, Ammari A, Alkuraya FS. Ciliary genes TBC1D32/C6orf170 and SCLT1 are mutated in patients with OFD type IX. *Hum Mutat*. 2014;35(1):36-40.

51. Hong M, Srivastava K, Kim S, et al. BOC is a modifier gene in holoprosencephaly. *Hum Mutat*. 2017;38(11):1464-1470.

52. Tukachinsky H, Kuzmickas RP, Jao CY, Liu J, Salic A. Dispatched and Scube mediate the efficient secretion of the cholesterol-modified Hedgehog ligand. *Cell Rep*. 2012;2(2):308-320.

53. Jakobs P, Exner S, Schürmann S, et al. Scube2 enhances proteolytic Shh processing from the surface of Shh-producing cells. *J Cell Sci*. 2014;127(Pt 8):1726-1737.

54. Song JY, Holtz AM, Pinskey JM, Allen BL. Distinct structural requirements for CDON and BOC in the promotion of Hedgehog signaling. *Dev Biol*. 2015;402(2):239-252.

55. Borycki A, Brown AM, Emerson CP. Shh and Wnt signaling pathways converge to control Gli gene activation in avian somites. *Development*. 2000;127(10):2075-2087.

56. Goetz SC, Anderson KV. The Primary Cilium: A Signaling Center During Vertebrate Development. *Nat Rev Genet*. 2010;11(5):331-344.

57. Abdelhamed ZA, Wheway G, Szymanska K, et al. Variable expressivity of ciliopathy neurological phenotypes that encompass Meckel–Gruber syndrome and Joubert syndrome is caused by complex de-regulated ciliogenesis, Shh and Wnt signalling defects. *Hum Mol Genet*. 2013;22(7):1358-1372.

58. Romani M, Micalizzi A, Kraoua I, et al. Mutations in B9D1 and MKS1 cause mild Joubert syndrome: expanding the genetic overlap with the lethal ciliopathy Meckel syndrome. *Orphanet J Rare Dis*. 2014;9:72.

59. Dowdle WE, Robinson JF, Kneist A, et al. Disruption of a Ciliary B9 Protein Complex Causes Meckel Syndrome. *Am J Hum Genet*. 2011;89(1):94-110.

60. Tatin F, Taddei A, Weston A, et al. Planar Cell Polarity Protein Celsr1 Regulates Endothelial Adherens Junctions and Directed Cell Rearrangements during Valve Morphogenesis. *Dev Cell*. 2013;26(1):31-44.

61. Huang P, Schier AF. Dampened Hedgehog signaling but normal Wnt signaling in zebrafish without cilia. *Development*. 2009;136(18):3089-3098.

62. Simons M, Gloy J, Ganner A, et al. Inversin, the gene product mutated in nephronophthisis type II, functions as a molecular switch between Wnt signaling pathways. *Nat Genet*. 2005;37(5):537-543.

63. W Vogel T, Carter C, Abode-Iyamah K, Zhang Q, Robinson S. The role of primary cilia in the pathophysiology of neural tube defects. *Neurosurgical focus*. 2012;33:E2.

64. Badouel C, Zander MA, Liscio N, et al. Fat1 interacts with Fat4 to regulate neural tube closure, neural progenitor proliferation and apical constriction during mouse brain development. *Development*. 2015;142(16):2781-2791.

65. Liem KF, Ashe A, He M, et al. The IFT-A complex regulates Shh signaling through cilia structure and membrane protein trafficking. *J Cell Biol*. 2012;197(6):789-800.

66. Duran I, Taylor SP, Zhang W, et al. Destabilization of the IFT-B cilia core complex due to mutations in IFT81 causes a Spectrum of Short-Rib Polydactyly Syndrome. *Scientific Reports*. 2016;6:srep34232.

67. Gorivodsky M, Mukhopadhyay M, Wilsch-Braeuninger M, 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(1):24-32.

68. Yang T, Jia Z, Bryant-Pike W, 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(2):138-151.

69. Gibbs BC, Damerla RR, Vladar EK, et al. Prickle1 mutation causes planar cell polarity and directional cell migration defects associated with cardiac outflow tract anomalies and other structural birth defects. *Biology Open*. February 2016:bio.015750.

70. Yang T, Jia Z, Bryant-Pike W, 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(2):138-151.

71. Cao Y, Park A, Sun Z. Intraflagellar Transport Proteins Are Essential for Cilia Formation and for Planar Cell Polarity. *J Am Soc Nephrol*. 2010;21(8):1326-1333.

72. Wang L, Wang L, He F, et al. ABCB6 mutations cause ocular coloboma. *Am J Hum Genet*. 2012;90(1):40-48.

73.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. *Developmental Biology*. 2013;377(1):55-66.

74. Yadav SP, Sharma NK, Liu C, Dong L, Li T, Swaroop A. Centrosomal protein CP110 controls maturation of the mother centriole during cilia biogenesis. *Development*. 2016;143(9):1491-1501.

75. Garcia-Gonzalo FR, Corbit KC, Sirerol-Piquer MS, et al. A transition zone complex regulates mammalian ciliogenesis and ciliary membrane composition. *Nat Genet*. 2011;43(8):776-784.

76. Thomas S, Legendre M, Saunier S, et al. TCTN3 mutations cause Mohr-Majewski syndrome. *Am J Hum Genet*. 2012;91(2):372-378.

77. Cameron DA, Pennimpede T, Petkovich M. Tulp3 is a critical repressor of mouse hedgehog signaling. *Dev Dyn*. 2009;238(5):1140-1149.

78. Romanel A, Zhang T, Elemento O, Demichelis F. EthSEQ: ethnicity annotation from whole exome sequencing data. *Bioinformatics*. 2017;33(15):2402-2404.