

Genetics of Congenitally Corrected Transposition of the Great Arteries: next generation sequencing shows a mutation load effect for 156 genes involved in cardiac patterning

Yasmine Benadjaoud^{1,2}, Sabina Benko¹, Fabienne Jabot Hanin³, Enrique Audain⁴, Jean François Deleuze⁵, Anne Boland⁵, Christine Bole³, Marc Bras³, Fanny Bajolle², Morgan Derouk¹, Christopher T. Gordon¹, Marc Phillip Hitz⁴, Damien Bonnet^{2*}, Stanislas Lyonnet^{1*}

1 Laboratory of Embryology and Genetics of Malformations, INSERM UMR 1163, Paris Cite University, Imagine Institute, Paris, France

2 Pediatric Cardiology Department, Reference Centre M3C, Necker-Enfants Malades Hospital, Paris Cite University, Paris, France

3 Bioinformatics platform, INSERM UMR 1163, Paris Cite University, Imagine Institute, Paris, France

4 Department of Congenital Heart Disease and Pediatric Cardiology, Kiel University, Kiel, Germany

5 Centre National de Génotypage (CNG), Evry, Paris, France

* Equal contributed authors

Corresponding author:

Stanislas Lyonnet

INSERM U1163, 24 bd du Montparnasse, 75015 Paris France

stanislas.lyonnet@inserm.fr

Abstract

Congenital heart disease (CHD) is a major public health issue. It is considered as a major cause of infant mortality today.

We focused on a complex and rare CHD, the congenitally corrected transposition of the great arteries (CCTGA).

The aim of this study was to identify the genes that control the alignment of cardiac chambers that are abnormal in CCTGA to further understand the mechanism of this CHD.

We analyzed a cohort of 43 CCTGA cases (41 sporadic cases and 2 familial cases) of isolated CCTGA by next generation sequencing analysis (whole-genome sequencing and whole-exome sequencing).

Under the hypothesis of Mendelian model including *de novo* genomic alterations, no major gene effect involved in the disease could be identified.

Under the hypothesis of a complex model of inheritance, we highlighted a group of 156 cardiac mutated genes in our patients, in which we found a significant enrichment of rare variants in patients compared to controls in a replication cohort. The highly heterogeneous combinations of susceptibility rare variants, mostly inherited from the healthy mother and father respectively, are correlated with the CCTGA phenotype, any given deleterious variant combination within the CCTGA gene set being specific to the affected individual.

Taken together, the cases could be explained by a mutation load of segregated alleles at loci mostly involved in heart tube *looping*, outflow tract morphogenesis and establishment of left-right asymmetry. Our data suggest that CHD are not Mendelian traits, but rather of polygenic origin. In the majority of cases, both parents harbour a set of susceptibility alleles for the CHD that are inherited by the offspring in a combination that confers the risk of CCTGA.

Key words: congenital heart defects, CCTGA, next generation sequencing

Introduction:

Congenital heart disease (CHD) is the most common human birth defects, occurring in 8/1000 live births^{1 2}. It is still considered as a major cause of infant mortality today³.

The defects comprise an assortment of structural malformations, ranging from minor septal defects to complex life-threatening malformations, which require highly specialized medical care.

Some category of CHDs is due to a misalignment of cardiac chambers (ventricles and great arteries).

We focus on a complex and rare CHD: the congenitally corrected transposition of the great arteries (CCTGA). Its incidence is 1/33000 live birth, accounting for 0.05% of all CHDs⁴. It

implies both atrio-ventricular discordance (the right atrium is emptying into the left ventricle (LV), and left atrium emptying into the right ventricle (RV)) and ventriculo-arterial discordance (the anterior and leftward positioned aorta arises from the RV, while the pulmonary artery, positioned posteriorly and rightward, arises from the LV) ⁵ .

Embryologically, CCTGA is considered to be consequential to the abnormal leftward looping of the primitive heart tube, which results in the morphological LV being positioned to the right of the morphological RV ⁶ .

The genetic bases of the CCTGA are still unknown ⁷ .

It is most often of sporadic appearance. However, there are some rare familial cases reported in the literature ^{7 8} . These reported pedigrees suggest a common pathogenetic pathway involving laterality genes in the pathophysiology of CCTGA. Globally, the risk for a recurrence in subsequent pregnancy for healthy parents with a first child with a CHD is estimated at 1-3% ⁹ .

The aim of this study is to better understand the genetic mechanisms that control the alignment of cardiac chambers and that lead to the CCTGA.

We performed whole-genome and whole-exome sequencing in a cohort of CCTGA. Our data favour a polygenic origin of the disease in the majority of cases, where there is addition of pathogenic variants inherited from the father and the mother, and this addition of variants lead to CCTGA.

Material and methods:

1. The CARREG Collection

All the DNAs of the patients and their parents were obtained thanks to a bio-collection of congenital heart defects in children, based on the collection of DNA by blood sampling. This DNA library, called CARREG (CARDiac congenital defects and REGulation genes) was created by Pr. Damien Bonnet from the M3C-Necker pediatric cardiology department of Necker-

Enfants Malades Hospital, and contains more than 4000 referenced samples of patients with different congenital heart defects.

2. DNA extraction

Patients and parents DNAs were extracted from a leukocyte pellet. It is an extraction method without organic solvent, by sodium dodecyl sulphate (SDS) deionization, proteinase K digestion, denaturation and protein precipitation by saturated NaCl. Precipitation of the DNA is carried out by adding absolute ethanol at -20 ° C. Excess NaCl is removed by addition of 70% ethanol at 4° C. Finally, the DNA is suspended in the water.

3. Next Generation Sequencing (NGS)

a) Whole Genome Sequencing: WGS

The DNAs were purified on AMICON® Ultra columns and their quality verified on an agarose gel. Libraries were created after ultrasonic DNA fragmentation (Covaris E210) using the Illumina TruSeq v3 kit. Complete genome sequencing (WGS) was performed in paired end (2x100 bases) on an Illumina HiSeq2000 platform. The mean depth of sequencing was on the order of 30 reads, except for coverage risks due to the intrinsic quality of the sample.

b) Whole Exome Sequencing: WES

8 µg of genomic DNA has been purified on AMICON® Ultra columns. Its quality was verified on a 2% agarose gel and the concentration of each sample was measured by EXPOSE®. For the production of the exome library, 3 µg of genomic DNA was used with the Agilent SureSelect v5 all Human Exon Kits. The exon multiplex bank was sequenced on a HiSeq2500 machine (Illumina) in "paired-end 76 + 76" mode.

4. Polymerase Chain Reaction (PCR)

PCR (Polymerase Chain Reaction) reactions, optimized on control DNA, were performed using GoTaq DNA polymerase (Promega®). The final volume of PCR mixture was 25 µl. PCR reactions were performed using the DNA of the index individuals extracted from blood or saliva, in an Eppendorf® thermal cycler. PCR products were visualized on a 1% agarose gel with bromophenol blue as a loading buffer. To verify the size of the amplicons we used the Roche Applied Science® DNA VIII Molecular Weight Marker.

5. Sanger sequencing

The amplicons obtained by PCR were subjected to sequencing by the Sanger method. For this, the PCR products were first purified of excess primers and nucleotides using ExoSap-IT (Affimetrix®): 5µl of PCR product+ 2µl of ExoSap. Sanger sequencing makes it possible to read the DNA sequence using fluorophores. When a ddNTP is added to the DNA chain, it stops its elongation and emits a coloured light that is specific to it: we can then identify the nucleotide. These multiple elongations are performed simultaneously on many fragments, thus obtaining the complete sequence. Thus, we used the Big Dye® Terminator v3.1 Cycle Sequencing kit (Applied Biosystems®) to generate the fluorescent fragments, which are then purified on Sephadex G-50 Fine gel to remove low molecular weight impurities. The product is then deposited on the Applied Biosystems 3500xl Genetic Analyser automatic sequencer. The results are analyzed using Sequencing Analysis software (Sequencing Analysis V6.0). The sequences obtained are finally compared with the reference sequence of the human genome (hg19 assembly).

6. Cohort description

- Primary cohort :

Patients with CCTGA were recruited in Necker-Enfants Malades Hospital thanks to the CARREG program. A total of 18 trios (2 parents without CCTGA and 1 child), 17 families with 2 parents, sibs and 1 child affected, 4 duos (1 parent without CCTGA and 1 child), 1 family with a parent and a child both affected, 1 family with 2 affected were recruited.

The inclusion criteria were isolated CCTGA patients without extra-cardiac anomalies. Associated cardiac anomalies were accepted.

502 ethnically matched internal “controls” were also included in the study. They were patients studied at the Imagine Institute, Paris, France, and suffering mainly from immune related diseases.

- Replication cohort :

We used a cohort from 86 unrelated CCTGA patients and 5163 controls from the Wellcome Sanger Institute for replication, all from Caucasian origin. The individuals were sequenced with Agilent SureSelect capture kit V3, V4 or V5. Cases and controls were aligned with BWA software and jointly called with GATK haplotypcaller standard pipeline. Variant effect

predictor tool (VEP, API 94) was used to annotate the dataset, and all quality controls and analyses were performed in HAIL 0.2.

Only bi-allelic and exonic variants were considered, with a mean DP > 15, a mean GQ > 30 and call_rate > 0.95 over the whole cohort. Individual genotypes were set up to missing if Allelic Balance < 20% or > 80% for heterozygous.

7. Bioinformatics analysis

The downstream analyses of primary cohort's sequencing data were performed at the Bioinformatics platform of the University of Paris, France. After demultiplexing, paired-end sequences were aligned to the reference human genome (GRCh37) with the Burrows-Wheeler Aligner ¹⁰. The mean depth of coverage obtained for each sample was greater than 150× for WES and greater than 30x for WGS, and more than 96% of the exome was covered at least 15× in each WES or WGS.

Downstream processing was carried out with the Genome Analysis Toolkit (GATK) ¹¹, SAM tools ¹², and Picard according to documented best practices. GATK Haplotype caller was used for variant calling.

Patients-based *de novo* variants analysis was performed with an in-house software system (PolyWeb) which allows annotation and filtering of variants according to relevant genetic models. All *de novo* variants were then confirmed by Sanger sequencing.

Cases and controls of primary cohort were jointly called with GATK Haplotype caller. We used Variant Quality Score Recalibration (VQSR) to filter the joint VCF file following GATK recommendations for parameter settings: HapMap 3.3, Omni 2.5, dbSNP 138, 1000 Genome phase I for SNPs training sets, and Mills- and 1000 Genome phase I data for indels, and we kept only variants with a VQSLOD score above the tranche sensitivity threshold of 99%. Individual genotypes with a DP <8x, GQ <20 and minor allele ratio < 20% were set to missing.

8. Ethnicity inference

For selection of individuals from European or North African ancestry, we extracted 60 000 single nucleotide polymorphisms with a MAF > 5% in 1000 Genomes Project Phase3 covered

by the WES capture kit and performed a Principal Component Analysis on 1000 Genomes data with projection of patients and controls on the graph, thanks to Plink1.9 software ^{13 14}.

43 patients from the primary cohort and 502 controls were first selected based on this analysis to include a maximum number of patients (**supplementary figure 1**). The individual number of synonymous variants (assumed not to be associated with the phenotype) was compared between cases and controls to check the comparability of data (**supplementary figure 2**).

9. Association analysis

After quality filtering, all variants from the primary cohort were annotated using SnpEff ¹⁵. Only variants with a global call rate > 95% and annotated as missense, frameshift, disruptive inframe deletion/insertion, splice donor or acceptor, or stop gain as most severe consequence among analysed transcripts, and with a minor allele frequency below 0.1% in gnomAD database were kept in the analysis. Missense variants pathogenicity were analysed by an *in silico* pathogenicity criteria based on the CADD score ¹⁶ and/or on SIFT and PolyPhen softwares ¹⁷.

The evolutionary conservation at each residue was also investigated using multiple sequence alignments from the UCSC genome browser.

10. Family-based design

We investigated potential risk alleles transmission in the 43 families through the Rare-Variant Generalized Disequilibrium Test ¹⁸. This method is an extension of the Transmission Disequilibrium Test (TDT) allowing the grouping of information across multiple rare variants distributed within a genomic region. In this analysis, we considered rare functional variants (as previously defined) with or without a CADD score >15, and aggregated them at the gene level.

11. Case - control Design

An enrichment in rare variants was investigated at the individual gene level and at cardiac genesets level. Two different methods were used: the CAST method comparing the number of carriers of selected variants between cases and controls to test if a gene or group of genes is preferentially mutated among cases, and a logistic regression method, comparing the number of alleles in the considered genes or genesets between the 2 groups of individuals, to test if an accumulation of variants in a gene or group of genes could explain the phenotype.

12. Cardiac genes

We selected a list of cardiac genes, described in the literature to play a role in cardiac morphogenesis in human and mouse. We used Gene Ontology (GO) ^{19 20 21} Consortium and Mouse Genome Informatics (MGI) ^{22 23}. The terms used in GO were: heart morphogenesis, heart tube formation, outflow tract development, directing laterality specification. Genes expressed during embryonic heart development were also added. A total of 1708 genes were selected.

Results:

We sequenced by NGS a cohort of 43 families with 44 patients with CCTGA, from the CARREG program.

The cohort was composed of: 17 families with 2 healthy parents and siblings, 18 trios (healthy parents-patient), 4 duos (healthy mother-patient), 2 simplex cases (single patient) and 2 family cases: 1 case of recurrence of CCTGA in siblings (affected brother and sister) and 1 case of vertical transmission of CCTGA (affected father and patient).

The cohort included 18 girls and 26 boys (the boy/girl sex ratio is 1.5).

All patients had isolated CCTGA, with no extracardiac anomalies. The CCTGA could be simple or associated with other cardiac anomalies such as ventricular septal defects, Ebstein morphology of the tricuspid valve, or pulmonary stenosis.

The demographic and anatomical description of the cohort is shown in **Table (1)**.

de novo analysis:

Applying the *de novo* mode of inheritance in Polyweb, we selected a list of 20 genes with a *de novo* variant (loss of function or missense) in 17 different families that meet the selection criteria for frequency and nature of variation. 6 variants are loss of function (frameshift insertion or deletion, splice acceptor donor, start-stop lost, STOP gain) and 14 are missense variants.

This list of genes with the description of the variant, its CADD score and the probability to loss of function intolerance (PLI score) is presented in **table (2)**. The presence of another variant inherited from one of the parents in each gene is also detailed.

Among the 20 genes with *de novo* variation, 15 have a cardiac function described in the literature (GO, MGI). These genes are: *ATP2B1*, *EPHB2*, *FLII*, *HAND1*, *HSPA2*, *LRP1*, *MYOCD*, *PCSK5*, *POU5F1*, *PTCH1*, *RMND5A*, *SMAD2*, *THBD*, *VTA1* and *WDR4*.

If we consider that 1708 genes / 19370 protein-coding whole genome genes have a role cardiac development of function, and in our cohort 15 genes / 20 genes with *de novo* variant have also cardiac function, by performing a hypergeometric test, we found a significant enrichment of cardiac genes in the *de novo* genes ($p < 10E-10$).

To investigate possible variant enrichment in *de novo* genes, we analyzed variants annotated by SnpEff as disruptive inframe insertion/deletion, frameshift, missense, splice acceptor, splice donor, stop gained, stop lost.

We analyzed the 26 families in whom no *de novo* variant was found. We do not find any enrichment of variants in the 27 cases compared to the 502 controls with a p value=0.83 for a MAF <1%. The same results are seen on the list of 15 genes with a cardiac function.

Considering the enrichment in cardiac genes in *de novo* analysis, we focused on the list of 1708 cardiac genes.

We thus made a comparative analysis of the number of variants in these 1708 genes in patients and controls.

We observed a trend towards an enrichment in the number of very rare variants (MAF <0.01%) in cardiac genes in patients compared to controls ($p=0.025$). By adjusting for the first 5 principal components (PC), the significance of this enrichment increases (p value 0.005).

These results are shown in **Table (3)**.

To go further, we looked among the 1708 cardiac genes, if some played a more important role than the others and presented a higher number of variants in the cases than in the controls.

To do this, we performed statistical analysis using the CAST method and logistic regression tests.

We analyzed 2 types of variants: PTV (protein truncating variant) represented frameshift_variant, stop_gained, splice_donor_variant, splice_acceptor_variant, and PAV (protein altering variant) represented missense variant, inframe deletion, inframe insertion, start lost, stop retained variant, stop lost, protein altering variant, start retained variant .

No gene passed the significance threshold.

Analysis by group of genes: the polygenic model

By analyzing the families of patients, among the 1708 cardiac genes, a group of 156 genes seemed to group interesting variants meeting the defined selection criteria.

This list of 156 genes is shown in **Table (4)**.

Each patient carries an association of variants (between 3 and 10 variants), inherited from the father and the mother, in addition of a *de novo* variant.

The genotype of each of the 43 families in the CCTGA cohort (inherited and *de novo* variants) is represented by the pedigrees in **Figure (1)**.

To support the hypothesis of a preferential transmission of variants in candidate genes from parents to sick children, we performed a family analysis with RV-GDT (Rare Variant Generalized Disequilibrium Test).

There was a significant transmission of rare variants in the 156 genes to the sick child ($p=0.026$).

To validate that these 156 genes are indeed linked to the disease, we tried to replicate our results on a German cohort of 86 index cases of non-syndromic CCTGA, and 5163 controls.

We used a logistic regression on the number of variants within the group of genes, according to the type of variants (PTV=protein truncating variant and PAV=protein altering variant).

We found a significant enrichment ($p\ value=0.04$, OR=1.5 for MAF 1%) in the replication cohort of PTV variants in the 156 genes of interest.

Table (5) summarizes these results.

Discussion:

CHD is a large global problem in child health ²⁴. Progress in the diagnosis and management of congenital heart disease is leading to reduced mortality and increased prevalence of CHD in adults. It is estimated that currently 85% of children with a complex CHD will reach adulthood. The population of adults with CHD is growing approximately 5% / year ²⁵. This highlights the need to lay a molecular foundation for these diseases. Basically, they represent the first step before physiopathological studies that will allow better understanding of the normal and pathological cardiac development leading to the main anomalies. Understanding the underlying genetic, molecular, and cellular mechanisms of CHDs will have real consequences for clinical practice, such as improving diagnosis, improving guidelines before surgery, and improving genetic counselling.

In this study, we sequenced a cohort of 44 cases of a complex and rare CHD, the congenitally corrected transposition of the great arteries, by whole genome and whole exome sequencing. Our data contradict a monogenic mode of inheritance, favouring a polygenic origin of CHD. In the majority of cases of our series, the parents harbour a set of susceptibility alleles for the disease that are inherited by the offspring in a combination that confers the risk of CCTGA.

Evidence exists for both inherited as Mendelian traits and *de novo* mutations associated to CHD. Yet, large multigenerational pedigrees demonstrating Mendelian inheritance are exceptional. Autosomal recessive inheritance of CHD has also been documented and these most extensively in populations with a high degree of consanguinity. Most of the reported multigenerational pedigrees of X-linked and dominantly inherited CHDs include either non-lethal CHDs such as atrial septal defect or those with variable phenotypes that include survivable CHDs and more severe manifestations of the same spectrum ²⁶.

Up to now CHDs most often occurs sporadically, suggesting a role of *de novo* mutations in their pathogenesis. In 2013, the Paediatric Cardiac Genomics Consortium performed whole exome sequencing of 362 parent-offspring trios ²⁷. This study found *de novo* point/insertion-deletion mutations in several hundred of genes that collectively contribute to at least 10% of severe

CHDs. Another study reported by Bruencker et al in 2017²⁸ and encompassing to date the largest genetic investigation of a single CHD cohort with 2871 probands, shows that the genetic contribution of *de novo* variants in CHDs was 8%. These elegant study do not demonstrate any causal relationship between the identifying variants and the occurrence of a CHD. Another hypothesis for this frequent sporadic occurrence of CHD is the limited role of genetic variations in causing the risk of CHD in the general population. This hypothesis is, with a very high probability, the reason for the scarce association of CHD with single gene mutation. A more holistic approach of the development of CHD should be developed without excluding the role of genes variations but considering that the cardiac morphogenesis would be disturbed when a combination of “cardiac” gene variations will exceed a certain threshold. This is the polygenic hypothesis for the cause of CHD. The hypothesis of polygenic inheritance of CHD is not new and is supported by the landscape of recurrence both in proportion of all CHDs’ incidence, and in the viewpoint of anatomical concordance. The number of genes’ variations in different “cardiac” genes may depend on the population considered. Further, it is also obvious that the set of genes to test to support this hypothesis is arbitrarily defined, and this is a limitation for describing the associations of genes conferring a risk for CCTGA. Here, we chose to have a minimal set of three “cardiac” genes to support the probability of an inherited combination $\leq 1\%$.

We thus built a set with 156 “cardiac” genes. Variants identified in these genes are characterized by their damaging effect and their shared role in heart development and/or have an embryological cardiac expression. Many interact with each other in the same signalling pathways. For most of them, mouse models show that mutations of these genes are associated with anomalies in the heart looping and the left-right asymmetry.

We managed to demonstrate that the CCTGA follows a complex polygenic model of inheritance, which consists of addition of susceptibility alleles inherited from the both healthy mother and father and that finally leads to a mutation burden in genes playing a role in heart morphogenesis in the affected sib.

We finally show through family segregation studies that a disease associated combination of variants is not replicated in healthy individuals in the index family.

Because of the complexity of our genetic polygenic model, no functional analysis could be conducted.

Conclusion:

Our data contradict a monogenic mode of inheritance, favouring an oligogenic origin of CCTGA, where both parents harbour genetic predisposition variants for the disease (susceptibility alleles) and both transmit this load of variants to the affected offspring, explaining the low recurrence risk.

Figures and tables:

N°	Family code	Sex patient	Complete cardiologic diagnosis	Cardiac arrhythmias
1	A001	F	SS, mesocardia, CCTGA, large VSD, pulmonary atresia	0
2	A002	M	SS, levocardia, CCTGA, Ebstein disease	WPW
3	A003	M	SS, levocardia, CCTGA, Ebstein disease	CAVB
4	A004	M	SS, levocardia, CCTGA	0
5	A005	F	SS, dextrocardia, CCTGA, VSD, pulmonary atresia	postoperative CAVB
6	A006	M	SS, mesocardia, CCTGA, VSD, pulmonary atresia	0
7	A007	F	SS, levocardia, CCTGA, Ebstein disease	0
8	A008	F	SS, levocardia, CCTGA, VSD, mitral cleft, Ebstein disease	CAVB
9	A009	M	SS, levocardia, CCTGA, VSD, pulmonary stenosis	0
10	A010	F	SS, mesocardia, CCTGA, large VSD, pulmonary stenosis, bicuspid aortic valve	WPW
11	A011	M	SS, levocardia, CCTGA, VSD	0
12	A012	F	SS, levocardia, CCTGA, multiple VSDs, coarctation of the aorta	0
13	A013	M	SS, levocardia, CCTGA, multiple VSDs, coarctation of the aorta, tricuspid straddling	0
14	A014	M	SS, levocardia, CCTGA	CAVB
15	A015	M	SS, levocardia, CCTGA	0
16	A016	M	SS, levocardia, CCTGA	0
17	A017	M	SS, levocardia, CCTGA	0

18	A018	M	SS, levocardia, CCTGA, restrictive VSD, pulmonary stenosis	CAVB
19	A019	F	SS, mesocardia, CCTGA, VSD, pulmonary stenosis	0
20	A020	M	SS, levocardia, CCTGA, VSD, hypoplasia of RV	JT, CAVB
21	A021	M	SS, levocardia, CCTGA, hypoplasia of RV, VSD, pulmonary stenosis	CAVB
22	A022	M	SS, levocardia, CCTGA, pulmonary stenosis	0
23	A023	F	SS, levocardia, CCTGA, VSD	CAVB
24	A024	F	SS, levocardia, CCTGA, restrictive VSD, Ebstein disease	0
25	A025	F	SS, dextrocardia, CCTGA, VSD, pulmonary stenosis	0
26	A026	M	SS, levocardia, CCTGA, large multiple VSDs	0
27	A027	F	SS, levocardia, CCTGA	WPW
28	A028	M	SI, dextrocardia, CCTGA, VSD, pulmonary stenosis, mitral straddling	0
29	A029	M	SS, mesocardia, CCTGA, VSD, pulmonary atresia	0
30	A030	M	SS, mesocardia, CCTGA, large VSD, pulmonary atresia	0
31	A031	F	SS, levocardia, CCTGA	CAVB
32	A032	F	SS, levocardia, CCTGA	0
33	A033	F	SS, levocardia, CCTGA, VSD	CAVB
34	A034	M	SS, dextrocardia, CCTGA	0
35	A035	F	SS, dextrocardia, CCTGA, VSD	0
36	A036	M	SS, levocardia, CCTGA, hypoplasia of RV, 2 larges VSDs, LSVC	0
37	A037	F	SS, dextrocardia, CCTGA, VSD, coarctation of the aorta	0

38	A038	M	SS, levocardia, CCTGA, VSD, tricuspid straddling	0
39	A039	F	SS, levocardia, CCTGA	CAVB
40	A040	M	SI, dextrocardia, CCTGA, hypoplasia of LV	0
		F	SS, levocardia, CCTGA	0
41	A041	M	SS, dextrocardia, CCTGA	0
42	A042	M	SS, levocardia, CCTGA	0
43	A043	M	SS, levocardia, CCTGA, VSD	0

CAVB: complete atrio-ventricular block, CCTGA: congenitally corrected transposition of the great arteries, JT: junctional tachycardia, LSVC: left superior vena cava, LV: left ventricle, RV: right ventricle, SS : situs solitus, SI: situs inversus , VSD : ventricular septal defect, WPW: Wolf Parkinson White

Table 2: *de novo* variants in CCTGA cohort

Gene	OMIM	PLI score	Nature of variation	Family	gDNA change	CADD score	Inherited variant
<i>ATP2B1</i>	108731	1	Missense	A003	12:90020324 C/T	24	1
<i>CALML6</i>	610171	0	Missense	AO19	1:1848211 G/A	15	1
<i>EPHB2</i>	600997	1	Missense	A015	1:123037486 G/A	17	0
<i>FLII</i>	600362	0	Missense	A026	17:18151140 C/A	26	3
<i>HAND1</i>	602406	0,1	Missense	A029	5:153857402 A/C	27	0
<i>HSPA2</i>	140560	0.05	Missense	A007	14:65009037 C/G	25	0
<i>IQSEC3</i>	612118	0	Deletion frameshift	A025	12:284120 TACAA/-	-	0
<i>IRF2</i>	147576	0,8	Missense	A029	4:185310196 T/G	24	0
<i>LRP1</i>	107770	1	Missense	A038	12:57605109 A/G	19	2
<i>MYOCD</i>	606127	0.33	Missense	A015	17:12649301 T/C	23	0
<i>MYOZ1</i>	605603	0	Missense	A036	10:75391917 G/A	24	0
<i>NES</i>	600915	0	Insertion frameshift	A024	1:156640659 G/GC	-	0
<i>PCSK5</i>	600488	0	Missense	A002	9:78937968A/C	26	3
<i>POU5F1</i>	164177	0.95	Splice donor/acceptor	A016	6:31133347 C/G	35	0
<i>PTCH1</i>	601309	1	Missense	A018	9:98229667 C/T	28	1
<i>RMND5A</i>	618964	0.9	Missense	A002	2:86998763 A/G	26	0
<i>SMAD2</i>	601366	1	Stop gained	A020	18:45395639 G/T	35	1
<i>THBD</i>	188040	0.05	Missense	A023	20:23029339 G/A	26	1
<i>VTA1</i>	610902	0	Start-stop lost	A012	6:142468504 A/G	27	0
<i>WDR4</i>	605924	0	Stop gained	A004	21:44296816 T/A	41	0

Table 3 : statistical analysis case/control of variants in the 1708 cardiac genes in CCTGA cohort

	Number of variants	Ncarriers_controls / 502	Ncarriers_cases/ 43	Total alleles Nb controls	Total alleles Nb cases	Alleles Nb / control	Alleles Nb /case	pvalueLRT	pvalueWald
MAF1%, CADD>15	13853	502	43	24016	2152	47.8	50	0.44	0.43
MAF 0.1%, CADD>15	9951	502	43	11586	1076	23.1	25	0.21	0.21
MAF0.01%, CADD >15	6326	500	43	6252	627	12.5	14.6	0.025	0.022
MAF<0.01%, CADD>15 Adjusted on the 5th first PCs								0.005	0.004

Table 4: List of 156 cardiac genes with pathogenic variant in CCTGA cohort

<i>ACAD11</i>	<i>COL4A3BP</i>	<i>FOXP1</i>	<i>MAP3K3</i>	<i>POU5F1</i>	<i>TMEM2</i>
<i>ALDH1A2</i>	<i>CREBBP</i>	<i>GBX2</i>	<i>MATR3</i>	<i>PRICKLE1</i>	<i>TMEM67</i>
<i>APC</i>	<i>DAAM2</i>	<i>GDF1</i>	<i>MECOM</i>	<i>PRRT2</i>	<i>VTA1</i>
<i>ARID1A</i>	<i>DDX11</i>	<i>GPR161</i>	<i>MEGF8</i>	<i>PSEN1</i>	<i>WDR90</i>
<i>ARID1B</i>	<i>DISP1</i>	<i>HAND1</i>	<i>MESP1</i>	<i>PTCH1</i>	<i>WDR35</i>
<i>ARID2</i>	<i>DLC1</i>	<i>HSPA2</i>	<i>MESP2</i>	<i>PTPN11</i>	<i>WDR4</i>
<i>ARID3B</i>	<i>DMRT2</i>	<i>HSPG2</i>	<i>MICAL2</i>	<i>R3HCC1</i>	<i>WDR62</i>
<i>ARL13B</i>	<i>DNAAF1</i>	<i>IFT122</i>	<i>MKL1</i>	<i>RERE</i>	<i>WNT11</i>
<i>ASXL1</i>	<i>DNAH1</i>	<i>IFT140</i>	<i>MYBPC3</i>	<i>RMND5A</i>	<i>WTIP</i>
<i>ASXL2</i>	<i>DNAH11</i>	<i>IGHMBP2</i>	<i>MYH14</i>	<i>ROBO1</i>	<i>XIRP2</i>
<i>ASPH</i>	<i>DNAH5</i>	<i>INTU</i>	<i>MYOCD</i>	<i>ROBO2</i>	<i>ZEB1</i>
<i>ATP2B1</i>	<i>DNAH9</i>	<i>JARID2</i>	<i>NCOR2</i>	<i>ROR2</i>	
<i>AXIN2</i>	<i>DNAI1</i>	<i>KDM6A</i>	<i>NCSTN</i>	<i>RPGRIP1L</i>	
<i>BBS4</i>	<i>DNHD1</i>	<i>KDM6B</i>	<i>NF1</i>	<i>RYR1</i>	
<i>BCAS3</i>	<i>DOCK1</i>	<i>KIAA0319</i>	<i>NODAL</i>	<i>RYR2</i>	
<i>BCOR</i>	<i>DSCAM</i>	<i>KIAA0586</i>	<i>NOS2</i>	<i>SEMA3A</i>	
<i>BMP10</i>	<i>DVL2</i>	<i>KMT2A</i>	<i>NOS3</i>	<i>SHOX2</i>	

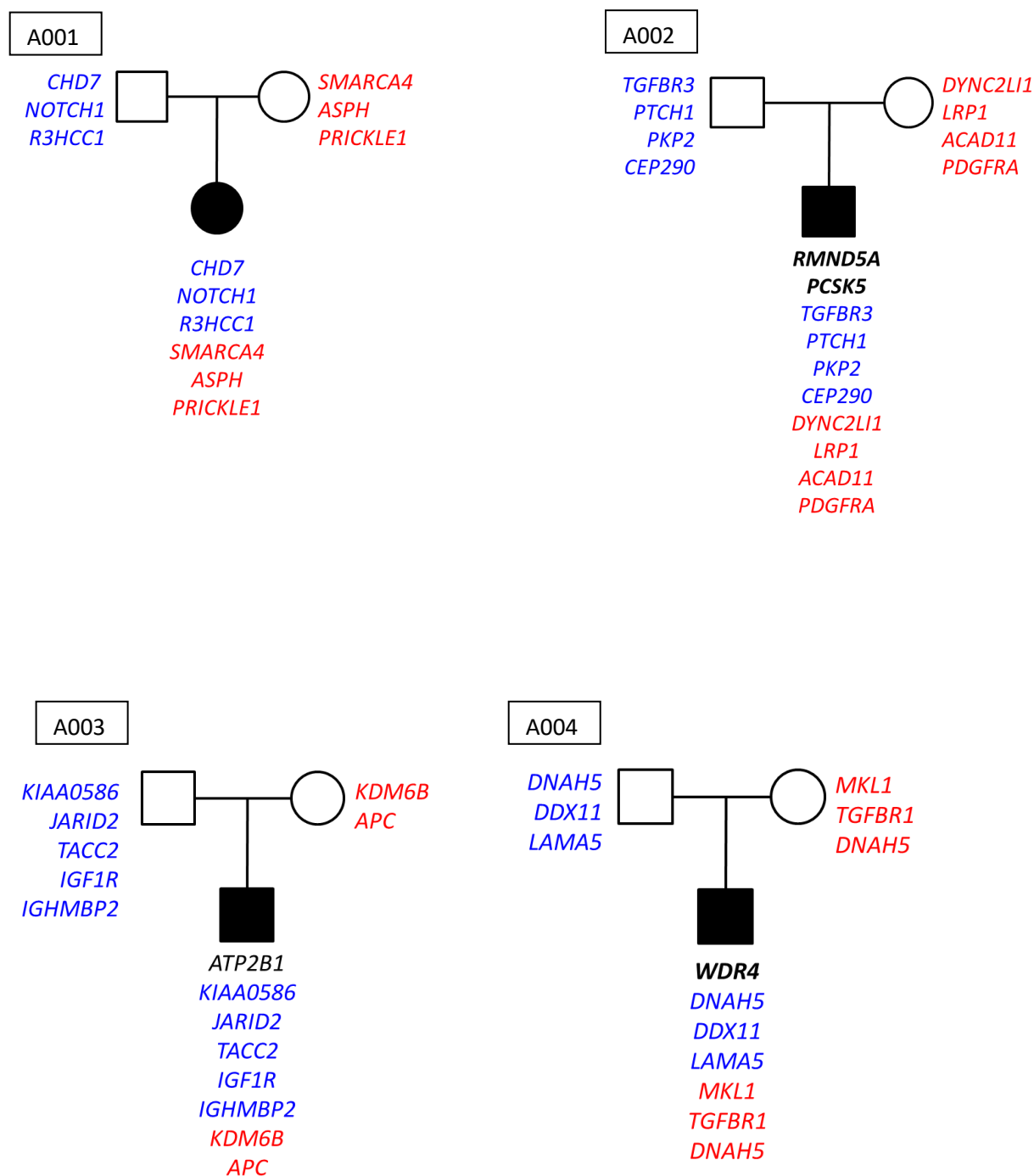
<i>BMP4</i>	<i>DVL3</i>	<i>KMT2B</i>	<i>NOTCH1</i>	<i>SMAD2</i>
<i>CACNA1C</i>	<i>DYNC2H1</i>	<i>KMT2D</i>	<i>NOTCH2</i>	<i>SMARCA4</i>
<i>CASZ1</i>	<i>DYNC2LI1</i>	<i>LAMA2</i>	<i>PATZ1</i>	<i>SMARCD3</i>
<i>CCDC114</i>	<i>EP300</i>	<i>LAMA5</i>	<i>PBX2</i>	<i>SMG1</i>
<i>CCDC39</i>	<i>EPHB2</i>	<i>LEFTY2</i>	<i>PCSK5</i>	<i>SNED1</i>
<i>CENPF</i>	<i>EPHB4</i>	<i>LMNA</i>	<i>PDGFRA</i>	<i>SNX17</i>
<i>CEP120</i>	<i>ETV5</i>	<i>LRP1</i>	<i>PDLIM7</i>	<i>TACC2</i>
<i>CEP290</i>	<i>FGFR1</i>	<i>LRP2</i>	<i>PDLIM3</i>	<i>TAL1</i>
<i>CHD7</i>	<i>FHOD3</i>	<i>LRP6</i>	<i>PKD1</i>	<i>TBX5</i>
<i>CHD8</i>	<i>FLII</i>	<i>LTBP1</i>	<i>PKD1L1</i>	<i>TGFBR1</i>
<i>CHRD</i>	<i>FOXC2</i>	<i>LTBP3</i>	<i>PKP2</i>	<i>TGFBR3</i>
<i>CNPY1</i>	<i>FOXD3</i>	<i>LTBP4</i>	<i>PLXND1</i>	<i>THBD</i>

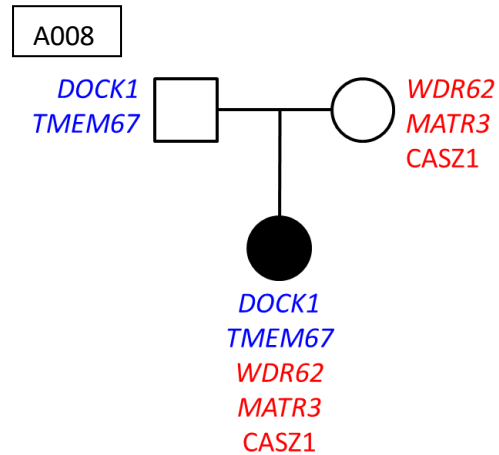
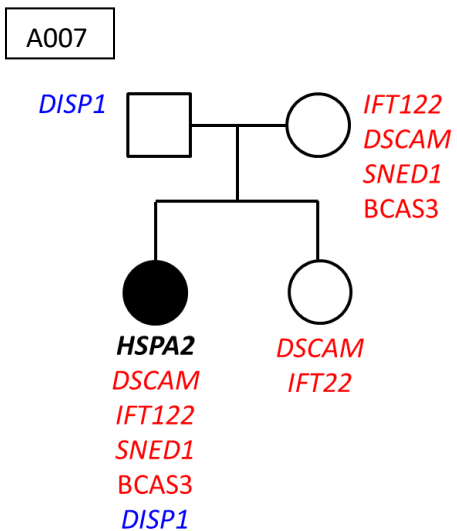
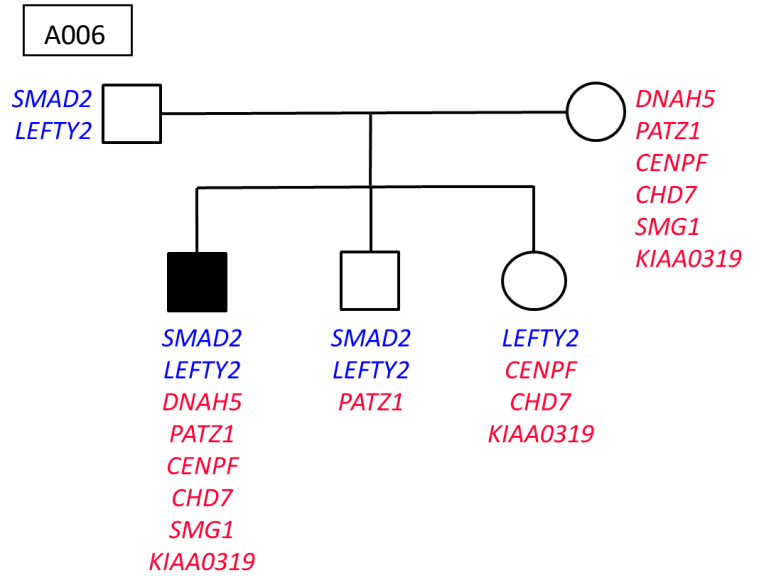
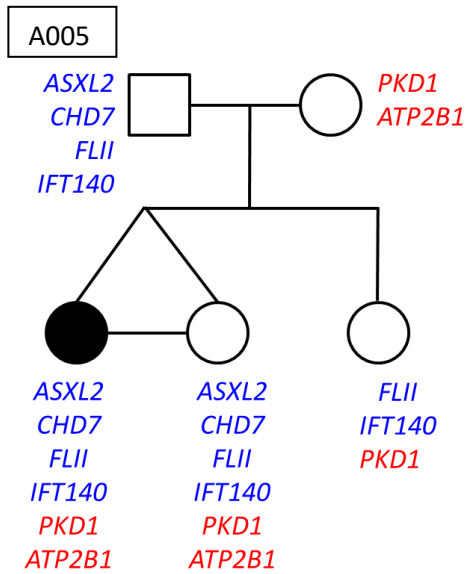
Table 5: Replication cohort : statistical analysis

Cluster name	MAF	Type of variant	Total cases	Total controls	N_hets_cases	N_hets_controls	Nb var/cases	Nb var/controls	p_value	Odds radio
Geneset-156	1,00 ^E -02	PAV	86	5163	479	31417	5,57	6,09	0,03	0,92
Geneset-156	1,00 ^E -02	PTV	86	5163	18	716	0,21	0,14	0,04	1,5
Geneset-156	5,00 ^E -03	PAV	86	5163	362	23776	4,21	4,61	0,05	0,92
Geneset-156	5,00 ^E -03	PTV	86	5163	17	660	0,20	0,13	0,04	1,54

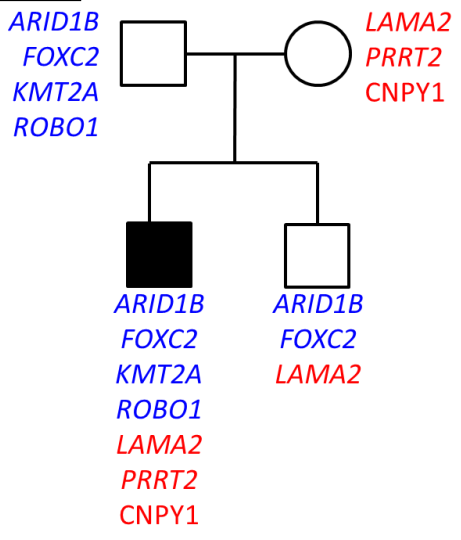
Figure (1): pedigrees of the 43 CCTGA families with the genotype of cases, parents and siblings

Color code: blue: variants inherited from the father. Red: variant inherited from the mother. Black: *de novo* variant. Grey: unknown inheritance

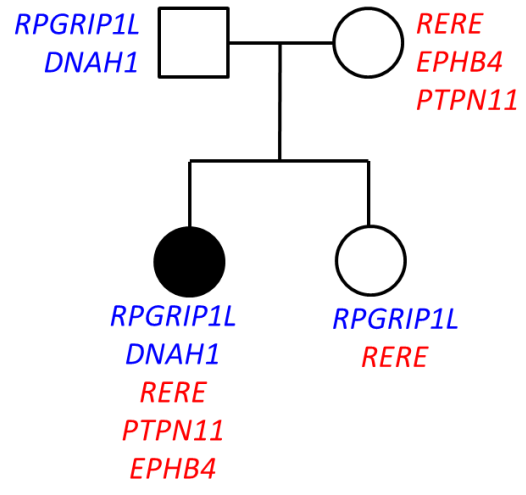




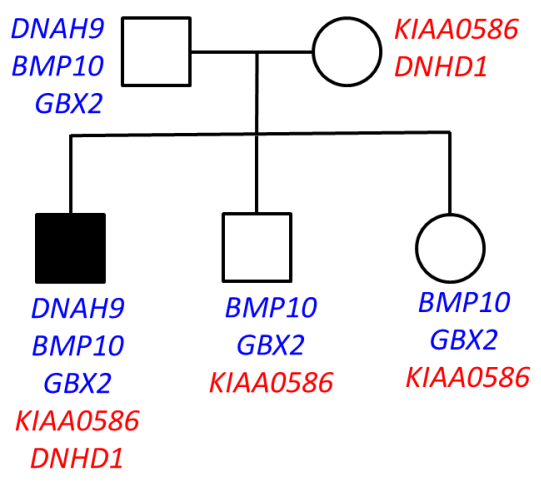
A009



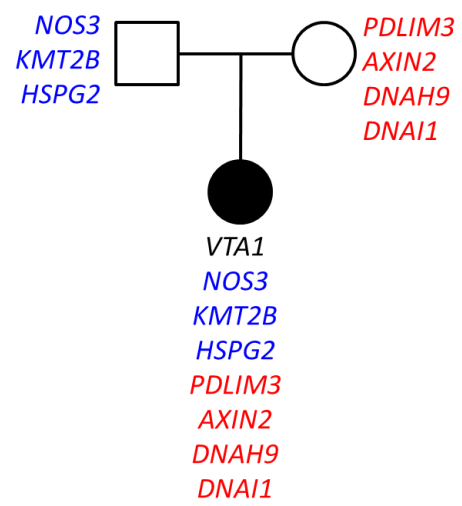
A010

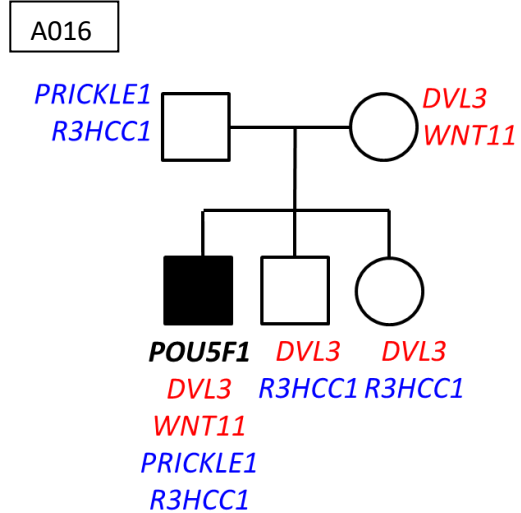
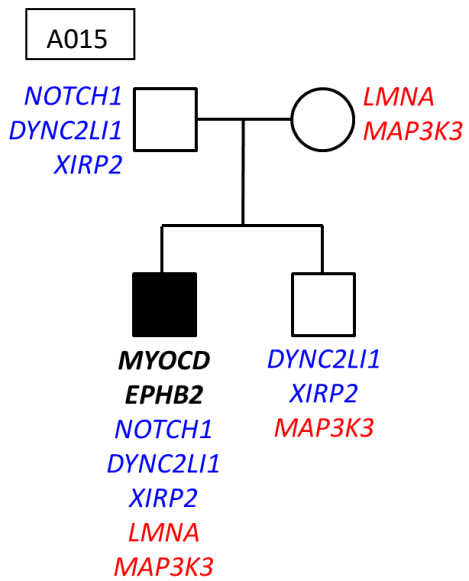
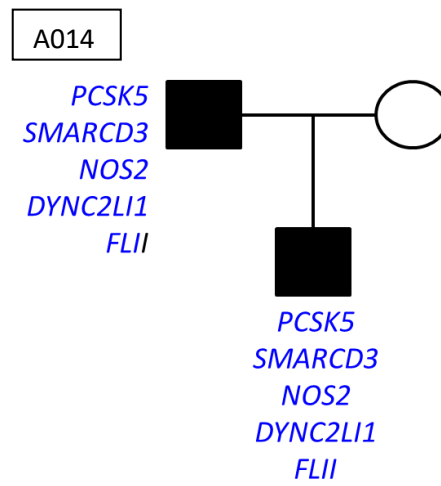
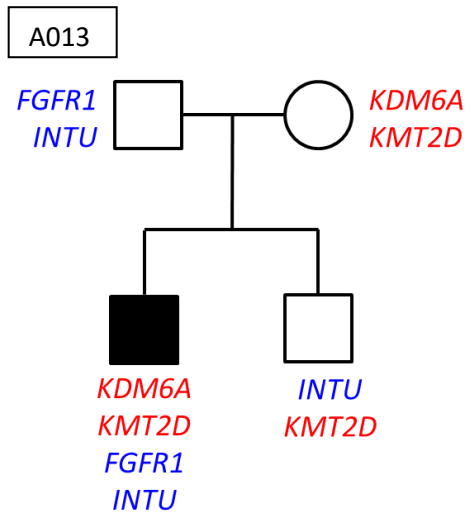


A011

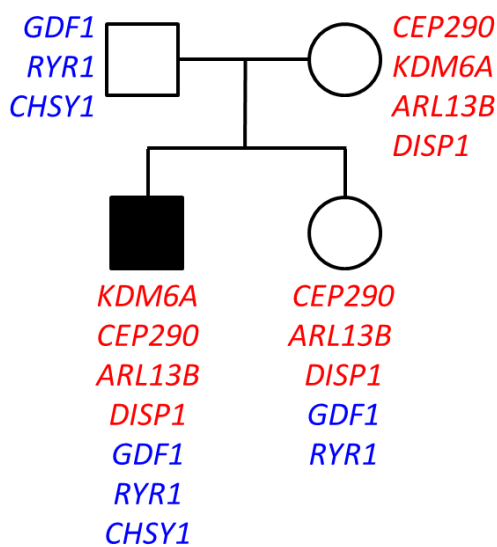


A012

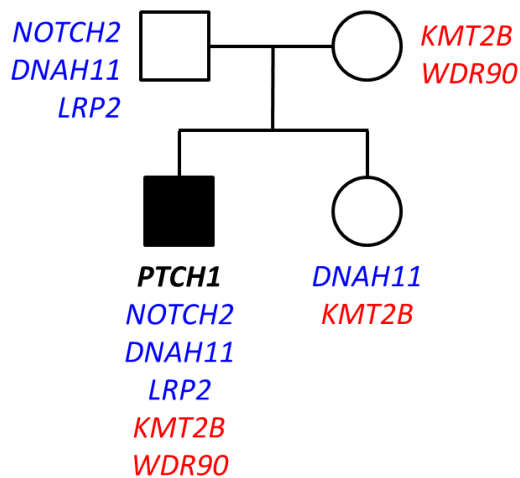




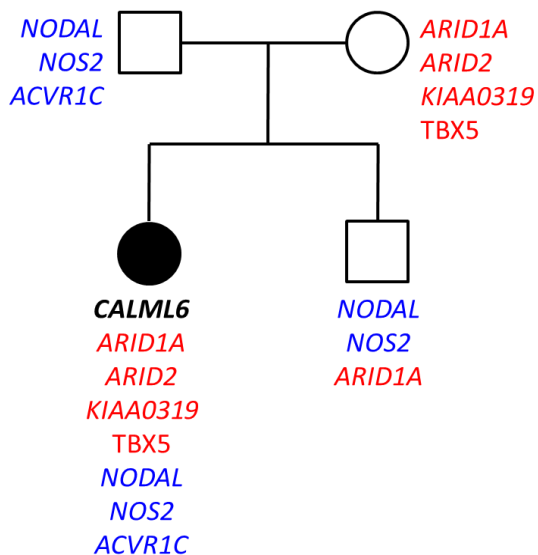
A017



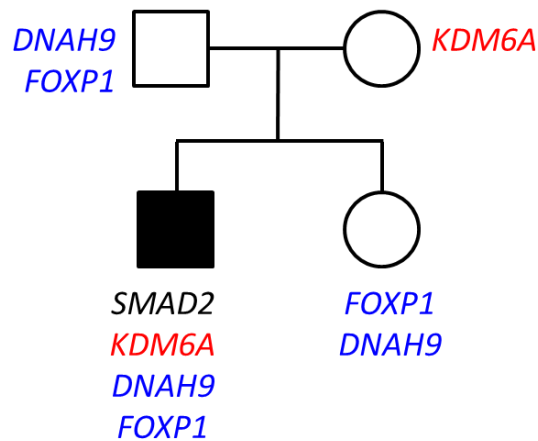
A018



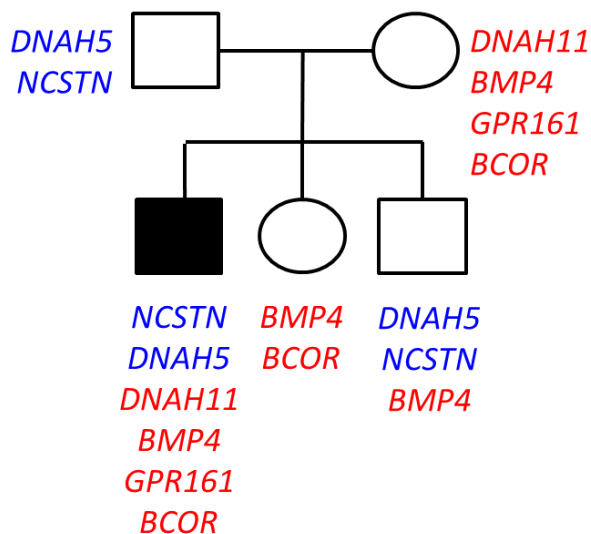
A019



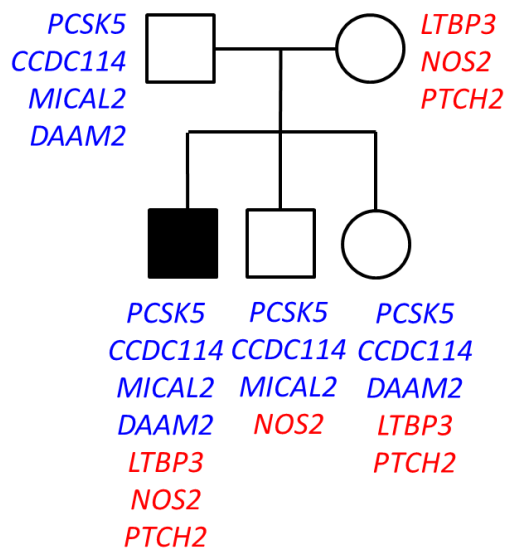
A020



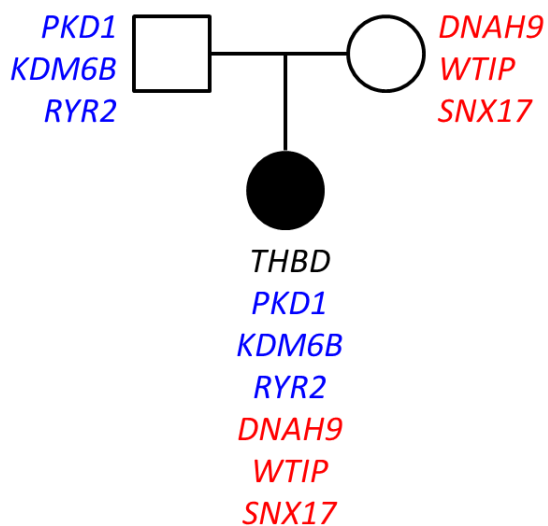
A021



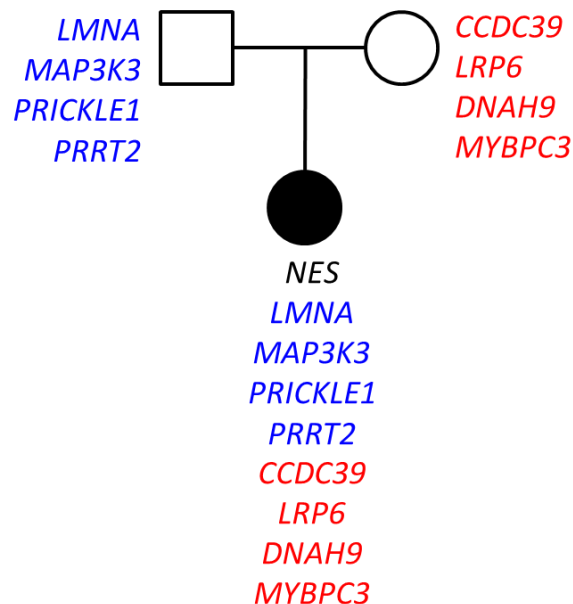
A022

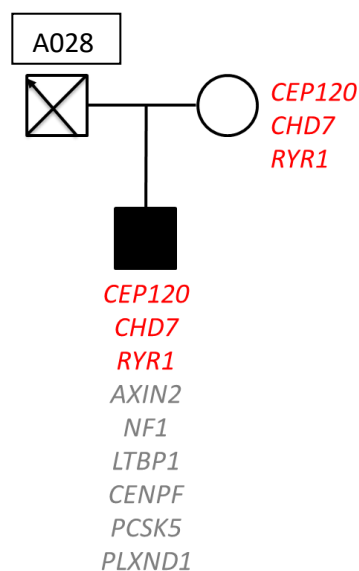
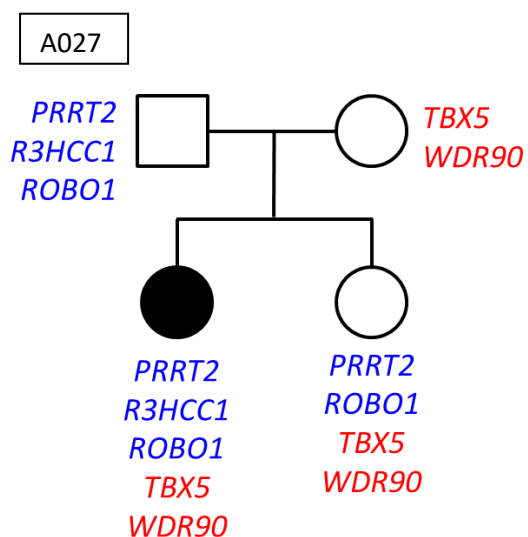
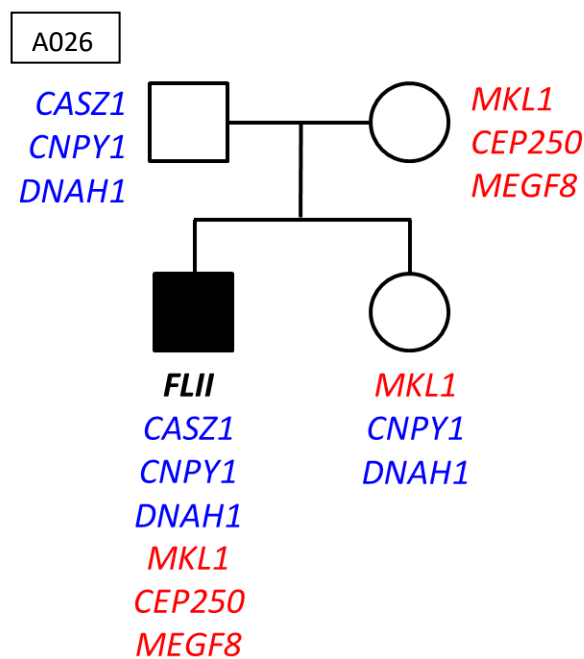
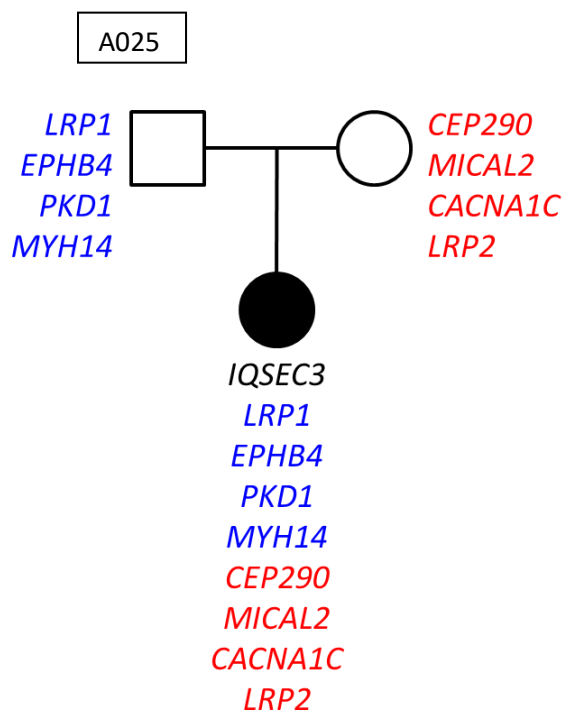


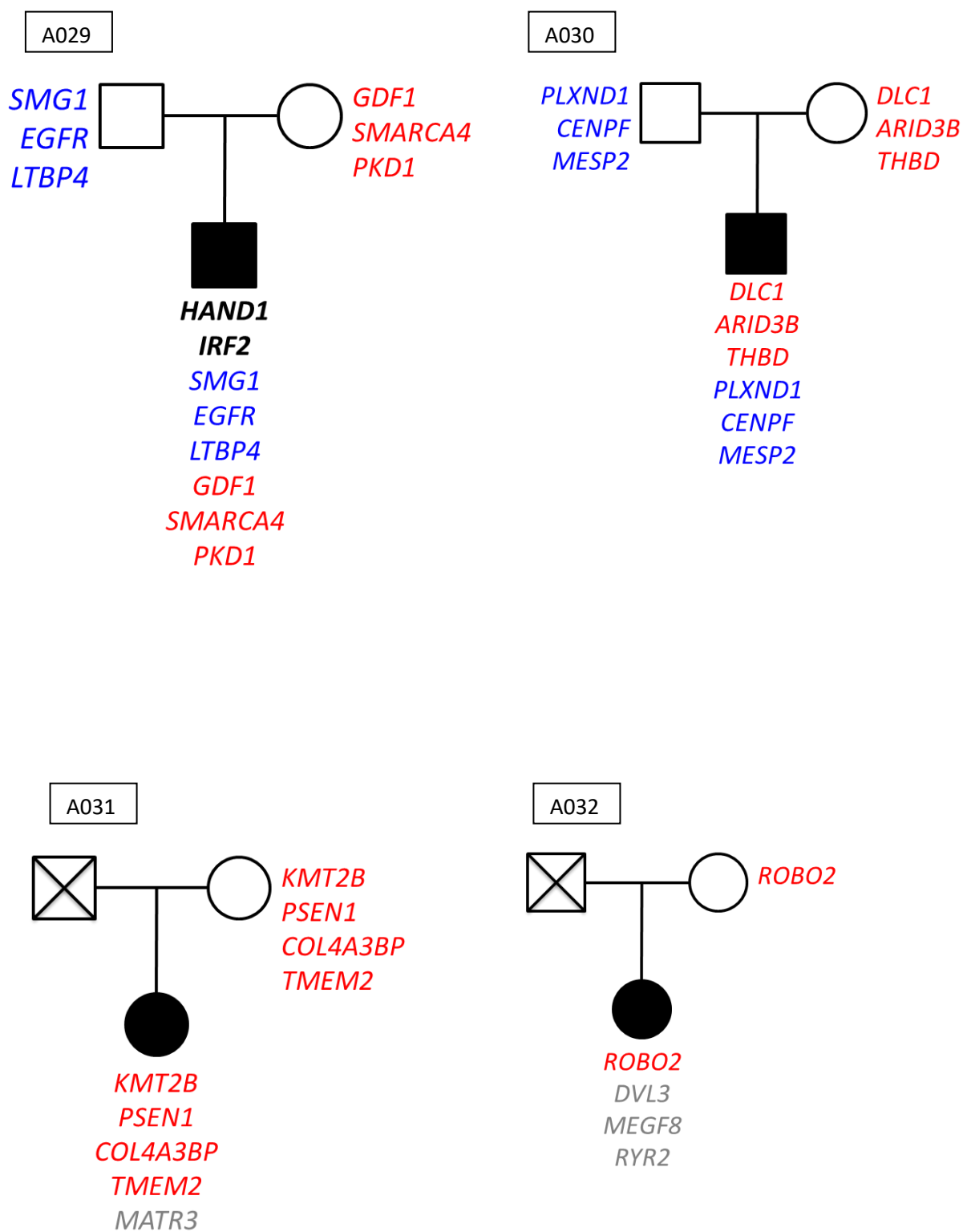
A023

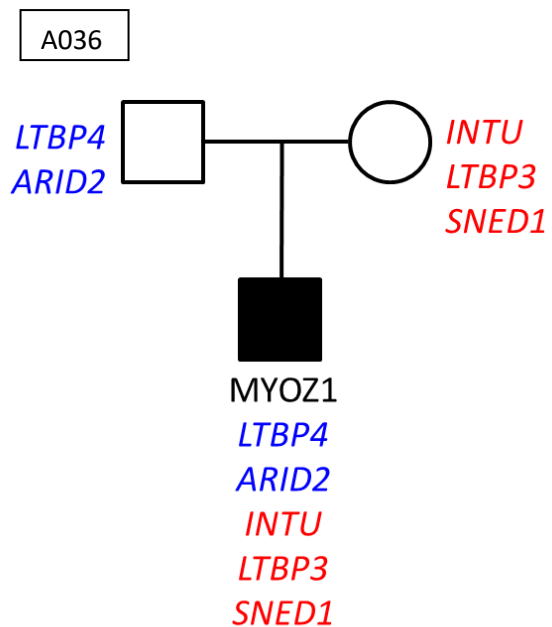
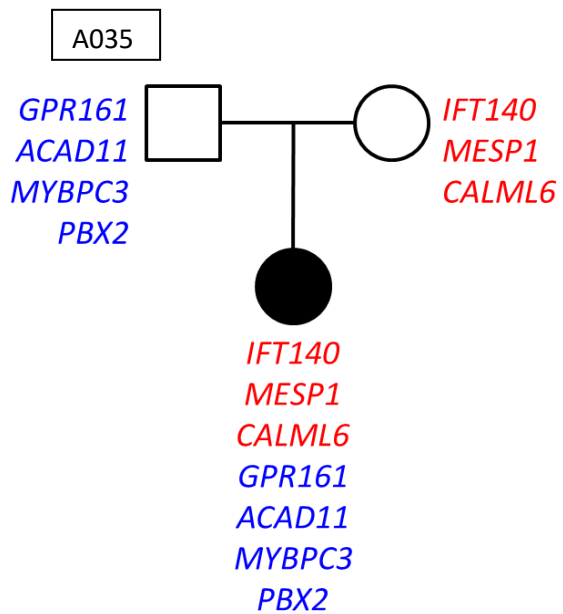
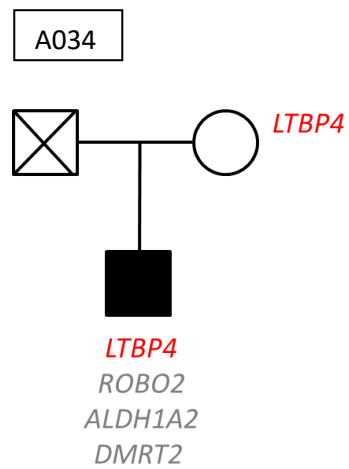
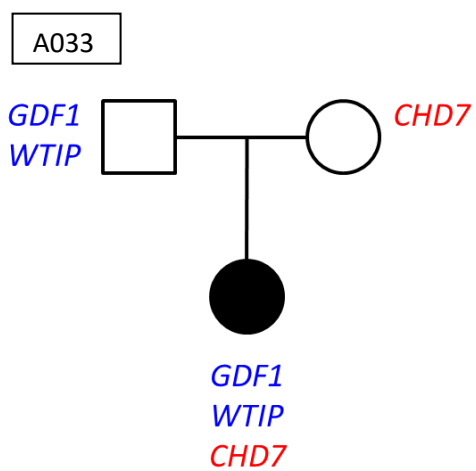


A024





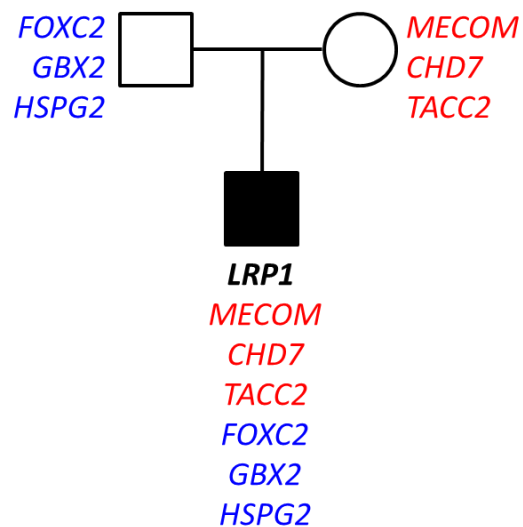




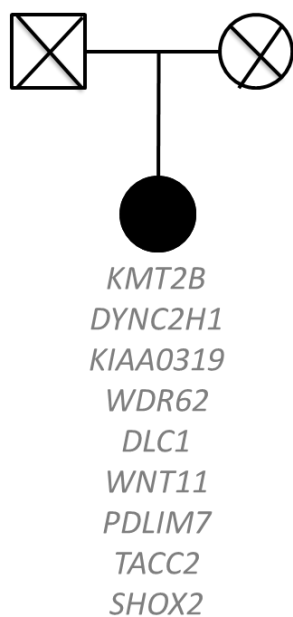
A037



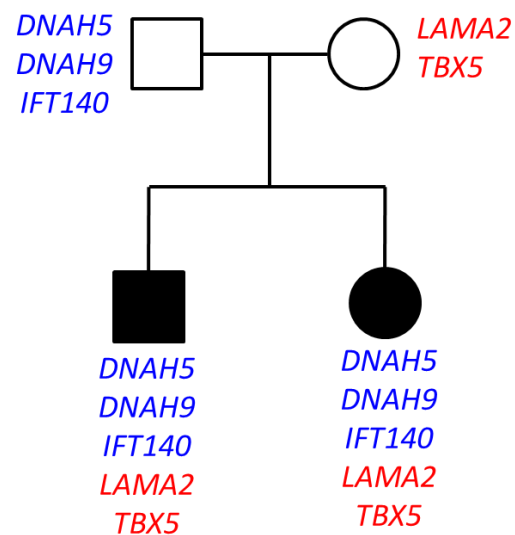
A038

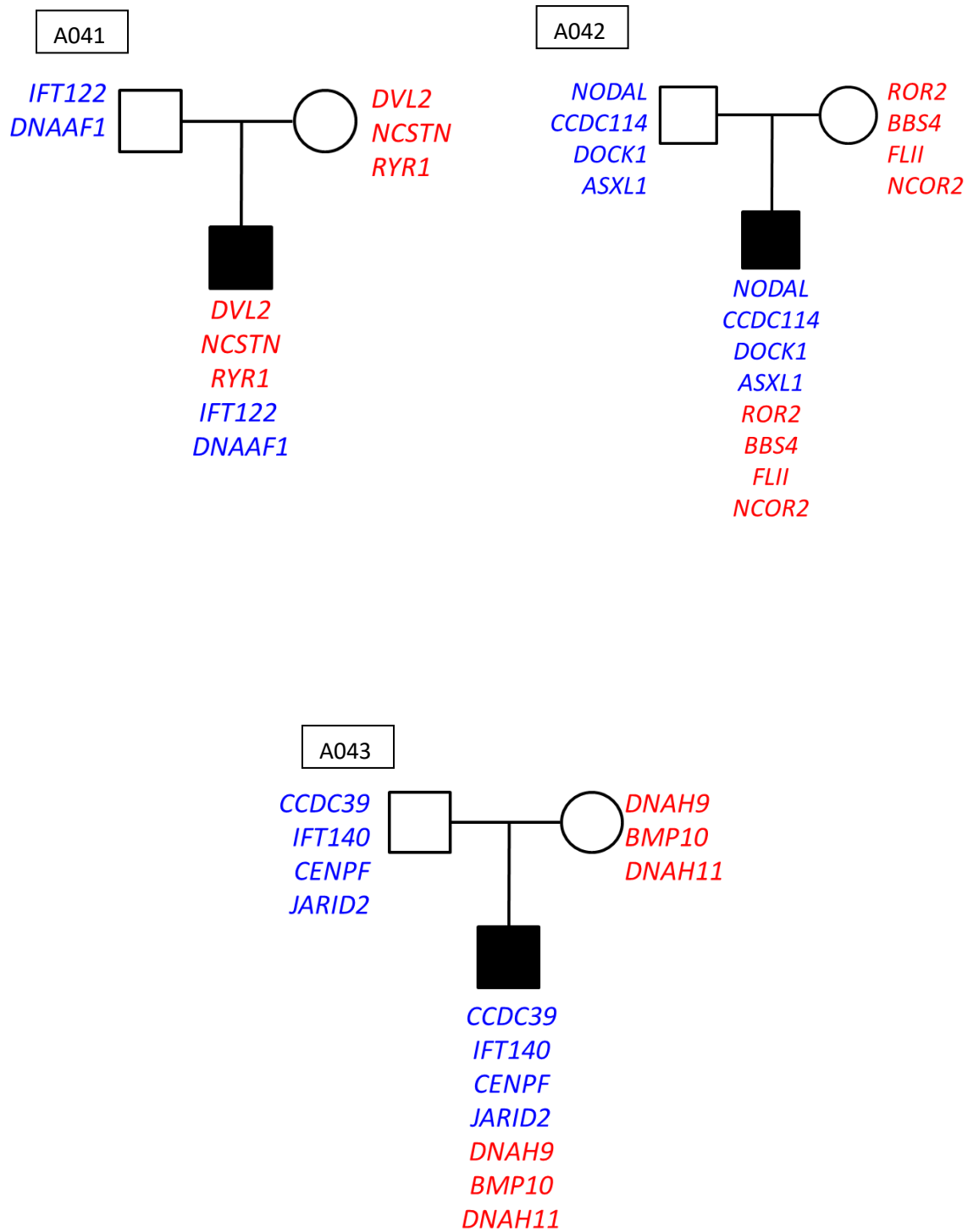


A039



A040





Acknowledgements:

We thank the families for their participation. We thank Anna PELET for technical assistance. This work was supported by grants from the Agence Nationale de la Recherche (ANR-Heart Asymmetry) and the Fondation pour la Recherche Médicale (FRM).

Competing financial interests:

The authors declare no competing financial interests.

Ethical and Legal Issues:

I- For the human DNA studies, a particular agreement for the present study has been obtained with our Paris Descartes University (Île-de-France II) Ethical Committee in January 2010 (Comité de Protection des Personnes (CPP) agreement N° AOM 95224, P959892), according to which, each person involved in the study signed an informed consent; as for under-age individuals, the consent was signed by the parents as required by national laws. For obtaining written informed consent, each participating family has received an explanation (both oral and written) about the aims and objectives of the study.

II- Biobanking of patient materials (e.g. blood DNA) is achieved in our ISO9001-certified DNA bank, within the Imagine building, in accordance with national, and international rules. The leukocyte DNA storage and automated extraction was performed at the Necker CRB-ADN bank, Centre de Ressources Biologiques (CRB-ADN). DNA is stored in two aliquots, under anonymous codes. Of important note, our ongoing CARREG cohort sampling, has obtained an agreement from the CNIL (Commission Nationale de l'Informatique et des Libertés).

References

1. Ferencz C, Rubin JD, Mccarter RJ, et al. Congenital heart disease: prevalence at livebirth. *Am J Epidemiol.* 1985;121(1):31-36. doi:10.1093/oxfordjournals.aje.a113979
2. Reller MD, Strickland MJ, Riehle-Colarusso T, Mahle WT, Correa A. Prevalence of Congenital Heart Defects in Metropolitan Atlanta, 1998-2005. *J Pediatr.* 2008;153(6):807-813. doi:10.1016/j.jpeds.2008.05.059
3. van der Linde D, Konings EEM, Slager MA, et al. Birth Prevalence of Congenital Heart Disease Worldwide. *J Am Coll Cardiol.* 2011;58(21):2241-2247. doi:10.1016/j.jacc.2011.08.025
4. Šamánek M, Voříšková M. Congenital Heart Disease Among 815,569 Children Born Between 1980 and 1990 and Their 15-Year Survival: A Prospective Bohemia Survival Study. *Pediatr Cardiol.* 1999;20(6):411-417. doi:10.1007/s002469900502
5. Wallis GA, Debich-Spicer D, Anderson RH. Congenitally corrected transposition. *Orphanet J Rare Dis.* 2011;6(1):22. doi:10.1186/1750-1172-6-22

6. Kutty S, Danford DA, Diller GP, Tutarel O. Contemporary management and outcomes in congenitally corrected transposition of the great arteries. *Heart*. 2018;104(14):1148-1155. doi:10.1136/heartjnl-2016-311032
7. Tortigue M, Nield LE, Karakachoff M, et al. Familial Recurrence Patterns in Congenitally Corrected Transposition of the Great Arteries: An International Study. *Circ Genomic Precis Med*. 2022;15(3). doi:10.1161/CIRCGEN.121.003464
8. Piacentini G, Digilio MC, Capolino R, et al. Familial recurrence of heart defects in subjects with congenitally corrected transposition of the great arteries. *Am J Med Genet A*. 2005;137A(2):176-180. doi:10.1002/ajmg.a.30859
9. Becker TA, Van Amber R, Moller JH, Pierpont MEM. Occurrence of cardiac malformations in relatives of children with transposition of the great arteries. *Am J Med Genet*. 1996;66(1):28-32. doi:10.1002/(SICI)1096-8628(19961202)66:1<28::AID-AJMG7>3.0.CO;2-S
10. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma Oxf Engl*. 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324
11. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297-1303. doi:10.1101/gr.107524.110
12. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinforma Oxf Engl*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352
13. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4:7. doi:10.1186/s13742-015-0047-8
14. www.cog-genomics.org/plink/1.9/.
15. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6(2):80-92. doi:10.4161/fly.19695
16. 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. *Nat Genet*. 2014;46(3):310-315. doi:10.1038/ng.2892
17. Yu B, Hinchcliffe M, eds. *In Silico Tools for Gene Discovery*. Vol 760. Humana Press; 2011. doi:10.1007/978-1-61779-176-5
18. He Z, Zhang D, Renton AE, et al. The Rare-Variant Generalized Disequilibrium Test for Association Analysis of Nuclear and Extended Pedigrees with Application to Alzheimer Disease WGS Data. *Am J Hum Genet*. 2017;100(2):193-204. doi:10.1016/j.ajhg.2016.12.001
19. Ashburner M, Ball CA, Blake JA, et al. Gene Ontology: tool for the unification of biology. *Nat Genet*. 2000;25(1):25-29. doi:10.1038/75556
20. The Gene Ontology Consortium. Gene Ontology Annotations and Resources. *Nucleic Acids Res*. 2012;41(D1):D530-D535. doi:10.1093/nar/gks1050
21. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res*. 2019;47(D1):D330-D338. doi:10.1093/nar/gky1055
22. Blake JA, Bult CJ, Kadin JA, Richardson JE, Eppig JT, the Mouse Genome Database Group. The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Res*. 2011;39(Database):D842-D848. doi:10.1093/nar/gkq1008

23. Bult CJ, Kadin JA, Richardson JE, Blake JA, Eppig JT, the Mouse Genome Database Group. The Mouse Genome Database: enhancements and updates. *Nucleic Acids Res.* 2010;38(suppl_1):D586-D592. doi:10.1093/nar/gkp880
24. Mandalenakis Z, Giang KW, Eriksson P, et al. Survival in Children With Congenital Heart Disease: Have We Reached a Peak at 97%? *J Am Heart Assoc.* 2020;9(22):e017704. doi:10.1161/JAHA.120.017704
25. van der Bom T, Zomer AC, Zwinderman AH, Meijboom FJ, Bouma BJ, Mulder BJM. The changing epidemiology of congenital heart disease. *Nat Rev Cardiol.* 2011;8(1):50-60. doi:10.1038/nrcardio.2010.166
26. Yuan S, Zaidi S, Brueckner M. Congenital heart disease: emerging themes linking genetics and development. *Curr Opin Genet Dev.* 2013;23(3):352-359. doi:10.1016/j.gde.2013.05.004
27. Pediatric Cardiac Genomics Consortium, Writing Committee:, Gelb B, et al. The Congenital Heart Disease Genetic Network Study: Rationale, Design, and Early Results. *Circ Res.* 2013;112(4):698-706. doi:10.1161/CIRCRESAHA.111.300297
28. Jin SC, Homsy J, Zaidi S, et al. Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat Genet.* 2017;49(11):1593-1601. doi:10.1038/ng.3970