Sara Melo Dias

Licenciada em Ciências Biomédicas

Characterization of a chromosome rearrangement associated with cardiopathy and autism

Dissertação para obtenção do Grau de Mestre em
Genética Molecular e Biomedicina

Orientador: Doutor Dezsö David, Investigador Auxiliar, Departamento de Genética Humana do Instituto Nacional de Saúde Doutor Ricardo Jorge, I.P.

Júri:

Presidente: Prof. Doutora Maria Alexandra Núncio de Carvalho Ramos Fernandes

Arguente: Prof. Doutor Pedro Miguel Ribeiro Viana Baptista

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Dezembro de 2017

Spine (Lombada)

Characterization of a chromosome rearrangement associated with cardiopathy and autism

Sara Melo Dias

2017

Sara Melo Dias

Licenciada em Ciências Biomédicas

Characterization of a chromosome rearrangement associated with cardiopathy and autism

Dissertação para obtenção do Grau de Mestre em
Genética Molecular e Biomedicina

Orientador: Doutor Dezsö David, Investigador Auxiliar, Departamento de Genética Humana do Instituto Nacional de Saúde Doutor Ricardo Jorge, I.P.

Júri:

Presidente: Prof. Doutora Maria Alexandra Núncio de Carvalho Ramos Fernandes

Arguente: Prof. Doutor Pedro Miguel Ribeiro Viana Baptista

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Dezembro de 2017

Identification of candidate gene for Cognitive Deficiencies through characterization of chromosome rearrangements

Acknowlegments

Abstract

Chromosomal rearrangements have been associated with multiple congenital abnormalities, including malformative syndromes and global developmental delay.

The aim of this study was identification of candidate genes for a complex phenotype characterized by cardiopathy and autism, identified in an individual with a chromosome translocation t(4;7)(q21.1;p21.2). Since classical and molecular cytogenetic analyses have low resolutions, large-insert whole-genome sequencing (liWGS) was applied for identification and mapping of structural chromosomal alterations. By this approach, the 4q21.1 breakpoint was identified between genomic positions chr4:73,918,924-74,049,529 on 4q13.3, whereas the 7p21.2 breakpoint between chr7:13,184,731-14,536,001 [GRCh38/hg38]; suggesting the occurrence of deletions at both breakpoints. Additionally, a 473Kb deletion on 2p16.3 was also identified in the proband. Nucleotide-level resolution of the breakpoints and familial segregation analysis were carried out by amplification of the junction fragments and Sanger sequencing. At the 4q13.3 breakpoint, the 130Kb deletion erases four genes *PF4*, *PPBP*, *CXCL5* and *CXCL3*, whereas at 7p21.2, the 1351Kb deletion removes the entire *ETV1* and disrupts *DGKB* and the long non-coding intergenic (Linc) RNA *AC011288.2*. Furthermore, at this breakpoint region, genomic array analysis identified in the proband's father a 742Kb deletion comprising *DGKB* and *ETV* but not the *LincRNA AC011288.2.* The maternally inherited 473Kb deletion on 2p16.3 removes the first 5 exons of *NRXN1*, a gene associated with Pitt-Hopkins like syndrome (OMIM #614325), susceptibility to schizophrenia and chromosome 2p16.3 deletion (OMIM #614332). Similar deletions have been reported with incomplete penetrance and variable expressivity. Several genes from the 7p21.2 breakpoint region and especially those affected by the deletion, *DGKB*, *ETV1* and *LincRNA AC011288.2*, have been linked with cognitive, speech, language and auditory disorders. In conclusion, coinheritance of the maternally derived deletion on 2p16.3 and the deletion at the breakpoint of the der(7) on 7p21.2 appear to be the most contributively alterations for the proband's phenotype. At the time, *NRXN1*, *DGKB*, *ETV1* and *LincRNA AC011288.2* are the most likely genes to be responsible for the proband's phenotype, being those mainly characterised by cardiopathy and autism.

Keywords: Cardiopathy, Autism, Chromosomal Rearrangement, *NRXN1*, *DGKB*, *ETV1*

x

Resumo

Os rearranjos cromossómicos estruturais têm sido associados a anomalias congénitas incluindo síndromes malformativas e atraso global do desenvolvimento. O objetivo deste estudo foi identificar os genes candidatos para um fenótipo caracterizado por cardiopatia e autismo, presente num individuo com o rearranjo cromossómico t(4;7)(q21.1;p21.2). Visto que as análises de citogenética clássica e molecular têm baixa resolução, a sequenciação pangenómica de grandes insertos foi a metodologia usada para mapear as alterações cromossómicas estruturais. Através desta abordagem determinou-se que o ponto de quebra previamente identificado em 4q21.1, está na região chr4:73.918.924-74.049.529 em 4q13.3, e que o ponto de quebra de 7p21.2 está na região chr7:13.184.731-14.536.001, [GRCh38/hg38] sugerindo a ocorrência de deleções nos pontos de quebra. Adicionalmente, foi identificada uma deleção de 473Kb em 2p16.3 no probando. A determinação dos pontos de quebra com resolução nucleotídica e os estudos de segregação familiar foram feitos por amplificação e sequenciação *Sanger* dos fragmentos de junção. No ponto de quebra em 4q13.3 a deleção de 130Kb elimina 4 genes, *PF4*, *PPBP*, *CXCL5* e *CXCL3*, enquanto que em 7p21.2 a deleção de 1351Kb remove *ETV1* e interrompe *DGKB* e o *LincRNA AC011288.2*. em 7.21.2 o *array* genómico identificou uma deleção de 742Kb no pai do probando, envolvendo apenas *DGKB* e *ETV1*. A deleção de 473Kb, de origem materna, elimina os 5 exões iniciais de *NRXN1*, um gene associado a síndrome de Pitt-Hopkins (OMIM #614325), suscetibilidade para esquizofrenia e a deleções em 2p16.3 (OMIM #614332). Deleções semelhantes foram reportadas como tendo penetrância incompleta e expressividade variável. Vários genes em 7p21.1 têm sido associados a desordens cognitivas, de linguagem e auditivas. Em suma, A co-herança da deleção materna em 2p16.3 e a deleção associada ao ponto de quebra do der(7) em 7p21.2, aparentam ser as alterações cromossómicas com maior contribuição para o fenótipo. Assim, *NRXN1, DGKB, ETV1* e *LincRNA AC011288.2* foram identificados como sendo os genes mais prováveis de serem causadores do fenótipo de cardiopatia e autismo apresentado pelo probando.

Palavras chave: Cardiopatia, Autismo, Rearranjo Cromossómico, *NRXN1*, *DGKB*, *ETV1*

Table of Contents

List of Figures

List of Tables

List of Abbreviations

3C – Chromosome Conformation Capture

3D – Three Dimensional

aCGH – Comparative genomic hybridization array

ACTC1 – Actin alpha cardiac muscle

AFP – Alpha-fetoprotein

AIFM1 – Apoptosis inducing factor mitochondria associated 1

ALB – Albumin

ASD – Autism Spectrum Disorder

BACs – Bacterial artificial chromosomes

BCR – Balanced Chromosomal Rearrangements

CCDC88C – Coiled-coil domain containing 88C

CCR – Complex Chromosomal Rearrangement

CD36 – CD36 molecule

CGH – Comparative Genomic Hybridization

CHD – Congenital Heart Defect

chr – Chromosome

CIT – Citron Rho interacting serine/threonine kinase

CITED2 – Cbp/P300 interacting transactivator with Glu/Asp rich carboxy terminal domain 2

CMV – Cytomegalovirus

CNV – Copy Number Variation

COL1A1 – Collagen type I alpha 1 chain

COL1A2 – Collagen type I alpha 2 chain

CRTAP – Cartilage associated protein

CXCL3 – Chemokine, C-X-C motif, ligand 3

CXCL5 – Chemokine, C-X-C motif, ligand 5

CXCL8 – Chemokine, C-X-C motif, ligand 5

del – Deletion

der – Derivative Chromosome

DGKB – Diacylglycerol beta, 90kD

DMSO – Dimethyl sulfoxide

DNA – Deoxyribonucleic Acid

dNTPs – Deoxynucleotides

dup – Duplication

EBV – Epstein-Barr virus

EDTA – Ethylenediamine tetraacetic acid

ELN – Elastin

ETV1 – E-twenty-six variant gene 1

EUROCAT – European surveillance of congenital anomalies

FISH – Fluorescence *in situ* Hybridization

FLNA – Filamin A

FMR1 – Fragile X mental retardation 1

FSHR – Follicle stimulating hormone receptor

GABRB3 – Gamma-aminobutyric acid type A receptor beta 3 sebunit

GATA4 – GATA binding protein 4

GTL – Giemsa-Trypsin-leishman

Hi – Haploinsufficiency index

hTERT – Telomerase catalytic subunit gene

ID – Intellectual Disability

ins – Insertion

INSA, I.P. – Instituto Nacional de Saúde Doutor Ricardo Jorge, Instituto Público

inv – Inversion

IQ – Intelligence Quotient

JAG1 – Jagged 1

KRAS – KRAS proto oncogene, GTPase

LB – Lysogeny broth

LCL – Lymphoblastoid Cell Line

LHGCR – Luteinizing hormone/ choriogonadotropin receptor

liWGS – Large Insert Whole-Genome Sequencing

LOH – Loss of heterozygosity

low-TE – Low-Tris- Ethylenediamine tetraacetic acid

MECP2 – Methyl CpG binding protein 2

MEF2C – Myocyte enhancer factor 2C

MYH6 – Myosin heavy chain 6

NCBI – National Center for biotechnology information

NDD – Neurodevelopmental Disorder

NF1 – Neurofibromin 1

NGS – Next Generation Sequencing

NLGN2 – Neuroligin 2

NLGN3 – Neuroligin 3

NOTCH1 – Notch 1

NOTCH2 – Notch 2

NRAS – NRAS proto-oncogene, GTPase

NRXN1 – Neurexin 1

NX2-5 – NK2 Homeobox 5

PACs – P1 artificial chromosomes

PBMCs – Peripheral Blood Mononuclear Cells

PBS – Phosphate Buffered Saline

PF4 – Platelet factor 4

PMSE – Proteasome activator subunit 4

PNPT1 – Polyribonucleotide nucleotidyltransferase 1

PPBP – Pro-platelet basic protein

PRSS50 – Protease serine 50

PTEN – Phosphatase and tensin homolog

PTH1R – Parathyroid hormone 1 receptor

RPMI – Roswell ParK memorial institute

RPS27A – Ribosomal protein S27A

RTN4 – Neurite outgrowth inhibitor (reticulon 4) clathrin heavy chain linker domain containing I

SDS – Sodium dodecyl sulphate

SHANK3 – SH3 and multiple Ankyrin repeat domains 3

SNP – Single Nucleotide Polymorphism

SNV – Single Nucleotide Variant

SOS1 – SOS RAS/RAC guanine nucleotide exchange factor 1

SOX3 – SRY-box 3

SV40 – Simian virus 40

TAD – Topological association domain

TAE – Tris-Acetate-EDTA

TBE – Tris-Borate-EDTA

TBX20 – T-box 20

TBX5 – T-box 5

TEAD1 – TEA domain transcription factor 1

TLL1 – Tolloid like 1

TNFSF11 – TNF superfamily member 11

TPP2 – Tripeptidyl Peptidase 2

TSC1 – Tuberous sclerosis 1

TSC2 – Tuberous sclerosis 2

TUBA1A – Tubulin alpha 1 a

TUBB3 – Tubulin beta 3 class III

UBE3A – Ubiquitin protein ligase E3A

UCSC – University of California, Santa Cruz

UV – Ultraviolet

WGS – Whole genome sequencing

WT1 – Wilms Tumor 1

List of Symbols

~ – Approximately

% – Percentage

%(v/v) – Volume per volume percentage

%(wt/v) – Weight per volume percentage

bp – Base pair

F – Forward

Kb – Kilo-base pair

M – Molar

Mb – Mega-base pair

mM – Millimolar

n – Haploid number, refers to the number of chromosomes (n=23)

OD – Optical density

p – Chromosome short arm

q – Chromosome long arm

R – Reverse

rpm – Rotations per minute

t – chromosomal translocation

U – Enzyme units

X – Chromosome X

xg – Relative centrifugal force

Y – Chromosome Y

1. Introduction

1.1. Considerations about the Human Genome and its organization

The Human Genome is the set of all the genetic information related to *Homo Sapiens* specie, codified on more or less 3 billion DNA base pairs allotted in 46 chromosomes. (Gardner, R.J. Mckinlay; Sutherland, GranT R.; Shaffer 2012; National Library of Medicine 2017). These 46 chromosomes are organized in 22 matching pairs and two sex chromosomes. One homolog of each pair is inherited from the mother and the other one is inherited from the father. (Gardner, R.J. Mckinlay; Sutherland, GranT R.; Shaffer 2012)

The 22 matching pairs correspond to Autosomes and the remaining ones correspond to the Sex Chromosomes or Genosome. Although autosomes' homologs have the same morphology in both sexes, sex chromosomes are different in males and females. Females have a pair of X while males have a X and a Y chromosome. A single set of homologs, one of each autosome, plus one of the sex chromosomes, is the haploid set (n=23). That only exists, as such, in gametocytes. All the other cells in the organism, somatic cells, have the diploid set (2n=46). (Gardner, R.J. Mckinlay; Sutherland, GranT R.; Shaffer 2012)

Chromosomes are formed by chromatin, a complex of DNA and proteins, in its condensed state. There are different condensed forms of chromatin: euchromatin corresponds to the less condensed form, and constitutes coding DNA – genes; on the other hand, heterochromatin is the most condensed form, and constitutes the non-coding DNA present in centromeres and telomeres. DNA associated proteins, histones and non-histones, provide scaffolding of chromosomes. (Gardner, R.J. Mckinlay; Sutherland, GranT R.; Shaffer 2012) Histones are strongly conserved alkaline proteins, their positive charges allow them to associate with DNA, which is negatively charged, and help condense it into chromatin fibres. Those fibres loop around some histones that function like spools and form nucleosomes. (Nature Education 2014)

Each chromosome has 2 arms, capped at terminal extremities by telomeres, and divided by the centromere. The shorter arms are designated as p while the longer arms are named as q. The centromere is a specialized region of non-coding DNA that, at mitosis, provides the site at which the spindle apparatus can be anchored and draw each separated chromatid to the opposite poles of the dividing cell. Telomeres are specialized non-coding DNA sequences comprising many repeats of the sequence "TTAGGG" that confers stable ends to chromosomes preventing its fusion with the chromatin of other chromosomes. (Gardner, R.J. Mckinlay; Sutherland, GranT R.; Shaffer 2012; Nature Education 2014)

Chromosomes are classically distinguishable on the basis of: size, centromere position and banding pattern. An Ideogram is the diagrammatic or schematic representation of a chromosome. A karyotype is the representation of the chromosomal constitution of one individual. (Gardner, R.J. Mckinlay; Sutherland, GranT R.; Shaffer 2012)

1.2.   Chromosomal abnormalities

Since chromosomes contain the genetic information essential to form an healthy human being, any change in the number or structure of chromosomes may lead to death, mental retardation, a variety of genetic syndromes, infertility and spontaneous abortions. (Barbara H. Czepulkowski 2001; Baht and Wani 2017)

The structural features and the architecture of the human genome can induce genomic instability due to the presence of specific regions with high susceptibility to mutations and rearrangements. (Lupski and Stankiewicz 2005)

Is important to note that certain parts of some chromosomes may show variation in population, a phenomenon known as heteromorphism. Watson-Crick base pair changes, that originate single nucleotide polymorphisms (SNPs), and by copy number variations (CNVs) are causes of Heteromorphism. (Lupski and Stankiewicz 2005; Zhang et al. 2009)

Chromosomal rearrangements occur by both homologous and non-homologous recombination mechanisms. Homologous mechanisms appear to be the predominant pathway for the formation of genomic rearrangements. (Lupski and Stankiewicz 2005)

Chromosomal abnormalities, in the form of deviations from the normal state of diploidy or in the form of structural rearrangements may lead to genetic imbalance, ending up in a variety of syndromes. (Baht and Wani 2017) Chromosomal abnormalities are classified as numerical or structural and each one of them be constitutional or acquired. An abnormality is considered constitutional if it arises during the embryonary development, being present in cells throughout the organism, and it is considered acquired if the individual was born without any cytogenetic alteration, having a normal karyotype and the chromosomal alteration appears in a subset of cells or tissues. Acquired chromosomal alterations are characteristic from processes like tumorigenesis. (Barbara H. Czepulkowski 2001; Gardner, R.J. Mckinlay; Sutherland, GranT R.; Shaffer 2012)

If an abnormality is large enough to be visible using the light microscope, being bigger than 4Mb in length, it is termed a cytogenetically visible. (Barbara H. Czepulkowski 2001) The different forms of observed chromosomal alterations, numerical and structural, are summarized in figure 1.1.
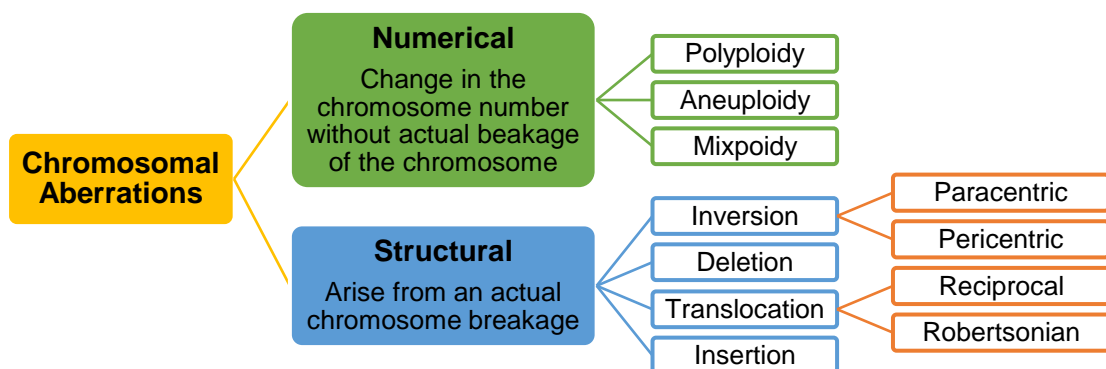
Figure 1.1 – Classification of Chromosomal Aberrations.

Like previously said chromosomal rearrangements may lead to genomic imbalance. The pathogenic mechanisms that arise from chromosomal abnormalities, are:

1. Gene dosage effect, caused by a lack (deletion, disruption of a gene) or excess (duplication) of chromosomal material, whether for a whole chromosome or a part of a chromosome;

2. Meiotic impairment, due to incongruent paternal origin of a chromosome or chromosomal segment;

3. A position effect, whereby a gene loss its normal functions when placed in a new chromosomal environment, this is caused for example by a translocation;

4. Combinations of the aforementioned. (Gardner, R.J. Mckinlay; Sutherland, GranT R.; Shaffer 2012; Baht and Wani 2017)

Genomic disorders are a class of human conditions that result from rearrangements of the human genome rather than from DNA sequence base changes. Those rearrangements convey traits or susceptibility to traits. (Lupski 2009)

1.2.1. Numerical Chromosomal abnormalities

Numerical chromosomal changes involve an alteration of chromosome number without actual breakage of the chromosome. These are subdivided in three different classes of alterations:

- Polyploidy (more than two paired homologous sets, this condition is not compatible with life);

- Aneuploidy (extra copies of a single chromosome or loss of one of both homologous);

- Mixploidy (someone who possesses two or more genetically different cell lines from a single zygote or occasionally because of chimerism). (Barbara H. Czepulkowski 2001)

1.2.2. Structural Chromosomal abnormalities

Structural chromosomal alterations are the result of chromosomal breakage in one or more chromosomes. There are lots of ways in which a chromosome can be altered after a breakage, so there are a large variety of structural chromosome changes. Breaks near the centromere are rapidly repaired by reparation enzymes. When the chromosome breaks in distal fragments terminal deletions are formed, because repair enzymes are not able to fix the breakage. (Barbara H. Czepulkowski 2001)

Breaks in more than one chromosome tend to disrupt the correct function of repair enzymes, which tend to join the wrong ends of chromosomes, being that the basis of the remaining structural chromosome anomalies. (Barbara H. Czepulkowski 2001) There are four types of alterations in chromosome structure, such as, deletions (del), duplications (dup), inversions (inv) and translocations (t). (Baht and Wani 2017)

An **inversion** is the result of chromosomal breakage in two different places followed by a 180 degree (º) rotation of the chromosomal fragment, and its re-joining to the remaining chromosome, where the shape and size of the original chromosome are conserved, only the position of the centromere can be altered. Inversions are subdivided in **paracentric**, if only fragments of the chromosomal arms are inverted or **pericentric** if the centromere is also inverted. Inversions cause little problem for an organism

under homozygous condition, while heterozygotes often have pairing difficulties in meiosis, leading to formation of inversion loops. (Barbara H. Czepulkowski 2001; Baht and Wani 2017)

A **deletion** represents a missing segment on a chromosome. It occurs when chromosomal arms split, forming a chromosomal fragment and the remaining chromosome re-joins. The chromosomal fragment formed from the breakage is posteriorly degraded. (Barbara H. Czepulkowski 2001; Baht and Wani 2017)
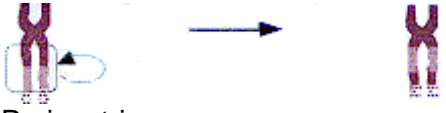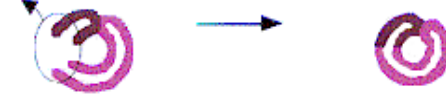
A **ring chromosome** is formed in a situation of breakage occurs in either side of the centromere, and the two ends of the segment are joined together, forming a circular chromosomal fragment. (Barbara H. Czepulkowski 2001)

Translocation phenomenon takes place in a situation of breakage in two or more chromosomes, at the same time, and fragments from different chromosomes are relocated and re-joined to other chromosomes, resulting in abnormal chromosomes with genetic material from more than one chromosome. Translocations are subdivided in distinct classes: **reciprocal translocations** (translocations where the broken acentric fragments, distal to the two breakpoints, are exchanged) and **centric or Robertsonian translocations** (two acrocentric chromosomes break near the centromere and the large fragments of these chromosomes fuse together.). (Barbara H. Czepulkowski 2001; Baht and Wani 2017)

Insertions are the result of three breaks, two in one chromosome, producing an acentric fragment, which is then, inserted at a break formed in another chromosome. (Barbara H. Czepulkowski 2001). Table 1.1 summarizes and illustrates the different types of structural chromosomal aberrations.

Table 1.1 – List of most common structural chromosomal abnormalities.

Two Breaks in a single chromosome	Two or more breaks in different chromosomes
Inversion Paracentric Pericentric	Reciprocal Translocation
Interstitial Deletion	Centric (Robertsonian) Translocation
Ring Chromosome	Insertion

Complex chromosomal rearrangements (CCRs) are rare, structural chromosomal abnormalities involving three or more cytogenetic breakpoints located on two or more chromosomes. (Hemmat et al. 2014; Pan et al. 2016)

CCRs can be balanced (if there is no gain or loss of chromosomal material) or unbalanced (if there is gain or loss of chromosomal material), and are usually associated with significant risk of mental retardation and phenotypic anomalies attributable to gene disruption, cryptic imbalances, or position effects. (Hemmat et al. 2014)

Balanced chromosomal rearrangements (BCR) occur sporadically in the population or are segregate within families, with a frequency of 1 in every 2000 live births. The risk of developing congenital anomalies or neurodevelopmental disorders (NDDs) has been estimated to be 6.1 percent (%) for a *the novo* BCR. (Utami et al. 2014)

Up to 5% of reference haploid human genome is copied in DNA fragments bigger than 1 kilo-base pair (Kb) in size and with higher than 90% sequence identity to the human genome reference sequence. These fragments are termed Low copy repeats (LCRs). LCRs longer than 10Kb and of over approximately (~) 97% sequence identity can lead to local genomic instability, and stimulate CNV formation. (Stankiewicz and Lupski 2010)

A CNV is a structural alteration that implies an imbalance on human genome. Its length may vary between 1Kb to several mega-base pair (Mb). CNVs perturb the normal biological balance of the diploid state at any given locus. Deletions, duplications, triplications, insertions, or translocations can result in CNVs. CNVs can be inherited or sporadic, and large *de novo* CNVs are thought more likely to be disease causative than the small ones. Phenotypic effects of CNVs are sometimes unclear and depend mainly on whether dosage-sensitive genes or regulatory sequences that are affected by the genomic rearrangement. (Stankiewicz and Lupski 2010)

1.3.   Initial considerations about Congenital Anomalies

Congenital anomalies, also known as birth defects, congenital disorders or congenital malformations, can be defined as structural or functional anomalies that occur during intrauterine life and that can be detected prenatally, at birth or sometimes only later in infancy. (World Health Organization 2017) Major Congenital Anomalies are those which are lethal, carry high mortality or have other serious medical functional consequences. Congenital malformations are a major cause of early spontaneous abortions, because some foetus malformations are incompatible with intrauterine survival. (Dolk et al. 2011; World Health Organization 2017)

Approximately 50% of all congenital anomalies cannot be linked to a specific cause, although there are some identified causes or risk factors associated with the development of these defects. (World Health Organization 2017) Both genetic and environmental factors are involved in the causation of congenital anomalies. (Dolk et al. 2011)

Genes play an important role in many congenital abnormalities, through parentally inherited genomic alterations, or through de novo mutations in genes. (World Health Organization 2017)

Structural chromosomal rearrangements, including CCRs, have been associated with multiple congenital abnormalities, including malformative syndromes and global developmental delay. (Houge et al. 2003; Kloosterman and Hochstenbach 2014). Usually there is a correlation between the number of chromosomal breakpoints and the severity of the phenotype, although is the genomic location of these breakpoints that accounts most for the phenotypical consequences. (Houge et al. 2003; Kloosterman and Hochstenbach 2014)

A genetic abnormality is diagnosed by the clinician with the help of genetic tests and family history, where the results of the genetic tests explain why the children is malformed. Genetic syndromes, associated with chromosomal anomalies, microdeletions, single gene mutations or by genetic imprinting, account for less than one fifth of the total reported cases of congenital anomalies. (Dolk et al. 2011)

The Maternal exposure to certain pesticides and other chemicals, as well as certain medications, alcohol, tobacco, radiation and some pathogens during pregnancy, may increase the risk of having a foetus or neonate affected by congenital anomalies. Advanced maternal age is also a risk factor for abnormal intrauterine fetal development. (World Health Organization 2017)

In the vast majority of cases, the cause cannot be identified as one single factor, these include congenital anomalies of multifactorial origin. In these cases many genetic and environmental factors contribute additively, such that, the individual embryo/foetus surpasses a "threshold" beyond which it can no longer self-regulate to follow the normal developmental pattern. (Dolk et al. 2011)

The main challenge for genetic research is the interaction between genetic factors and environmental factors, such that, specific environmental exposures are only teratogenic in the presence of specific predisposing genetic factors. (Dolk et al. 2011)

European surveillance of congenital anomalies (EUROCAT) is the network of population-based registers of congenital anomalies in Europe, with a common protocol and data quality review, covering 1.5 million annual births in 22 countries. (Dolk et al. 2011) EUROCAT recorded a total prevalence of major congenital anomalies of 23.9 per 1000 births for 2003 to 2007, in which 80% were live births. The prevalence of chromosomal anomalies was 3.6 per 1000 births. Congenital heart defects (CHD) were the most common non-chromosomal subgroup, at 6.5 per 1000 births, followed by limb defects (3.8 per 1000) anomalies of urinary system (3.1 per 1000) and nervous system defects (2.3 per 1000). (Dolk et al. 2011)

1.3.1. Congenital Heart diseases

Congenital heart defects (CHD) are problems linked to the heart's structure that are present at birth. These defects can involve the interior walls of the heart, the valves inside the heart or the arteries and veins that carry blood to the body or to the heart. CHD might change the normal flow of blood through the heart. (National Heart)

CHD are the major class of birth defects and accounts for more deaths in the first year of life than any other condition (when infectious aetiologies are not counted). (Richards and Garg 2010). There are

many types of CHD that range from simple defects (with no symptoms) to complex defects (with severe, life-threatening symptoms). (National Heart) The incidence of these type of malformation is 19-75 per 1000 live births, and are also present in an even greater proportion of miscarriages. (Richards and Garg 2010)

Although medical and surgical techniques for diagnosis and treatment of CHD have advanced through time, CHD aetiology is not completely understood. (National Heart; Richards and Garg 2010). Classical studies have found that CHD is multifactorial, accounting with genetic predisposition and environmental influence. (Richards and Garg 2010) However, sequencing the human genome and advances in molecular techniques led to increasing evidence of a stronger role of the genetic factors. (Richards and Garg 2010)

Non-genetic aetiologies of CHD include environmental teratogens (dioxins, polychlorinated biphenyls, pesticides), maternal exposures to alcohol, isotretinoin, thalidomide or anti-seizure medications, and infectious agents like rubella. Anti-retroviral and the epidemic of obesity with associated phenotypes of diabetes and hypercholesterolemia are nowadays recognized as emerging risk factors for CHD. (Akl C. Fahed et al. 2014)

Initial human genetics methodologies, like G-banding karyotyping, had poor resolution, which restricted analyses to inherited forms of CHD, and lead to a defective identification of the genes that might be linked to the presence of heart defects. (Akl C. Fahed et al. 2014; American Heart Association 2017)

Most children with CHD don't have other birth defects. 25 to 40% of cases of CHD are associated with other anomalies or as part of an identified genetic syndrome. About 30% of children with chromosomal abnormalities have CHD. (Richards and Garg 2010)

Interatrial communications are the most common of the congenital heart defects, accounting for 6-8% of malformed hearts. (Orphanet 2017) Having a communication between atrial chambers does not imply having a deficiency on atrial septum 3/4 of all cases are the ostium secundum defect, characterized by a large aperture on foramen ovale site. Ostium primum defect accounts for 1/6 of defects, and constitutes an atrioventricular septal defect with shunting only at atrial level. 1/10 of the cases are classified as sinus venosus defect, that is a hole outside the confines of the oval fossa, being frequently associated with anomalous connection of the right pulmonary veins. Coronary sinus defect, is the rarest defect. It is due to absence of the walls that separate the sinus from the left atrium, being found at the mouth of the coronary sinus. (Orphanet 2017)

Agenesis of the left pulmonary artery branch is caused by a malformation of the sixth aortic arch during embryonary development, and is frequently associated with other CHD like atrial septal defects. (Emren et al. 2015) Usually a consequence of this defect is left pulmonary hypoplasia, which is characterized by incomplete development of the lung tissue, and could be the base of the agenesis of the left lung, since it is caused by the failure of the development of the lung bud during embryogenesis. (Singh et al. 2015; Emren et al. 2015)

1.3.1.1. Chromosomal rearrangements and Genes associated with CHD

In the past couple of decades, there has been a greater understanding on molecular pathways regulating cardiac development. Through gene target technology mouse models of a multitude of CHD were created, which led to the identification of transcriptional regulators, signalling molecules and structural genes critical for normal cardiac morphogenesis. Multiple genes were also identified as being controlled by these highly conserved molecular pathways. These investigations have assisted in the identification of the genetic aetiology of CHD, and provide evidence that many genes may have etiologic roles in CHD. (Richards and Garg 2010)

Nowadays, there are several methods available for genetic research of CHDs, including cytogenetic analysis, linkage and association studies, CNV and DNA micro-array analysis, and whole exome sequencing. The use of new technologies has increased the possibility of identification of new genes and chromosomal loci in syndromic and non-syndromic CHDs. (Grollmuss et al. 2016)

Each chromosomal anomaly is preferentially associated with specific types of CHDs. (Grollmuss et al. 2016) Numerical alterations in chromosomes account for a significant portion of CHD cases. Chromosomal aneuploidy, the first recognized genetic cause of CHD, continues to be a major aetiology today. 50% of individuals born with trisomy 21 have CHD, ranging from atrial and ventricular septal defects to atrioventricular canal lesions; 80% of individuals born with trisomy 13 have CHD, with heterotaxy and laterality defects becoming more common; nearly all individuals born with trisomy 18 have CHD; ~1/3 of females with monosomy X (Turner syndrome) have CHD, most common diagnoses include bicuspid aortic valve, aortic stenosis, hypoplastic left heart syndrome and coarctation of the aorta; ~1/5 of males 47, XXY (Klienfelter syndrome) have CHD, with patent ductus arteriosus and atrial septal defects prevailing. (Richards and Garg 2010)

Submicroscopic chromosomal anomalies associated with syndromic CHDs are the 22q11.2 deletions, causative of DiGeorge/velocardiofacial with conotruncal heart defects, the deletions on 7q11.23 causative of Williams syndromes with supravalvular aortic stenosis, the 5p deletions causative of cri-du-cha syndrome, with atrial and ventricular septal defects and patent ducts arteriosus, the 4p deletions causative of Wolf-Hirschhorn syndrome, the 8p deletion syndrome, the 10p deletions, the 11q deletions causative of Jacobsen syndrome, the deletion and duplication on 1q21.1, the 1p36 deletion syndrome and the deletions on 20p12 causative of Alagille syndrome, all with different combination of various cardiac defects. (Pierpont et al. 2007; Grollmuss et al. 2016)

It is important to note that interatrial communications, are the cardiac defect most shared between the previous identified genomic syndromes.

The developments on the technologies for identification of the genetic cause of CHD demonstrate that 2/3 of the cases of interatrial communications are the result of spontaneous mutations on the genes actin, alpha cardiac muscle 1 (*ACTC1*), Cbp/p300 interacting transactivator with Glu/Asp rich carboxy terminal domain 2 (*CITED 2*), GATA binding protein 4 (*GATA4*), GATA binding protein 6 (*GATA6*), myosin heavy chain 6 (*MYH6*), NK2 homeobox 5 (*NX2-5*), T box 2 (*TBX20*) and toiled like 1 (*TLL1),* being those recognized as susceptibility genes for this defect. (Richards and Garg 2010; Orphanet 2017)

These developments on the technologies genomic technologies turned possible to identify that single-gene defects can lead to isolated CHD and reveal more about molecular pathways important for cardiac morphogenesis. (Richards and Garg 2010) On table 1.2 are summarized the genes associated to single-gene mutation induced CHD and the respective phenotype observed on patients.

Table 1.2 – Genes causative of congenital heart defects. (adapted from *"Genetics of Congenital Heart Disease"* - *Richards and Garg 2010)*

Phenotypes	Gene
Atrial septal defect, atrioventricular conduction delay, tetralogy of Fallot, tricuspid valve abnormalities	NKX2.5
Atrial septal defect, ventricular septal defect	GATA4
Atrial septal defect, hypertrophic cardiomyopathy	MYH6
Cardiac septation defects associated with pulmonary hypertension	BMPR2
Endocardial cushion defects	CRELD1, ALK2
Bicuspid aortic valve, early valve calcification	NOTCH1
D – transposition of the great arteries	PROSIT-240

It is not easy to define precisely the genetic defects underlying non-syndromic CHDs, due to the genetic and clinical heterogeneity of these malformations. (Grollmuss et al. 2016)

Through genotype-phenotype observations were concluded that CHD are not due to a global change in genomic content but rather from altered dose of specific genes. This concept gained more clarity with the development of methodologies, like array comparative genomic hybridization (aCGH) and whole genome sequencing, to define sub- chromosomal changes in genome structure, like CNVs. (Akl C. Fahed et al. 2014) Although is not easy to distinguish between pathogenetic CNVs and benign polymorphic variants, it was observed that the vast majority of individuals with CHD don't have only single gene defects, but have unique CNVs associated with cardiac malformations. (Richards and Garg 2010; Grollmuss et al. 2016)

While some of these CNVs occur in regions known to contain gene critical for cardiac development, like CNVs on 8p23.1 altering *GATA4*, on 20p12.2 altering jagged 1 (*JAG1*) and on 9q34.3 altering notch1 (*NOTCH1*), the majority occurred where no cardiac genes were located.(Richards and Garg 2010; Grollmuss et al. 2016) On table 1.3 are summarized some of the loci and respective genes where these CNVs were found.

Table 1.3 – Most recurrent CNVs associated with non-syndromic congenital Heart Defects. (adapted from "Genetics of Congenital Heart Disease: The Glass Half Empty" - Akl C. Fahed et al. 2014.)

Locus	Size Range (Kb)	Genes
1q21.1	418-3,981	PRKAB2, FM05, CHD1L, BCL9, ACP6, GJA5, CD160, PDZK1, NBPF11, FMO5, GJA8
3p25.1	175-12,380	RAF J, TMEM40
3q22.1-3q26.1	680-32,134	FOXL2, NPHP3,FAM62C, CEP70, FAIM, PIK3CB, FOXL2, BPESC1
4q22.1	45	PPM1K
5q14.1-5q14.3	4,937-5,454	EDIL3, VCAN, SSBP2, TMEM167A
5q35.3	264-1,777	CNOT6, GFPT2, FLT4, ZNF879, ZNF345C, ADAMTS2, NSD1
7q11.23	330-348	FKBP6
8p23.1	67-12,000	GATA4, NEIL2, FDFT1, CSTB, SOX7
9q34.3	190-263	NOTCH1, EHMT1
11p15.5	256-271	HRAS
13q14.11	555-1,430	TNFSF11
15q11.2	238-2,285	TUBGCP5, CYFIP1, NIPA2, NIPA1
16p13.11	1,414-2,903	MYH11
18q11.1-18q11.2	308-6,118	GATA6
19p13.3	52-805	MIER2, CNN2, FSTL3, PTBP1, WDR18, GNA11, S1PR4
Xp22.2	509-615	MID1

1.3.2. Neurocognitive and Neurodevelopmental Deficiencies

Cognition refers to the mental processes or thinking skills that allow people to learn and function in daily life, in society. These are the skills needed to process information, think, read, understand and solve problems. The main types of cognition include: attention, memory, processing information, solving problems, and planning/organizing. (Health)

NDDs are a group of conditions with onset on developmental period, with typical manifestation in early stages of development, and are characterized by developmental deficits that traduce impairments on personal, social, educational and occupational functioning. The range of these deficits varies from specific limitations, of learning, control or executive functions to global impairments of social skills or intelligence (American Psychiatric Association 2013)

Intellectual disability (ID), or Mental Retardation, is characterized by significant limitations in both intellectual functioning and in adaptive behavior, which covers many everyday social and practical skills. This disability originates before the age of 18. (American Association on Intellectual and Developmental Disabilities 2017)

It affects about 1-2% of the general population worldwide. ID is a broad diagnosis encompassing a wide variety of overlapping phenotypes and severities, it can range from mild to profound and can occur as an isolates phenotype or can be associated with either clinical symptoms or as part of a syndrome. (Vaidutis et al. 2016) Usually individuals with autism spectrum disorders also present intellectual disability. (American Psychiatric Association 2013)

The causes of ID vary with the severity of the disorder: - moderate to severe cases (Intelligence Quotient (IQ) score <50) are much more likely to be due to a single pathological cause than are mild cases (IQ score 50-70) which are thought to be multifactorial in origin. Chromosomal and genetic disorders account for 30-40% of causes of moderate to severe mental retardation, environmental factors explain a further 10-30% and 40% of the cases have unknown causes. (Knight et al. 1999)

ID is often overlooked because of its association with other neuropsychiatric disorders such as autism, epilepsy and schizophrenia, a fact that also contributes to its near invisibility in global health programs. (Ropers 2010)

It is estimated that ~1% of the human population has an autism spectrum disorder (ASD), being diagnosed in a 4/1 ratio male/female. (Zoghbi and Bear 2012; Samsam et al. 2014) ASD is characterized by persistent deficits in social communication and interaction across multiple contexts, excessively repetitive behaviors, restricted interests and insistence on sameness. (American Psychiatric Association 2013) The molecular pathogenesis of ASD is not already fully understood, since it has the contribution of a multitude of factors, constituting a multifactorial disease. (Samsam et al. 2014)

The evaluation of genetic causes of neurodevelopmental disorders remains challenging because these conditions are generally heterogeneous with many different genetic alterations resulting in clinically indistinguishable phenotypes. (Vaidutis et al. 2016)

1.3.2.1. Chromosomal Rearrangements and genes associated with Neurodevelopmental Disorders

Cytogenetically visible chromosome aberrations account for ~15% of all patients with severe ID. Submicroscopic deletions and duplications have recently emerged as an equally important, novel cause of ID. These are too small to be detectable under the microscope, and their identification is possible nowadays thanks to evolution of Molecular Cytogenetic Techniques and Next Generation Sequencing (NGS). (Ropers 2010) The usage of NGS methodologies like extensive gene panels, exome sequencing and the whole genome sequencing, allowed new ID genes to be identified in rapid succession. (Vissers et al. 2015)

There is some evidence that small chromosomal rearrangements involving the subtelomeric regions of chromosomes are also important causative of ID, being identified mostly on ID sporadic forms. (Rafati et al. 2012)

Nowadays, ASD and ID are named "synapsopathies", because several mutations on genes associated with synaptic structure and function were found to be causative of these disorders. The most common single-gene mutations described for ASD with ID are found on fragile X mental retardation 1 (*FMR1*), tuberous sclerosis 1 (*TSC1*), tuberous sclerosis 2 (*TSC2*), Neurofibromin 1 (*NF1*), ubiquitin protein

ligase E3A (*UBE3A*), methyl CpG binding protein 2 (*MECP2*), phosphatase and tensin homolog (*PTEN*) and SH3 and multiple Ankyrin repeat domains 3 (*SHANK3*) genes, being those genes associated to several syndromes. (Zoghbi and Bear 2012) Rare mutations on neuroligins, neuroligin 3 (*NLGN3*), neuroligin 2 (*NLGN2*) and on neurexin 1, neurexin 1 (*NRXN1*) were also described as being a cause of autism. (Zoghbi and Bear 2012)

Banerjee and his co-workers decribed the *de novo* translocation, t(1;2)(q31.3;p16.3)dn and the paternally inherited insertion, ins(16;2)(q22.1;p16.1p16.3)pat was being associated with ASD.

The presence of CNVs encompassing a variety of different genomic regions has been shown to be strongly associated with diseases such as autism, schizophrenia and developmental delay, along with being identified in individuals with multiple anomalies. (Richards and Garg 2010)

Performing comparative genomic hybridization array (aCGH) to scan the genome of autists and their healthy siblings', has shown *de novo,* but individually rare, CNVs are significantly associated with autism. (Richards and Garg 2010) And more recently the microdeletion or reciprocal microduplication of 16p11.2 and 15q13 were identified to be also significantly associated with autism, being present in more than 1% of the cases. (Richards and Garg 2010; Banerjee et al. 2014)

Similar studies made about schizophrenia found out that the CNVs detected in schizophrenic patients modify genes involved in neurodevelopment. Multiple studies detected that abnormal chromosomal copy number is also associated with mental retardation associated with dysmorphism. (Richards and Garg 2010)

Proving pathogenicity of mutations observed in patients with intellectual disability and establishing the candidate ID genes (ID-causing gene when mutated) is still complicated, even in this post-NGS era, since it is a polygenic disability maybe with hundreds of genetic loci giving small contributions to the phenotype. (American Psychiatric Association 2013) Replication, which consist in the observation of multiple unrelated but phenotypically similar patients with mutations that are predicted to be damaging in the same gene, is still a key requirement in the process of identifying the candidate ID genes. (Vissers et al. 2015)

The combination of CNV morbidity maps with lists of candidate genes that are affected by point mutations can be particularly powerful in this respect because both have an important role in ID. On figure 1.2 is shown a map of genic overlap for neurodevelopmental disorders obtained by statistical association of CNVs morbidity maps and disease by comparison of the frequency of these types of mutations in cases and controls.
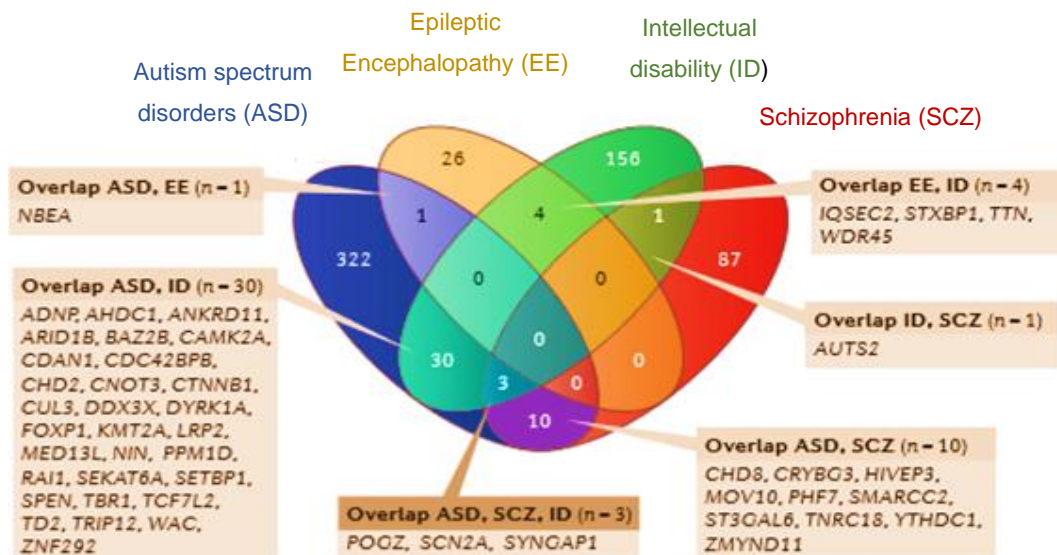
Figure 1.2 – Map of genic overlap for neurodevelopmental disorders obtained by replication process and combination of CNV morbidity maps. (adapted from "Genetic studies in intellectual disability and related disorders" - Vissers et al. 2015)

1.4.   Methods of identification of Chromosomal abnormalities

The characterization of constitutional chromosomal rearrangements is important for identification of the genetic aetiology of congenital abnormalities, and it is also fundamental for functional annotation of human genome. (David et al. 2013)

The combination of several molecular cytogenetic techniques with NGS constitutes a powerful methodology for characterization of chromosomal rearrangements. (Pan et al. 2016)

The advances on genomic technologies allows precise delineation of structural chromosomal rearrangements at nucleotide level, and since its results are becoming increasingly feasible, those techniques provide major improvements on genetic counselling. (Ordulu et al. 2014; Pan et al. 2016)

Karyotyping, aCGH and single nucleotide polymorphism array (array SNP) are the first-tier diagnostic tools to investigate chromosomal rearrangements. (Utami et al. 2014) BCRs are typically unidentified by low resolution methods, such as karyotyping. And since BCRs do not result in large gains or losses of genetic material at the breakpoint, molecular cytogenetic techniques, like microarrays, which detect CNVs at a considerable resolution fail in the identification on copy number neutral rearrangements. (Talkowski et al. 2011; Utami et al. 2014)

Most of times, interpretation of a rearrangement's clinical impact is based upon a broad chromosome region, insensitive to the specific gene(s) disrupted or deregulated and neglecting the sequence complexity that might underlie these rearrangements. (Talkowski et al. 2011) Nowadays, with the development of molecular cytogenetics techniques and the advent of NGS, it is possible to define chromosomal breakpoints at nucleotide resolution, and different approaches can be combined as shown in figure 1.3 in order to better understand the architecture of chromosomal rearrangements.
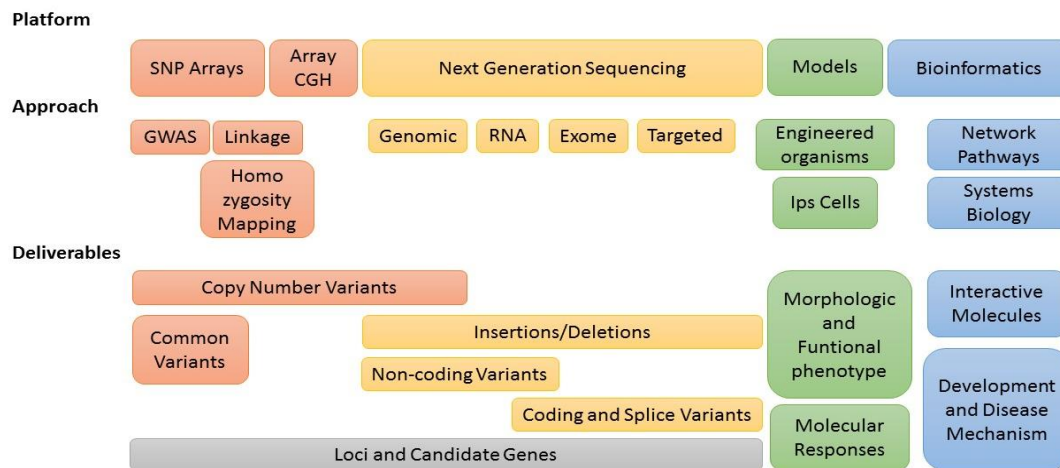
Figure 1.3 – Summary of the strategies used nowadays to define the genetic architecture of chromosomal rearrangement after Karyotyping. (adapted from *"Genetics of Congenital Heart Disease: The Glass Half Empty"* - Akl C. Fahed et al. 2014.)

1.4.1. Classical Cytogenetics

Conventional cytogenetics consists in the analysis of chromosomes through microscopy techniques. It allows scanning the genome in the pursuit for aberrations that involve both gains and losses of portions of the genome, as well as rearrangements within and among chromosomes. (Pinkel and LI 2005)

Karyotyping, is one of the most common and easily accessible assays used to determine the chromosomal constitution of human cells. (Bates 2011) However it is unable to resolve structural rearrangements to the nucleotide level, since its overall resolution is greater than 5Mb, and its breakpoint determination resolution varies to 5 to 15Mb, which makes it difficult to analyse the actual pathological burden of the rearrangement. (Ordulu et al. 2014)

The analysis by karyotyping is only possible recurring to banding techniques, which consists in staining of metaphasic or pro-metaphasic chromosomes with a suitable dye like Giemsa. Staining patterns produced on the chromosomes are sometimes ambiguous, and the resolution is limited, both by the optical characteristics of microscopes and the complex manner in which DNA is packaged into chromosomes. (Pinkel and LI 2005)

Although banding techniques represent the central theme in every cytogenetics laboratory it is mandatory to recur to methodologies of genomic analysis with better resolution and less limitation, since specific identification of disrupted genomic region(s) is critical in diagnosis and management of constitutional and acquired rearrangements, especially as annotation of the human genome accelerates. (Wan 2010; Ordulu et al. 2014)

1.4.2. Molecular Cytogenetics

Molecular cytogenetics arises in the context of clinical cytogenetics in order to facilitate the study of genetic disorders by the combination of classical cytogenetic and molecular techniques like Fluorescence in situ hybridization (FISH) and High resolution chromosomal microarray. (Pinkel and LI 2005)

FISH analysis was managed to overcome many of the limitations of karyotyping, it is sensitive rapid and a critical complement to conventional cytogenetics. It is based on the capacity of single stranded DNA sequences hybridizes with a complementary DNA sequence, and consists in the use of fluorescent specific probes for specific genetic loci, such as telomeres, centromeres or sites of known gain or loss of DNA sequences. It turns possible the detection of both numerical and structural chromosome alterations. (Pinkel and LI 2005; Wan 2010)

In diagnosis laboratory, the most useful FISH probes system is: centromere probes used for chromosomic enumeration (useful for the detection of numerical chromosomal alterations), chromosome painting probes (useful in deciphering major chromosomal rearrangements) and locus specific gene probes for gene fusions. (Wan 2010)

Although, FISH is an improvement to classical cytogenetic analysis, the complexity of the staining pattern that is produced is limited by the number of FISH probes that can be distinguished, and it has the same limitations of optical microscopy and chromosome structure that affect chromosome banding. (Pinkel and LI 2005)

Aberrations that involve gain or loss of chromosome segments are particularly important in medical cytogenetics. (Pinkel and LI 2005) Comparative genomic hybridization (CGH), is a technique based on quantitative tow colours fluorescence *in-situ* hybridization, and it allows the detection of genetic imbalances throughout the screening of the entire genome. (Pinkel and LI 2005; Wan 2010)

Using microarrays, the resolution of CGH can be much greater than with standard cytogenetics, by substitution of the hybridization targets on metaphasic chromosomes with genomic segments. These genomic segments can be bacterial artificial chromosome (BACs) or P1 artificial chromosome (PACs) clones for hybridization targets. Array CGH is a technology for "molecular karyotyping" with a resolution of 100Kb to 1Mb. This resolution can be improved by overlapping clones, turning possible to detect gains or losses as small as 40-80Kb. (Wan 2010)

The resolution and coverage of array CGH are dependent on the density of the array used. Microarrays with dense coverage of the entire genome allow screening for aberrations at any location without the need to have prior knowledge about where to test. (Pinkel and LI 2005; Wan 2010)

Thus, array CGH offers the ability to screen genomes of affected individuals to discover new aberrations, determine if apparently balanced translocations also involve gains or losses of DNA, establish phenotype-genotype relationships for conditions with variable aberrations, and provide comprehensive clinical diagnostic information. In combination with techniques such as chromosome microdissection or chromosome sorting, array CGH can map translocation breakpoints with high precision. (Pinkel and LI 2005)

SNP arrays consist in other methodology applied for the detection of CNVs. These contain 25-mer oligonucleotides originally designed to assess SNPs. This method has the advantage of measuring loss of heterozygosity (LOH) alongside with CNVs, and also allows sensitive and specific detection of single copy number changes at sub-microscopic level throughout whole genome. (Wan 2010)

FISH and/or High resolution chromosomal microarray studies have identified cryptic CCRs as a cause of abnormal phenotype in a significant number of patients with apparently balanced chromosomal rearrangements. (Hemmat et al. 2014)

1.4.3. Next Generation Cytogenetics

"Next generation cytogenetics" is an integration of traditional cytogenetic techniques and the NGS. (Ordulu et al. 2014)

Sequencing breakpoints of structural chromosome rearrangements has been possible since 1980s, however, precise localization of these aberrations to the nucleotide level in a genome wide context only became feasible with recent improvements of sequencing techniques. (Ordulu et al. 2014)

NGS is a powerful tool for genomics research. It has allowed rapid characterization of genomes, exomes and transcriptomes. Whole-genome sequencing (WGS) is the most comprehensive method for analysing the genome. (Liang et al. 2014; illumina 2017)

Although the cost of performing WGS has decreased in recent years, it remains costly compared with other methods like exome and ribonucleic acid (RNA) sequencing (RNAseq). This fact triggered for the need of identifying an alternative WGS strategy for identifying the breakpoints that characterize structural variants and copy number changes. (Wan 2010)

Analysing the whole genome using NGS delivers a base-by-base view of all genomic alterations, including single nucleotide variants (SNV), insertions and deletions, copy number changes, and structural variations. (illumina 2017)

Several different approaches were described to perform massively parallel sequencing of large fragments, such as large-insert sequencing, mate-pair sequencing, and jumping libraries, which all represent similar approaches to sequencing large genomic inserts. (Talkowski et al. 2011)

Next generation paired-end whole-genome sequencing involves sequencing both ends of a DNA fragment with predetermined size and yields millions of paired short reads from the ends of these fragments. This increases the likelihood of alignment to the reference genome and facilitates detection of genomic rearrangements, repetitive sequences, and gene fusions. (Talkowski et al. 2011; Liang et al. 2014)

The application of this methodology requires the preparation of is the use of Illumina's mate pair library. The standard protocol requires 10 milligrams (mg) of genomic DNA and supports the evaluation of regions spanning up to ~2–5Kb. (Liang et al. 2014)

1.5. C-technologies and Studies of 3D organization of Human Genome

As referred previously, the long strands of DNA of every cell's genome are package into chromatin in a very confined nuclear volume. The organization of human genome is extremely important to biological function at the gene level, as well as at global nuclear level. (Belton et al. 2013) The spatial organization of chromatin has been a topic of study for many years since, chromatin conformation and long range association between genes and other distal elements are thought to play important roles in gene

expression, regulation and other genomic activities. (Belaghzal et al. 2017) Unfortunately, the understanding of higher order genomic structure is coarse, fragmented or incomplete. (Dixon et al. 2012)

Preforming studies about three-dimensional (3D) conformation and organization of chromatin allow the better understanding of spatial aspects of gene regulation, chromosome morphogenesis and genome stability and transmission. Also enhance the understanding of the biophysics of chromatin and further enable the investigation of pathologies related to genome instability or nuclear morphology. (Belton et al. 2013)

Recent Studies have revealed the existence of millions of potential cis-regulatory elements in human genome, with a great number of them residing in intergenic regions and away from their target gene promoters. (Schmitt et al. 2016) These elements, which largely consist in enhancers, influence the transcription of target genes through looping of chromatin fibbers during animal development, which has been described for many enhancers. (Schmitt et al. 2016)

There are many techniques available to observe the spatial organization of chromatin and these are classified into microscopic and molecular assays. (Belton et al. 2013)

Electron microscopy is used to analyse the architecture of the nucleus in nanometre (nm) scale detail, while fluorescent (light) microscopy provides information on the shape and distribution of specific chromosomes and chromosomal loci, as well as the co-association of specific loci with sub-nuclear compartments, with resolution of 50-100nm. (Belton et al. 2013)

Microscopic techniques have the major disadvantage of lack of connectivity to the genomic sequence, i.e. incredible fine resolution architecture cannot be assigned to a specific location in the genome. (Belton et al. 2013). The use of sequence specific fluorescent probes in FISH assays allows to go through limitations of microscopy, but these methods are still limited in throughput, allowing the analysis of only a few loci simultaneously. (Belton et al. 2013)

Molecular assays are based in the concept that dense matrices of chromatin interactions could be used to determine the spatial organization of chromatin domains, chromosomes and ultimately entire genomes. (Belaghzal et al. 2017) This concept was explained with the development of the methodology of chromosome conformation capture (3C) where the genomic sequence itself is the output, completely connecting chromosome structure and genomic sequence. (Belton et al. 2013; Belaghzal et al. 2017)

3C and its derived 3C-based technologies are commonly used for studying chromatin interactions in eukaryotic cells. These techniques have uncovered general features of genome organization, which include the existence of hierarchical chromatin structures such as compartments, topologically associating domains (TADs), sub-TADs, insulted domains and chromatin loops. (Schmitt et al. 2016)

C-technologies measure the population-average frequency at which two DNA fragments physically associate in 3D space. (Belton et al. 2013) These methods combine protein crosslinking and proximity ligation of DNA to detect long range chromatin interactions between pairs of genomic loci. (Schmitt et al. 2016) Briefly nuclei are isolated from cultured cells, following are treated with formaldehyde, which crosslinks the chromatin proteins to their associated DNA to fixate the chromatin structure. (Schmitt et al. 2016) An efficient fixation of those interactions is crucial for the success of the analysis. The

crosslinked DNA is then digested using restriction enzymes and the ends of the obtained DNA fragments are re-ligated in diluted conditions that strongly favour ligation of the juxtaposed DNA fragments. The frequency of ligation between two genomic loci is then assed using polymerase chain reaction (PCR) or direct DNA sequencing. (Schmitt et al. 2016) On figure 1.4 is a schematic representation of the whole process of Hi-C method. The overhangs left after digestion will depend on the chosen restriction enzyme.
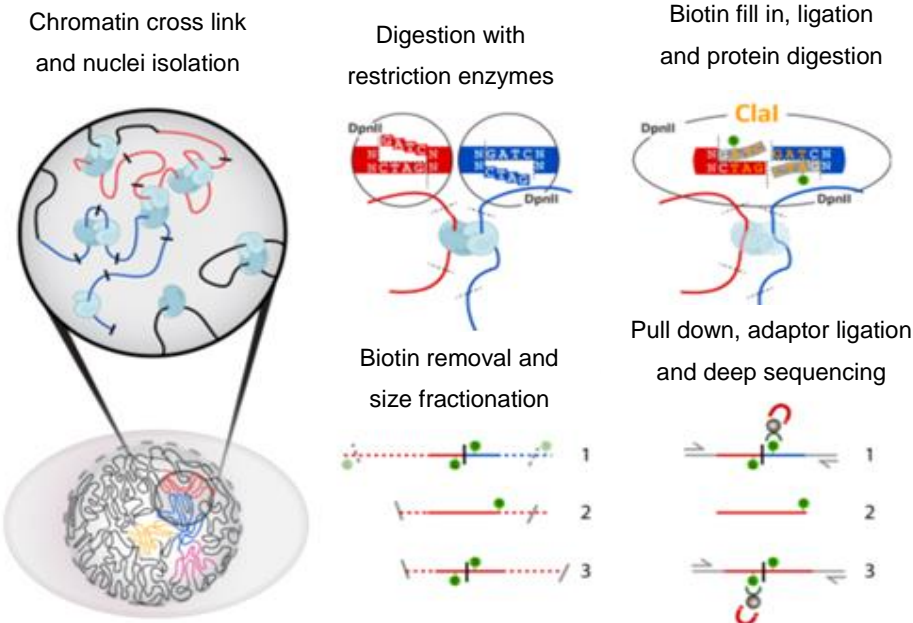
Figure 1.4 – Overview of Hi-C method. (adapted from *"Hi-C 2.0: An Optimized Hi-C procedure for High Resolution Genome Wide Mapping of chromosome conformation"* - Belaghzal et al. 2016)

2. Objectives

Having a proband with a complex phenotype, majorly characterized by cardiovascular and neurodevelopmental congenital anomalies, and with chromosomal alterations identified by classical and molecular cytogenetic analyses, the main goal of this master thesis was the establishment of the relation genotype-phenotype between proband's phenotypic characteristics and its genomic alterations, with determination of the chromosomal breakpoints with nucleotide resolution and identification of the candidate genes for the phenotype.

In order to achieve it a new approach, called Next Generation Cytogenetic, was put into practice, being the index' genome analysed through large insert whole genome sequencing, which allowed to have a much greater resolution on the genomic alterations that he is carrying. An unbalanced translocation, t(4;7)(q13.3;p21.2), with deletions on derivative chromosomes, der(4) and der(7)' breakpoints, and deletions, del, on chromosomes 2, 3, 9 and 11, were the genomic alterations detected with this methodology.

After the elucidation about all the genomic alterations presented, and to accomplish the main goal, the priority was identification of the candidate genes and clarification of molecular pathogenesis. For that, the genomic regions where the chromosomal breakpoints are localized were deeply studied and analysed, having in focus the topological association domains (TADs) where the affected genes that could contribute to the phenotype are located, and the susceptibility of those regions to copy number variations. The refinement of chromosomal breakpoints to nucleotide resolution was completed by PCR and Sanger sequencing. It was also necessary to perform familial studies with the intention to understand if the genomic alterations presented by the index were inherited or *de novo* alterations. Those were also carried out by PCR and Sanger sequencing, with verification of coincident junction fragments and chromosomal breakpoints on index' genome and on his parents' genomes. In the case of inherited alterations, with coincident junction fragments and coincident chromosomal breakpoints between proband and his parents' genomes, it was also important to understand how only the index presents phenotypic characteristics.

Finally, performing gene expression studies and chromatin conformation studies, Hi-C, were also key elements to reach the core objective of determining the genetic causes of the proband's phenotype, which might contribute to human genome annotation.

3. Analyses carried out prior to this study

3.1. Classical cytogenetic analysis

Karyotype analyses of index and his mother' samples were performed. Classical Cytogenetic analyses are usually performed on conventional giemsa-trypsin-leishman (GTL) banded metaphase chromosomes obtained from phytohemagglutinin-stimulated peripheral blood lymphocytes by standard techniques.

3.2. Molecular cytogenetic analysis

Molecular cytogenetic analyses were carried at Liverpool Women's NHS Trust, on the array CGH platform ISCA 8x60K cytochip (v2.0), from Blue Gnome. Usually the genomic DNA from the analysed individual, and reference control DNA were differentially labelled with two different fluorophores and then compared against the reference genome. Data were analysed using blue Fuse multi for microarrays (v3.2).

4. Material and Methods

The family under study was submitted to the research program "Citogenética de Próxima Geração Irrompe nos Cuidados de Saúde e Contribui para Anotação do Genoma Humano". Blood samples were collected after informed consent obtained from participants or their legal representative. The study was approved by the Ethics Committee of the National Institute of Health Doutor Ricardo Jorge and was carried out according to the Principles of the Declaration of Helsinki of the World Medical Association.

4.1. Blood Samples, RNA extraction and production of Lymphoblastoid Cell Lines (LCLs)

The blood samples, Sodium-heparin Blood Tubes, ethylenediamine tetraacetic acid (EDTA) Blood Tubes and PAXgene Blood RNA Tubes, were collected after informed consent.

Lymphoblastoid cell lines (LCLs) were established from peripheral blood mononuclear cells (PBMCs) according to the conventional protocol described by Neitzel (1986). PBMCs were isolated from whole blood samples recurring to a Ficoll gradient and centrifugation, with isolation of the buffy coat. The aspect of Ficoll gradient before and after centrifugation is represented on figure 4.1. The freshly isolated mononuclear leukocites were rinsed and incubated in medium containing Epstein Barr virus (EBV) to achieve cell immortalization, originating LCLs. The complete protocol for establishment of LCL from peripheral blood is described on appendix 1
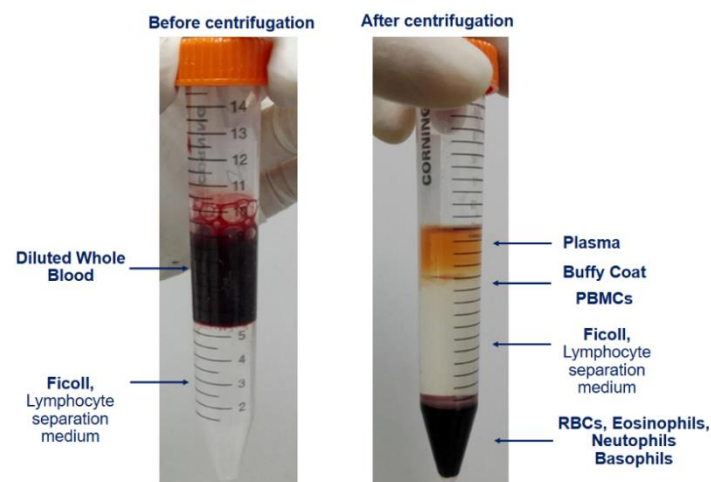
Figure 4.1 – Representation of Ficoll Gradient before and after centrifugation, for isolation of PBMCs to obtain LCLs. After spinning, is possible to isolate the Buffy Coat, using a Pasteur Pipette.

These were cultured in complete Dutch Modified Roswell Park Memorial Institute (RPMI) 1640 medium supplemented with 10% heat-inactivated foetal calf serum, 2 millimolar (mM) L-glutamine and antibiotics, at 37 Celsius degree (ºC) in a humidified atmosphere with 5% carbon dioxide (CO_2).

Prior to DNA and RNA extractions, LCLs were maintained at constant cell density between 0.7 to $0.8*10^6$/ millilitre (mL), at least 7 consecutive days. This was achieved by diluting the cells in the culture medium and dividing the culture in several culture flasks. For the extractions cell cultures were set up in duplicate at a density of $1*10^6$/mL, for DNA, and $0.7*10^6$/mL for RNA. Half of the culture medium was

replaced with fresh medium at 24 hours (h) and the cells were harvested after 48h. Cells were counted, using a Neubauer chamber and trypan blue as cellular viability marker. Cells were then pellet by centrifugation, rinsed in phosphate buffered saline (PBS), washed through centrifugation and the obtained cellular pellets $1*10^7$ and $0.8*10^7$ cells were transferred to 2mL microcentrifuge tubes. Cellular pellets for DNA extraction were stored at -20ºC, and RNA extraction was immediately carried out.

Total RNA extraction from peripheral blood samples, collected on PAXgene Blood RNA tubes, was performed with PAXgene Blood RNA extraction kit. RNA extraction was performed according to manufacturer's instructions, being the overall protocol schematized on appendix 2. In brief, the blood tubes were centrifuged, the obtained pellets were washed and dissolved in water and then subjected to cell lysis with proteinase K, for 15 minutes (min) at 55ºC. Lysates were pipette to shredder columns and centrifuged. RNA isolation from the flow throughs was made recurring to spin columns involving the steps of RNA binding to the spin column's matrix, DNA digestion with DNAse I, washes and RNA elution. The overall process is performed by centrifugation.

Extracted RNA was quantified on NanoDrop ND-1000 spectrophotometer with quality assessed by optical density (OD) 260/280 and OD260/230 ratios and through an analysis on Agilent 2100 bioanalyzer.

4.2.   DNA and RNA Extractions from LCL

Genomic DNA from peripheral blood lymphocytes and LCLs was extracted using the QIAamp DNA blood midi kit, from Qiagen. Extractions were carried out following the manufacturer's instructions. The protocol consists in cell lysis, through an incubation with proteinase K at 65ºC during 30 min on a water bath, followed by DNA binding, washing and elution, all performed in a spin column with addition different buffers in each step and centrifugations. The protocol workflow is schematized on Appendix 3.

Quality and quantity of the extracted DNA was assessed on NanoDrop ND-1000 spectrophotometer, with OD260/280 and OD260/230 rations. DNA integrity was evaluated through electrophoresis on an agarose gel 0.75% weight per volume (wt/v) Tris-Acetate-EDTA (TAE) ran during 3h at 45 volts (V).

Total RNA from LCL pellets was extracted by QIAamp RNA Blood, along with the manufacturer's instructions schematized on appendix 3. Cells were lysed by incubation with RLT buffer from the kit mixed with 2-mercaptoethanol. Lysates were transferred to Quiashredder spin columns and centrifuged. Ethanol was added to the flow-throughs and the isolation and purification of RNA was carried out on spin columns by centrifugation involving the steps of RNA binding to spin column's matrix, DNA digestion with DNAse I, Washes and finally RNA elution.

Extracted RNA was quantified on NanoDrop ND-1000 spectrophotometer with quality assessed by OD260/280 and OD260/230 ratios and through an analysis on Agilent 2100 bioanalyzer.

4.3.   Large Insert Whole Genome Sequencing (liWGS) and Bioinformatic Analysis

Large Insert Whole Genome Sequencing was carried out integrated in "Citogenética de Próxima Geração Irrompe nos Cuidados de Saúde e Contribui para Anotação do Genoma Humano" research project at Harvard Medical School.

According to the protocols described by Talkowski et al., 2011 and Hanscom and Talkowski, 2014, DNA libraries with large genomic inserts for 25 base pair (bp) paired end Illumina sequencing platform were generated.

Generally, genomic DNA is sheared to a targeted insert size, circularized, and then fragmented with the circularization junctions retained, thereby deriving short genomic fragments in which the ends are separated by the size of the circle.

The resulting data of the libraries sequencing, was submitted to a bioinformatics analysis, on INSA Genomic Disease Research Group. Usually this analysis involves several steps, including assessment of quality of the files, alignments and mapping of the proper pairs, identification of the improper pairs and filtering, in order to find the chromosomic rearrangements with nucleotide resolution.

4.4. Sequence-specific oligonucleotide design, amplification by PCR of Control and Junction fragments and Sanger Sequencing

4.4.1. Sequence-specific oligonucleotide design for the control and junction fragments of each chromosomal rearrangement

The design high quality sequence-specific oligonucleotides for the amplification of the control and junction fragments of each chromosomal rearrangement, the genomic regions of each alteration were observed and analysed on University of California, Santa Cruz (UCSC) genome browser, with a window of visualization defined by the genomic location of the breakpoints, with the tracks RepeatMasker and Interrupted Rpts activated. BLAT tool was used to identified non-repetitive but homologous elements on the region.

In figure 4.2 is a schematic representation of the analysis carried out for the deleted region identified by liWGS on index' chromosome 3. It included analysis of the repetitive elements, and homologies of the genomic region, and finally the location of the designed primers for amplification of control and junction fragments.

The primers were designed with Primer BLAST tool from National Center for Biotechnology Information (NCBI), and chosen taking in account its location – avoiding repetitive elements – sequence specificity – avoiding primers homologous with several regions – and annealing temperature – being preferred pairs of primers with close annealing temperatures. Sequence-specific oligonucleotides were designed as pairs for amplification of the control fragments, and allow combinations between each other for junction fragment amplification. All the sequence-specific primers, designed for amplification of control and junction fragments, and their specifications, of this case study are summarized on table 4.1 (Table 4.1 is integrated on INSA Genomic Diseases research group's primers tables, where all the designed primers for each studied case are summarized.)
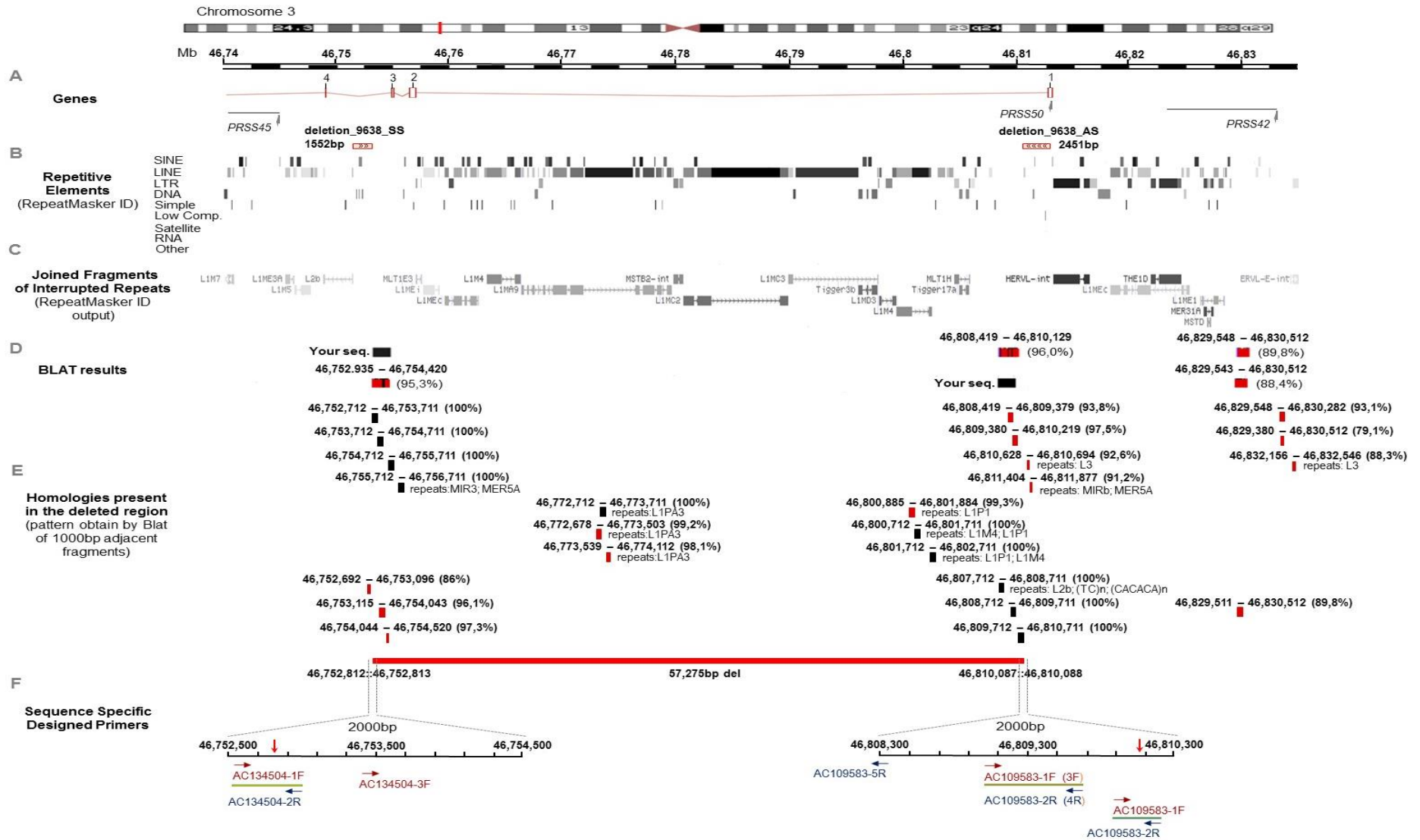
Figure 4.2 – Overview of the genomic region deleted on chromosome 3. In this scheme is possible to observe the summary of the analysis carried out in order to position the primers in locations favourable to PCR amplification, of control and junction fragments. This analysis included observation of the genomic region taking in consideration the disrupted and flanking genes (A), the repetitive elements (B), and of possible interrupted and rejoined repetitive elements (C). the interrupted and re-joined repetitive elements might cause trouble on PCR amplification and Sanger Sequencing. It also comprised a BLAT analysis with the aim of understand the homologies in the region of the sequences flanking the clusters of liWGS (D), that allowed the identification of the deleted region, and from the deleted region itself divided in 1000bp fragments (E), where the queried sequences are represented in red, and the homologies are represented in black. The lower section (F) shows the genomic location of the designed primers, being the forward primers represented in red, and the reverse primers represented in blue.

Table 4.1 – Summary of the designed primers for this case study. Specifications of each primer, like its sequence, size, strand, position and annealing temperature are enumerated.

Alteration	Chromosome	Designation	Sequence	Primer Size (bp)	Position hg38	Strand	Ann T ºC (NCBI)	ΔT ºC (Oligo)	Fragment
t(4,;7)	chr7	AC005681-1F	GTCTCCAGGAAAAGGGGGCA	20	chr7:13,184,627-13,184,646	Forward	61,79		der(7) Distal CF
		AC005681-2R	GAGATCGGATCCCTTTTGGGT	21	chr7:13,184,792-13,184,812	Reverse	59,51		
		DGKB-1F	GGGTTTCACCATGTTGGCAG	20	chr7:14,535,777-14,535,797	Forward	59,68		der(7) Proximal CF
		DGKB-2R	GCCATCTGGTCCTGAGATTTTC	22	chr7:14,536,175-14,536,196	Reverse	59,05		
	chr4	AC097709-4F	GTTTCCCTGATGGTCGGTCT	20	chr4:74,048,726-74,048,745	Forward	59,39		der(4) Distal CF
		AC097709-2R	GCACTCAGCTACACAAAGCAAA	22	chr4:74,049,591-74,049,612	Reverse	59,97		
		ACO97709-1F	GCTTGGGAATGAAGGTGAAAGC	22	chr4:73,918,724 -73,918,745	Forward	60,35		der(4) Proximal CF
		AC097709-3R	TCCCAAAATCAATCGCCTCCT	21	chr4:73,919,166 -73,919,186	Reverse	59,72		
del2	chr2	NRXN1-1F	TGCTTATTCCTATCCCCTCGGT	22	chr2:50,842,953-50,842,974	Forward	60,43		del2 Distal CF
		NRXN1-2R	GTGGGACAATTCCATAAAAGGCAAA	25	chr2:50,843,523-50,843,547	Reverse	60,51		
		NRXN1-3F	GGCTCTTCTTCCACAAGCCAT	21	chr2:50,843,207-50,843,227	Forward	60,62		
		AC007682-1F	CCCGATTTTCCTGGTGCTGT	20	chr2:51,318,599-51,318,618	Forward	60,61		del2 Proximal CF
		AC007682-2R	ACCTCACGTGCATACAGAGGA	21	chr2:51,319,244-51,319,264	Reverse	60,89		
		AC007402-1R	AAATGTAAAGACCATCAAGGCC	22	chr2:51,316,803-51,316,825	Reverse		64,80	
		AC007402-2R	CTTAAAGCCAAGCCTCGGGA	20	chr2: 51,317,538 -51,317,557	Reverse	60,04		Sequencing primer
del3	chr3	AC134504-1F	CTAGAAGGGCAAAGGGCAAGA	21	chr3:46,752,523-46,752,543	Forward	60		del3 Distal CF
		AC134504-2R	TCACTGGCGAGTTTGGTTGAA	21	chr3:46,753,002-46,753,022	Reverse	60,41		
		AC134504-3F	TCTAACGCTTGCTTCACCGT	20	chr3:46,753,391-46,753,410	Forward	59,97		
		AC109583-1F	CCTGAGTACACTGCCTGTCA	20	chr3:46,809,866-46,809,885	Forward	59,03		del3 Proximal CF
		AC109583-2R	GGTCAAGAAAGGGACCATCTGT	22	chr3:46,810,189-46,810,210	Reverse	59,96		
		AC109583-3F	CCAGAACCACGTCTAACGCT	20	chr3:46,808,976-46,808,995	Forward	60,04		
		AC109583-4R	GGAAGCCCCCTTAGTGCTTA	20	chr3:46,809,674-46,809,693	Reverse	59.08		
		AC109583-5R	GGGATGAGGAGCTTGTCTGC	20	chr3:46,808,337 -46,808,356	Reverse	60.46		

4.4.2. PCR amplification of Control and Junction Fragments

Sequence-specific PCR protocol using the designed primers for each chromosomal alteration and Amplitaq DNA polymerase was carried out for the amplification of the control and junction fragments of each chromosomal rearrangement, like described by David et al., 2013.

For each rearrangement were amplified two control fragments and the junction fragment. Control fragments are placed in the flanking regions of the breakpoints and the junction fragments are the result of the chromosomal breakage associated to the chromosomal rearrangement. Sequence-specific oligonucleotides were designed as pairs for amplification of the control fragments. The pair of primers used for junction fragment amplification is a different combination of the available primers for the control fragments of each rearrangement, usually the primer forward of distal control fragment and the primer reverse of proximal control fragment.

PCR is performed on a thermocycler and general conditions for the amplification of specific medium length fragments from genomic DNA, in this context, usually involve an initial denaturation, performed at 94ºC during 3 min, 35 cycles of denaturation, at 94ºC, annealing, at the optimal annealing temperature of the primers, and elongation, at 72ºC, during each step 45 seconds (s), and a final elongation of 3 min at 72ºC. The program hold is performed at 4ºC. Times of each step can be adjusted taking in account the length and the DNA sequence of fragment that is going to be amplified. (Life Technologies)

In order to use the same pre-mix in all polymerase chain reactions 20mL of a fresh PCR master-mix were made and aliquoted on 1mL aliquots, as summarized on table 4.2.

Table 4.2 – Composition of PCR master-mix made to this study.

Reagents	Volume
bi-distilled water (ddH$_2$O)	18mL
Amplitaq buffer I	2mL
ATP, 100mM	40µL
CTP, 100mM	40µL
GTP, 100mM	40µL
TTP, 100mM	40µL

4.4.2.1. Translocation t(4;7)(q13.3;p21.2)

Taking in account the translocation t(4;7)(q13.3;p21.2), AC097707-1F/AC005681-1F and AC097707-2R/DGKB-2R were the pairs of primers selected for the amplification of der(4) and der(7) junction fragments. On tables 4.3 and 4.4 are summarized the sets of primers used for the amplification of control and junction fragments of der(4) and der(7), respectively.

Table 4.3 – Primer combination for amplification of control and junction fragments of der(4).

der(4)		Amplified Fragment
CFrg	AC097707-4F AC097707-2R	Distal control fragment

	AC097707-1F	Proximal control fragment
	AC097707-3R	
JFrg	AC097707-1F	Junction fragment
	AC005681-1F	

Table 4.4 – Primer combination for amplification of control and junction fragments of der(7).

der(7)		Amplified Fragment
CFrg	AC005681-1F	Distal control fragment
	AC005681-2R	
	DGKB-1F	Proximal control fragment
	DGKB-2R	
JFrg	AC097707-2R	Juntion fragment
	DGKB-2R	

To establish the ideal PCR conditions, and determine the optimal primer's annealing temperature, an experimental assay called temperature gradient was performed. DNA extracted from index LCL was used on this assay. Distal and proximal control fragments of der(4) and der(7) were amplified with annealing temperatures of 58ºC, 60ºC and 62ºC, since according to primer BLAST their annealing temperature is around 60ºC. Each reaction tube contained 25 microliters (µL) of PCR master-mix, 0.25µL of each primer, at a concentration of 300 nanograms (ng)/µL, 0.25µL of Amplitaq polymerase and about 100ng of DNA. 5% dimethyl sulfoxide (DMSO) can also be added to increase specificity. PCR cycling was similar to the standard cycling described above. The same assay was carried out to determine the best amplification conditions for the junction fragments of this derivative chromosomes.

The results of these assay were evaluated by a gel electrophoresis run. Amplified fragments mixed with bromophenol blue (loading dye), at 6:1 proportion, were loaded on an agarose gel 1%(wt/v) Tris-Borate-EDTA (TBE). The run was performed during 90 min at 60V.

In this context, valid PCR amplification of control and junction fragments for Sanger sequencing, needs to be carried out with DNA extracted from peripheral blood and the amplification must be performed on a total volume of 50µL per tube.

For amplification of each fragment 50µL of PCR master-mix, 0.5µL of each primer, at a concentration of 300ng/µL, 0.5µL of Amplitaq polymerase and about 100ng of DNA were mixed on a 0,2mL tube. Amplification these fragments was done with the optimized parameters, an initial denaturation at 94ºC during 3 min, 35 cycles of denaturation, at 94ºC, annealing at 62ºC and elongation, at 72ºC, during each step 45s, and a final elongation of 3 min at 72ºC. The program hold is performed at 4ºC.

After amplification, an electrophoresis was performed, with 1% (wt/v) agarose in TBE buffer. 5µL of PCR product + 1µL of loading dye were loaded per well, and the gel was running at 60V for about 60 to 90 minutes.

In order to perform familial studies about the translocation between chromosomes 4 and 7, PCRs of junction and control fragments were also made with DNA samples from index' mother.

4.4.2.2.   Deletion 2p16.3

Taking in account the deletion on 2p16.3, NRXN1-1F/AC007682-2R was the pair of primers selected for the amplification of the junction fragment that resulted from this deletion. On table 4.5 are summarized the sets of primers used for the amplification of control and junction fragments of del2.

Table 4.5 – Primer combination for amplification of control and junction fragments of del2.

Del(2)		Amplified Fragment
CFrg	NRXN1-1F	Distal control fragment
	NRXN1-2R	
	AC007682-1F	Proximal control fragment
	AC007682-2R	
JFrg	NRXN1-1F	Juntion fragment 1
	AC007682-2R	

To establish the ideal PCR conditions, and determine the best primer's annealing temperature, temperature gradient assay was done. DNA extracted from index LCL was used on this assay. Distal and proximal control fragments were amplified with annealing temperatures of 58ºC, 60ºC and 62ºC, since according to primer BLAST their annealing temperature is around 60ºC. Each reaction tube contained 25µL of PCR master-mix, 0.25µL of each primer, at a concentration of 300ng/µL, 0.25µL of Amplitaq polymerase and about 100ng of DNA. Amplification cycling was: initial denaturation at 94ºC during 3 min, 35 cycles of denaturation, at 94ºC, annealing at 62ºC and elongation, at 72ºC, during each step 45s, and a final elongation of 3 min at 72ºC. The program hold is performed at 4ºC. The same assay was carried out to determine the best amplification conditions for the junction fragment.

The results of these assays were assessed by an electrophoresis run. Amplified fragments mixed with bromophenol blue, at 6:1 proportion, were loaded on an 1% (wt/v) agarose in TBE gel and the gel was left running for 90 minutes at 60V.

Due to the length of the obtained junction fragment was necessary to perform other PCR technique that allows amplification of longer fragments. The chosen kit was Expand Long Template PCR System from Roche, and the protocol was performed following the manufacturer's instructions. Before starting the mix of deoxynucleotides (dNTPs) (10 millimolar (mM)) was made and aliquoted and the 3 buffers from the kit were also aliquoted. Primers, that were at a concentration of 300ng/µL, were diluted to a final concentration of 60ng/µL, and an aliquot of buffer II was heated at 55ºC for 15 min to remove precipitates.

This protocol is optimized for 25µL of total reaction volume per tube and is divided into two phases, with preparation of two different premixes. The first mix contains 11.85µL of ddH$_2$0, 1.25µL of dNTP mix, 1µL of each diluted primer, 1.2µL of DMSO and 100ng of DNA extracted from index LCL per reaction tube. Since this premix contains the DNA each reaction tube should be prepared individually. The second premix contains, per tube, 6µL of ddH$_2$0, 2.5µL of buffer II, which is the selected buffer for this amplification, and 0.75µL of Taq mix. Both premixes were then mixtured on the reactions tubes where premixes 1 were prepared and quickly the PCR tubes were placed on the thermocycler. The performed

PCR cycle was: initial denaturation, performed at 94ºC during 3 min, 35 cycles of denaturation, at 94ºC for 30 seconds, annealing, with a temperature gradient from 59 to 63ºC for 45s, and elongation, at 68ºC during 1 min and 45s, and a final elongation of 7 min at 68ºC. program hold is performed at 4ºC.

The results of these assays were assessed by an electrophoresis run. Amplified fragments mixed with bromophenol blue, at 6:1 proportion, were loaded on an 1% (wt/v) agarose in TBE gel and ran during 90 min at 60V.

PCR amplification of control and junction fragments for Sanger sequencing was carried out with DNA extracted from peripheral blood. The amplification of the control fragments was carried out with the conventional PCR protocol and amplification of the junction fragment was made with Expand Long Template PCR System. PCR amplifications were done like previously referred, being the reaction volume per amplification tube of 50µL.

After amplification, an electrophoresis was performed, with 1% (wt/v) agarose in TBE buffer. 5µL of PCR product + 1µL of loading dye were loaded per well. The run was performed at 60V for about 60 to 90 minutes.

Although the del2 junction fragment was already amplified, due to limitations of Sanger sequencing technique in sequencing long fragments, an additional reverse primer was designed for the amplification of this junction fragment. NRXN1-1F/AC007402-2R was the new pair of primers used for the amplification like is summarized on table 4.6.

Table 4.6 – New Junction fragment amplification option with a new reverse primer only designed for reverse sanger sequencing of del2 junction fragment.

Del(2)		Amplified Fragment
JFrg	NRXN1-1F AC007402-2R	Junction fragment 2

To determine the optimal conditions of amplification of this fragment temperature gradient assays were performed like above described, using DNA extracted from LCL and following the standard PCR cycling.

Amplification of this fragment for Sanger sequencing, was carried out with DNA extracted from peripheral blood and performed on a total volume of 50µL per tube, with 50µL of PCR master-mix, 0.5µL of each primer, at a concentration of 300ng/µL, 0.5µL of Amplitaq polymerase and about 100ng of DNA per tube. Amplification these fragments was done with the optimized parameters, initial denaturation at 94ºC during 3 min, 35 cycles of denaturation, at 94ºC, annealing at 62ºC and elongation, at 72ºC, during each step 45s, and a final elongation of 3 min at 72ºC. The program hold is performed at 4ºC.

After amplification, an electrophoresis was performed, with 1% (wt/v) agarose in TBE buffer. 5µL of PCR product + 1µL of loading dye were loaded per well, and the gel ran at 60V for about 60 to 90 minutes.

To perform familial studies about the deletion on chromosome 2, amplification of junction and control fragments was also made with DNA samples from index' mother.

4.4.2.3. Deletion 3p21.1

For amplification junction fragment of del3, taking in account the high complexity of the genomic region, several combinations of different primers were made. The oligonucleotides used for PCRs are summarized on table 4.7.

Table 4.7 – Primer combination for amplification of control and junction fragments of del3.

Del(3)		Amplified Fragment
CFrg	AC134504-1F	Distal control fragment
	AC134504-2R	
	AC109583-1F	Proximal control fragment
	AC109583-2R	
	AC109583-2R	
JFrag	AC134504-1F	Junction fragment 1
	AC109583-2R	
	AC134504-1F	Junction fragment 2
	AC109583-4R	
	AC134504-1F	Junction fragment 3
	AC109583-5R	
	AC134504-3F	Junction fragment 4
	AC109583-2R	
	AC134504-3F	Junction fragment 5
	AC109583-4R	
	AC134504-3F	Junction fragment 6
	AC109583-5R	

To establish the ideal PCR conditions, and determine the best primer's annealing temperature, temperature gradient assay was made. DNA extracted from index LCL was used on this assay. Control and junction fragments were amplified with annealing temperatures of 58ºC, 60ºC and 62ºC, since according to primer BLAST their annealing temperature is around 60ºC. Each reaction tube contained 25µL of PCR master-mix, 0.25µL of each primer, at a concentration of 300ng/µL, 0.25µL of Amplitaq polymerase and about 100ng of DNA. Amplification cycling was: initial denaturation at 94ºC during 3 min, 35 cycles of denaturation, at 94ºC, annealing at 62ºC and elongation, at 72ºC, during each step 45s, and a final elongation of 3 min at 72ºC. The program hold is performed at 4ºC.

The results of these assays were assessed by an electrophoresis run. Amplified fragments mixed with bromophenol blue, at 6:1 proportion, were loaded on an 1% (wt/v) agarose in TBE gel and ran during 90 minutes at 60V.

Amplification of these fragments for Sanger sequencing, was carried out with DNA extracted from peripheral blood and performed on a total volume of 50µL per tube, containing 50µL of PCR master-mix, 0.5µL of each primer, at a concentration of 300ng/µL, 0.5µL of Amplitaq polymerase and about 100ng of DNA per tube. Amplification these fragments was done with the optimized parameters, initial denaturation at 94ºC during 3 min, 35 cycles of denaturation, at 94ºC, annealing at 62ºC and elongation, at 72ºC, during each step 45s, and a final elongation of 3 min at 72ºC.The program hold is performed at 4ºC.

After amplification, an electrophoresis was performed, with 1% (wt/v) agarose in TBE buffer. 5μL of PCR product + 1μL of loading dye were loaded per well. The run was carried out at 60V for about 60 to 90 minutes.

To perform familial studies about the deletion on chromosome 3, amplification of junction and control fragments was also made with DNA samples from index' mother.

4.4.3. PCR fragments purification and Automated direct dideoxy sequencing - Sanger sequencing adaptation

Control and junction fragments amplified on DNA extracted from peripheral blood were submitted to automated direct dideoxy sequencing.

Before sequencing these fragments were purified with Amicon Ultra – 0.5 Centrifugal Filter according to manufacturer's instruction. The 50μL of each PCR product were diluted in 450μL of low-Tris-EDTA (low-TE) buffer, transferred for the spin columns and centrifuged at maximum speed for 17 min. Additional 300μL of low-TE buffer were added to each spin column matrix, another centrifugation was performed for 12 min at maximum speed. For elution spin columns are flipped and placed on new sample tuples and centrifuged at 1000xg for 2 min. On Appendix 5 is a schematic representation of the protocol.

Automated direct dideoxy sequencing was carried out with BigDye terminator cycle sequencing kit and the PCR primers and reaction products were separated on the ABI Prism 3100 Genetic Analyzer, following the manufacturer's instructions. The protocol is adapted from Applied Biosystems chemistry guide – DNA sequencing by Capillary Electrophoresis.

This sequencing methodology is based on Sanger chain termination chemistry. It was performed within a cycle sequencing reaction using the previous purified PCR products as template DNA.

Each cycle sequencing reaction tube contains 2μL of BigDye cycle sequencing terminator mix, 3μL of BigDye terminator sequencing buffer, 1μL of sequencing primer dilution, 40ng of purified DNA template and ddH$_2$o up until 20μL total volume per tube.

This 25-cycle reaction, performed on a thermocycler, include DNA template denaturation, at 96ºC for 10s, primer annealing, at the specific oligonucleotide temperature for 5s and extension, at 60ºC for 4 min. On figure 4.3 is a diagram of the cycle sequencing reaction.
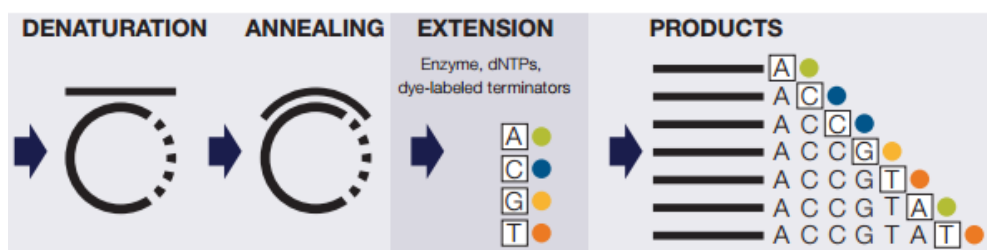
Figure 4.3 – Diagrammatic representation of cycle sequencing reaction, where the fluorescent ddNTPs were incorporated into the sequencing products. (adapted from Applied Biosystems chemistry guide – DNA sequencing by Capillary Electrophoresis)

After performing the cycle sequencing reaction, the sample tubes are delivered at INSA Unity of Technology and Innovation, where the remaining protocol of sequencing is carried out. Like published on Applied Biosystems chemistry guide – DNA sequencing by Capillary Electrophoresis, after cycle sequencing reaction, the products are purified and prepared for the separation on the sequencer. Resulting sequences were aligned and read against the human genome assembly GRCh38/hg38 for determination of the chromosomal breakpoints.

4.5. Electroelution, a method for DNA fragment purification

Electroelution is a simple technique that allows complete purification of DNA, RNA and proteins from agarose or polyacrylamide gels. It is performed on an elutor that has the capacity to purify 6 samples simultaneously. Elution technique is based on the use of a highly conductive salt barrier, so that by electrophoretic elution, molecules are concentrated in the high salt buffer, and are later precipitated. (Zarzosa-Alvarez et al. 2010)

A protocol of this technique was developed with the objective of isolate ultrapure DNA fragments for various applications, like Sanger sequencing, endonuclease digestion and electroporation of urinary cells. Alphagene Biometra HSB elutor manual was used as reference to establish the electroelution protocol.

Previous to the electroelution, is essential to electrophoresis an electrophoresis of the DNA fragments, in extraction agarose gel. After the run, the interest bands are isolated from the remaining gel on an ultraviolet (UV) light transilluminator.

The electroelution technique is based on the use of a highly conductive salt barrier, for that a solution of Ammonium Acetate, 7.5M NH_4AC with 0.2% volume per volume (v/v) bromophenol blue was made and used as high salt buffer. The Biometra HSB-elutor was filled with TAE buffer, and the v-shaped channels of the elutor carried with 200µL of high salt buffer. The interest bands excised from the agarose gel, were shopped in small cubes and placed in the sample cup of the elutor. Electroelution run was done at 125V during 1h, then salt cushion was removed using a syringe and needle, and an ethanol precipitation of the DNA was carried out.

Two volumes of ice cold absolute ethanol + 1 µL of a glycogen (5µg/ µL) solution, were added to sample tubes and left incubating at -20ºC for at least 1h, next sample tubes were centrifuged at maximum speed at 4ºC for 15 min, obtained pellets were washed twice with 300µL of ice cold 80% ethanol and centrifuged for 5 min at maximum speed. The supernatants were decanted and the pellets left air drying. After drying pellets were ressuspended in 50µL of TE buffer.

DNA pellets were quantified on NanoDrop ND-1000 and quality was assessed by OD260/280 and OD260/230 ratios. Integrity of the DNA was evaluated on agarose gel electrophoresis, with 0.8-1% (wt/v) agarose in TAE buffer, 100ng of DNA per well, and a 2h run at 45V.

4.6. Culture, Extraction and Characterization of plasmids for Establishment of Urinary Cell Lines

In an attempt to immortalize urinary primary cell lines, as a different source of RNA to perform expression studies, a pCI-neo-hEST2 stab culture, from Addgene, was acquired.

Bacterial culture to obtain isolated colonies was carried out following Addgene protocols and recommendations. Since the selection marker of competent bacteria is ampicillin, it was added to the culture medium at a final concentration of 100µg/mL.

Briefly Lysogeny broth (LB) Agar + ampicillin petri plates were made by pouring the LB agar media into the plate, around 1cm of media height, near a Bunsen burner. After dried and labeled, they were inoculated. In this process a toothpick was sterilized on the Bunsen burner, and with it the bacterial growth of the stab culture was touched and then the toothpick containing bacteria was run over a section corresponding to about one fourth of the petri plate, and then spreaded out to the rest of the plate. Inoculated plates were incubated overnight (12-18h) at 37ºC. The aspect of the plates after incubation is shown in figure 4.4, isolated colonies, whit similar diameter, should be observed.
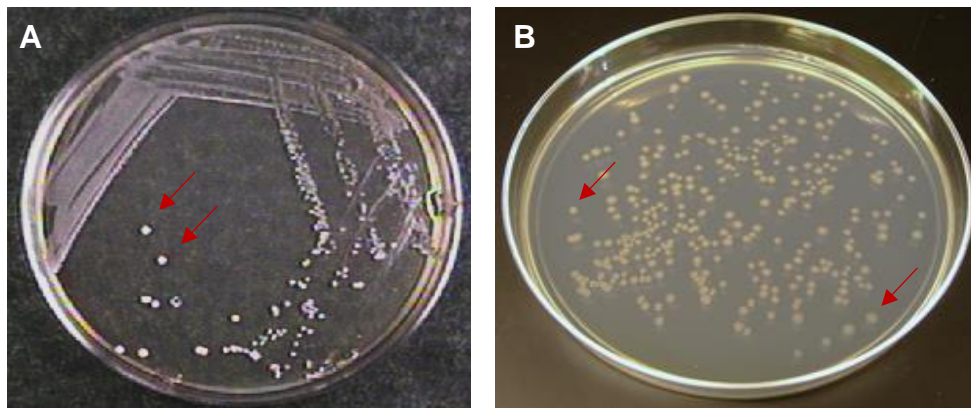
Figure 4.4 – Bacterial growth plates after overnight incubation. The red arrows indicate isolated colonies like the ones that were used to inoculate minipreps. Panel A shows the result of a direct inoculation by touching LB agar plate with the toothpick carrying bacteria, and Panel B shows the result of inoculation recurring to serial dilutions.

With the isolated colonies, miniprep cultures were established following Addgene guidelines. As previously, ampicillin was added to the culture medium at a final concentration of 100µg/mL. 3mL of liquid LB + ampicillin were added to 12mL culture tubes, then with the sterile toothpick an isolated colony from each plate was touch, then the toothpick was merged in the tube previously made. The inoculated minipreps were incubated at 37°C overnight (12-18h) in a shacking incubator at 300 rpm. The bacterial growth was observed by the hazing of the culture medium.

To know if the bacteria were competent with the interest plasmid, about 1.8mL of medium was collected for plasmidic DNA extraction, using the kit ZR Plasmid Miniprep™ – Classic, according to manufacturer's instructions. Basically, this protocol consists in cell lysis, followed by DNA binding, washing and elution, all performed in a spin column by addition of different buffers to the sample, recurring to centrifugation. The protocol workflow is schematized on Appendix 6.

In order to assess the quality and integrity of the extracted plasmids an agarose gel electrophoresis was performed, with 0.8-1% (wt/v) agarose in TBE buffer, with 100ng/µL plasmidic DNA per well, and a run at 45V for about 2h. Several plasmid conformations, linear, nicked (open-circular), supercoiled should appear separate on the gel, in the form of different bands with variable intensities.

After plasmidic DNA extraction, plasmid characterization assays were performed. Those consist in DNA digestion with restriction enzymes, in this case the digestion was carried out with *Hind III*, a two cutter in the interest DNA sequence, at 37ºC during 16h, following the manufacturer's protocol.

After plasmid characterization, the remaining volumes of miniprep cultures were used to inoculate starter cultures with a total volume of 10mL. These cultures were left incubating at 37°C overnight (12-18h) in a shacking incubator at 300 rpm. The use of starter cultures potentiates a faster bacterial growth.

After that period, the starter cultures were diluted 1/500 to 1/1000 into liquid LB + ampicillin medium, and since the pCI-neo-hEST2 DNA plasmid is a high copy number plasmid, midipreps were prepared by inoculation of 60mL of LB medium with the diluted starter culture. Midipreps were left incubating at 37°C overnight (12-18h) in a shacking incubator at 300 rpm.

Plasmidic DNA from 50mL of midipreps was extracted HiSpeed Plasmid Midi Kit (Qiagen) – Midiprep,at 4ºC, following the manufacturer's recommendations. The protocol is summarized in Appendix 7. It is a gravity methodology. Before starting, midipreps were centrifuged to obtain bacterial pellets, next these pellets are lysed by addition of several buffers, proteins and other cellular compounds were removed from the lysate by filtration, then were performed steps of binding, washing and elution of the DNA in resin matrix, followed by a DNA purification, which has started with isopropanol, with passage through a filter and washes with ethanol 70%. Finally, the ultrapure plasmidic DNA is eluted in low-TE buffer.

The remaining volume of midipreps was used to inoculate bacterial glycerol stocks of the plasmid, 500µL of the midipreps were added to 500µL of 50% glycerol in a 2mL cryovial and stored at -80ºC. Glycerol stocks are important for long-term storage of plasmids.

A part of the extracted plasmidic DNA of pCI-neo-hEST2 plasmid was subjected to digestion with the endonuclease *Hind III*, at 37ºC during 16h, following the manufacturer's protocol. It allows to separate the plasmid replication machinery in *E. coli* from the machinery of expression of human Telomerase reverse transcriptase (hTERT= in eukaryotic cells, since they have distinct molecular weights, and appear in two different bands on an agarose gel.

Assessment of quality of the extracted plasmids, was done through agarose gel electrophoresis, with an agarose gel 0.8-1% (wt/v) agarose in TAE buffer, and 100ng/µL plasmidic DNA per well. The run was performed at 45V for about 2h. Supercoiled conformation should be the dominant isoform of the plasmid, although other bands with lower intensities, which correspond to other plasmidic conformations, could appear on the gel.

Complete step by step protocol of plasmid culture and extraction is summarized on Appendix 8.

4.7.   Hi-C – Studies about Chromatin Conformation

With the aim of studying the impact of the genomic aberrations on chromatin conformation, interactions, and on topological domains, an *in situ* Hi-C protocol is being created and is on experimentation phase. This protocol was established until recircularization step, which is a break point step, and is based on Belaghzal et al. 2016; Rao et al. 2014 and Ramani et al. 2016 protocols. The main goal of this methodology is to study chromatin conformation and interactions. For that, chromatin interactions of cultured cells are fixed with formaldehyde, and then nuclei are isolated, and a Hi-C library is created. Since it is an in situ Hi-C protocol is crucial to have intact nuclear membranes with the less attached cytoplasmic fragments, since they might inhibit the forthcoming reactions.

Like described on literature, the protocol was developed and experimented on LCLs cultured in complete Dutch Modified RPMI 1640 medium supplemented with 10% heat-inactivated FBS, 2mM L-glutamine and antibiotics, at 37ºC in a humidified atmosphere with 5% CO_2, at a cell density of $5*10^6$/mL.

Cells were pelleted by centrifugation and the pellets washed with PBS. To perform chromatin crosslinking pellets were ressuspended at a density of 0.5×10^6 cells/mL in PBS, formaldehyde 37% was added to a final concentration of 1%, and the cells suspensions left incubating 10 min at room temperature. After the incubation, formaldehyde was quenched by the addiction of glycine 2.5M to a final concentration of 0.15M, and by incubating 5 min at room temperature and at least 15 min on ice.

Subsequent steps constituted the nuclei isolation methodology, and should be completed as quickly as possible.

Previous to the mechanic nuclei isolation with a Potter homogenizer, or with a glass Pasteur pipette, cellular suspensions were centrifuged and ressuspended in 1mL of ice cold Hi-C lysis buffer. Samples were left incubating on ice for a total of 15 min. In the middle of incubation time, mechanic cell homogenization was carried out with a Potter homogenizer, in two rounds with a resting pause of 1 min on ice between each round of homogenization.

To access the quality of isolated nuclei, optical microscopy fresh preparations coloured with trypan blue, giemsa-staining and light green were made, and observed at maximum ampliation, 1000x. The complete, step by step, protocol of the partial assay for assessment of the quality of the isolated nuclei is summarized on appendix 9.

To access integrity of the chromatin, additional samples of each tube, before and after homogenization were collected for posterior extraction of DNA, with classic phenol: cloroform technique, and an agarose gel electrophoresis, with 0.8-1% (wt/v) agarose in TAE buffer, containing about 100ng/µL of genomic DNA per well, and a 2 hours run at 45V, was done.

To maintain the integrity of nuclear membranes, Hi-C lysis buffer was removed by centrifugation, and nuclear pellets were washed, ressuspended in NEB buffer 2.1 and samples were placed on ice for 1 hour. The buffer chosen for this step should be the optimal buffer for the chosen restriction enzyme, which would digest the genomic DNA.

Before endonuclease digestion for genomic DNA fragmentation, sodium dodecyl sulphate (SDS) 1%, to a final concentration of 0.1% was added to sample tubes and left incubating for 10 min at 65ºC. SDS was quenched with 10% Triton X-100. Then that samples were washed and ressuspended with NEB 2.1.

Restriction reactions, with 400 enzyme units (U) of *HindIII*, were performed overnight at 37ºC, with 300rpm incubation. *HindIII* is a type II restriction endonuclease, which recognizes DNA palindromic sequence AAGCTT and cleaves between the two AAs, originating 5' overhangs, in the presence of the cofactor magnesium ion (Mg^{2+}). After 14 hours of digestion, an additional 100U of *HindIII* were added to the sample tubes. This guarantee the complete digestion of genomic crosslinked DNA.

The following steps correspond to fill-in of the DNA fragments with biotin and *in situ* ligation of the proximal ends (re-circularization).

Restriction enzyme inactivation was done by incubating the sample tubes at 80ºC during 20 min, and then place them on ice. During that time the fill-in mix was prepared. Next it was added to the sample tubes, and those ones were left incubating for 1 h at 37ºC with 300rpm agitation. In fill-in step, 5' overhangs produced by *HindIII* during the digestion, are filled in with a biotinylated deoxyribonucleotide, biotin-14-dCTP, which allows marking the digestion site. It is this specific fill-in that separates Hi-C from other chromosome conformation capture based methods.

Finally, ligation mix was prepared and added to sample tubes, which were left incubating for 4 h, at 16ºC with 300rpm agitation. Now that interacting loci are ligated into chimeric pieces of DNA, proteins that hold interacting fragments in close proximity can be removed. So that, to conclude the protocol an overnight digestion with proteinase K, called reverse crosslink was carried out.

The complete protocol of the partial *in situ* Hi-C assay performed is described on appendix 10.

5. Results and Discussion

5.1. Clinical description

The proband is the only child of a healthy non-consanguineous couple. There is no family history of any genetic disease, although his paternal grandfather died of cerebral stroke at age 40.

At 23 weeks of gestation the proband presented with pyelocalicial dilation also known as **fetal renal pelvic dilation** (HP:0010945). The rest of the pregnancy went, apparently, without any complications. Delivery was uneventful at 37 weeks of gestation.

Apgar scores were 8 and 10 at one and five minutes, respectively. At birth, his weight was 3910g, he was 51.5cm long and his cephalic perimeter was on the normal range.

At 6 months of age our proband was diagnosed with a complex cardiopathy characterized by **absent** left **pulmonary artery** (HP:0004960), stenosis of the right superior lobar branch of the pulmonary artery and a small **atrial septal defect** (HP:0001630). Additionally, our proband presented with left **(unilateral) lung agenesis** (HP:0030707). The bronchoscopy revealed a left **bronchial atresia** (HP:0030715) and a typically left bronchial tree segmentation on the right bronchial tree, maybe due to the right superior lobar branch of the pulmonary artery stenosis.

Our patient also presents with **autistic behavior** (HP:0000729), **absent speech** (HP:0001344), **moderate global developmental delay** (HP:0011343) and a **hearing impairment** (HP:0000365) that needs further clarification. Cranial Nuclear Magnetic Resonance revealed moderately **enlarged ventricles** (HP:0002119) and **sylvian cistern** (HP:0100952), **brain stem Asymmetry** due to a reduction of the left side of the pons thickness and **Dysmorphic temporal horns.**

No major dimorphism is present. His growth development was stable and his weight and height are now within the 20th percentile. He has been included in a psychomotor intervention program and speech therapy with great results.

5.2. Cytogenetic analyses

The karyotype of G-banded metaphase of the proband revealed a balanced translocation between the long arm of chromosome 4 and the short arm of chromosome 7, t(4;7)(q21.1;p21.2).

On the karyotype of G-banded metaphase of the mother no alterations were found, presenting a normal feminine sex individual karyotype, 46, XX.

Array CGH analyses of the index and both parents' genomes were also carried out for identification of genomic imbalance within the breakpoint region or elsewhere in the genome. BlueGnome ISCA 60K oligo CytoChip (v2) platform combined with Blue Fuse Multi for microarrays (v3.2) were used to carry out microarray analyses.

BlueGnome ISCA 60K oligo CytoChip (v2) platform consist of 60,000 oligonucleotide probes and allows the evaluation of the whole genome with an effective backbone resolution of 170Kb. (Liu et al. 2011)

According to the medical report whole genome microarray analysis of index DNA sample has shown a deletion of 742.0Kb on short arm of chromosome 7, encompassing 9 probes, located at 7p21.2, chr7:13,800,087-14,542,100 [GRCh38/hg38]. This deletion overlaps 2 OMIM referenced genes, E-Twenty-Six variant 1 (*ETV1*) and diacylglycerol kinase beta (*DGKB*). The significance of this deletion is uncertain at the time of the analysis. Additionally, a deletion of 490.0Kb on short arm of chromosome 2, encompassing 24 probes, located at 2p16.3 chr2:50,839,470-51,329,548 [GRCh38/hg38]. 2p16.3 deletion partially overlaps the OMIM referenced gene *NRXN1*. This deletion appears to be likely pathogenic.

In case of both mother and father, whole genome microarray analyses were not performed. The analyses only covered the genomic regions, 2p16.3 and 7p21.2, where unbalanced chromosomal alterations were find at index genome.

Mother's microarray analysis revealed a 490.1Kb deletion at 2p16.3, encompassing 24 oligo probes, like the one observed in the proband's results. So that, this deletion appears to be maternally inherited by the index subject. There was no evidence of any copy number change at 7p21.2 as seen in index genome.

On the other hand, Father's microarray analysis shown 742.0Kb deletion at 7p21.2, encompassing 9 oligo probes at 7p21.2 like the one found in proband's results. The deletion at 7p21.2 present in father's genome appears to be the same deletion present on index genome. It is therefore likely that the copy number change is a polymorphic variant and not pathogenic in causing index phenotype. There was no evidence of any copy number change at 2p16.3 as seen in index genome.

The summary of the results obtained through cytogenetic analyses are summarized on figure 5.1.
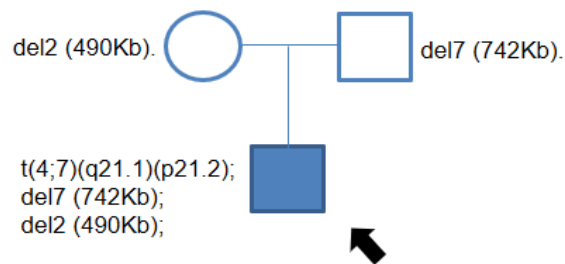
Figure 5.1 – Pedigree of the family in study with indication of the cytogenetic analyses results for each individual. The index patient with a phenotype defined by complex cardiopathy and autism is identified by a solid arrow, parents were reported as healthy individuals.

In conclusion, the identified 742Kb deletion within the cytogenetically established 7p21.2 breakpoint region and the 490Kb deletion at 2p16.3 were inherited from a phenotypically normal father and mother, respectively.

Subsequently the proband was submitted to the research project for identification of the translocation breakpoints at nucleotide resolution and determination of the candidate genes for the phenotype.

5.3. Identification of structural chromosome rearrangements by large insert whole genome sequencing (liWGS)

A sample of DNA from the proband was sent to Harvard Medical School, were the protocol of liWGS was performed. Sequencing results were sent to Genomic Diseases research group of INSA, where the bioinformatic analysis was carried out.

The sequencing achieved a raw yield of 148,555,914 paired-end reads with 25 base pairs of size. Quality assessment of data showed a high sequencing quality along the read length, although a 0.3% of contamination with illumina primer was detected. Nevertheless, the data was considered proper to prosecute the analysis.

The results of alignments and mapping of the proper pairs and identification of the improper pairs consisted on 97.51% of the initial reads successfully mapped against the reference human genome [GRCh38/hg38], where 5.46% of them were considered as improper pairs. Improper pairs are those with abnormal insert sizes or the reads of the same pair mapping into different chromosomes, or with alterations in the orientation.

Filtering of the improper pairs, allowed to determine 801, 234 and 16651 putative clusters for translocations, inversions and deletions clusters, respectively.

By Filtering of the previously identified clusters was possible to select 3 clusters that putatively identify translocations and 6 clusters that putatively identify deletions, which were submitted to further analysis. The filtered clusters passed the parameters for inversions.

After further analysis of the clusters for translocations and deletions, was possible to redefine the coordinates of the chromosomal alterations previously identified on karyotyping and array CGH, and additional deletions were discovered. The results are summarized on table 5.1.

Table 5.1 – Summary of the results from the bioinformatic analysis carried out on the data received from liWGS of proband's genome. (reference sequence: human genome assembly [GRCh38/hg38])

Cytogenetic karyotype	Type of alteration	Cytoband	Position [hg38]	Cytoband	Position [hg38]	Break point delimited region [hg38]	Alteration size (bp)
46,XY,t(4;7)(q21.1;p21.2)	Translocation	4q13.3	73,918,924	7p21.3	13,184,731	chr4:73,918,924 -74,049,529	---
	Breakpoint Deletion	4q13.3	73,918,924	4q13.3	74,049,529	chr4:73,918,924 -74,049,529	130 605
	Translocation	4q13.3	74,049,529	7p21.2	14,536,001	chr7:13,184,731 -14,536,001	---
	Breakpoint Deletion	7p21.3	13,184,731	7p21.2	14,536,001	chr7:13,184,731 -14,536,001	1 351 270
	Deletion	2p16.3	50,843,321	2p16.3	51,318,853	chr2:50,843,321 -51,318,853	475 532
	Deletion	3p21.1	46,752,812	3p21.1	46,810,087	chr3:46,752,812 -46,810,087	57 275
	Deletion	9p21.3	25,258,355	9p21.3	25,352,944	chr9:25,258,355 -25,352,944	94 589
	Deletion	11p14.3	25,680,453	11p14.3	25,699,601	chr11:25,680,453-25,699,601	19 148

The translocation between chromosomes 4 and 7, identified on index' genome by cytogenetic analysis, initially defined as t(4;7)(q21.1;p21.2), was redefined to t(4;7)(q13.3;p21.2), with estimated regions for the breakpoints of chr4:73,918,924-74,049,529, on 4q13.3, and chr7:13,184,731-14,536,001, on 7p21.2 [GRCh38/hg38].

The supposedly paternal del7 of 742Kb within the 7p21.2 breakpoint identified by aCGH, was redefined for a deletion of 1.35Mb length.

A deletion of 130.6Kb within the 4q13.3 breakpoint was identified.

The apparently maternal del2 490Kb identified by array CGH, was redefined for a deletion of 475.7Kb.

Deletions of 57.2Kb, 94.5Kb and 19.1Kb were identified on chromosomes 3, 9 and 11 respectively.

5.4.   Determination of the chromosomal breakpoints with nucleotide resolution

Refinement of the exact location of the breakpoints, within the region delimited by aCGH analysis and posteriorly redefined by liWGS, was carried out by PCR and Sanger sequencing, through automated direct dideoxy sequencing technique. For each breakpoint were amplified control and junction fragments. The control fragments, placed in the flanking regions of the breakpoints, should be amplified in carriers of the chromosomal alteration, as well as in healthy individuals. The junction fragments, which result from the chromosomal breakage associated to the chromosomal rearrangement, should only amplify in carriers of the chromosomal alteration. In healthy individuals primers are placed at very large distances, could even be in different chromosomes, which turns amplification unfeasible. Sequencing the amplified junction fragments allows to take a closer look on DNA sequences of each alteration, with determination of the chromosomal breakpoints with nucleotide resolution and identification of SNPs and other small variations linked to the chromosomal breakpoint that might be useful for the determination of the molecular pathogenesis.

5.4.1. Translocation between 4q13.3 and 7p21.2

Control and junction fragments of der(4) and der(7) were amplified by PCR from proband's DNA extracted from peripheral blood, like previous described in methods section. On figure 5.2 is possible to observe a photograph of the electrophoresis gel were the amplified fragments were ran, is possible to observe that der(4) junction fragment is about 300-400bp long, and der(7) junction fragment is about 800-900bp long.
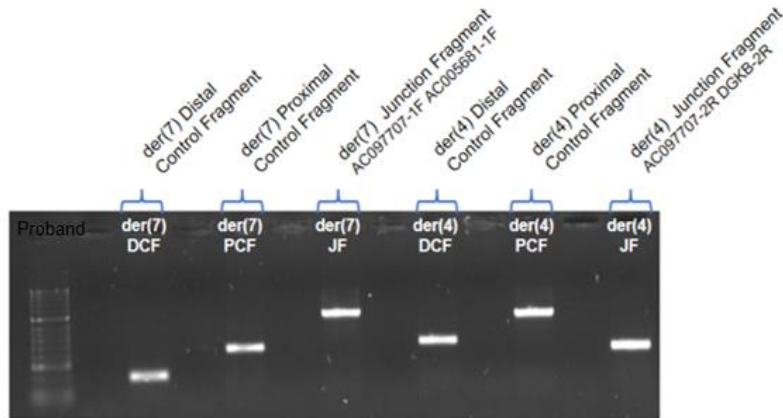
Figure 5.2 – PCR confirmation of t(4;7)(q13.3;p21,2) of the proband.

aCGH results had revealed a deletion on chromosome 7 of index' father overlapping the one found on index genome, with the major difference that his father is not carrier of the translocation t(4;7)(q13.3;p21.2). These same results do not reveal any alteration on chromosomes 4 and 7 from his mother. Since DNA samples of the father were not available, the familial study was only performed with his mother DNA sample. The pattern of bands revealed in the agarose electrophoresis gel confirmed the aCGH results, and like expect, it was not possible to amplify the junction fragments of der(4) and der(7) from mother's DNA sample. The amplified control fragments were coincident with those amplified from proband's DNA sample, like is possible to observe on figure 5.3. Amplified Control fragments of index' mother were not submitted to Sanger sequencing.
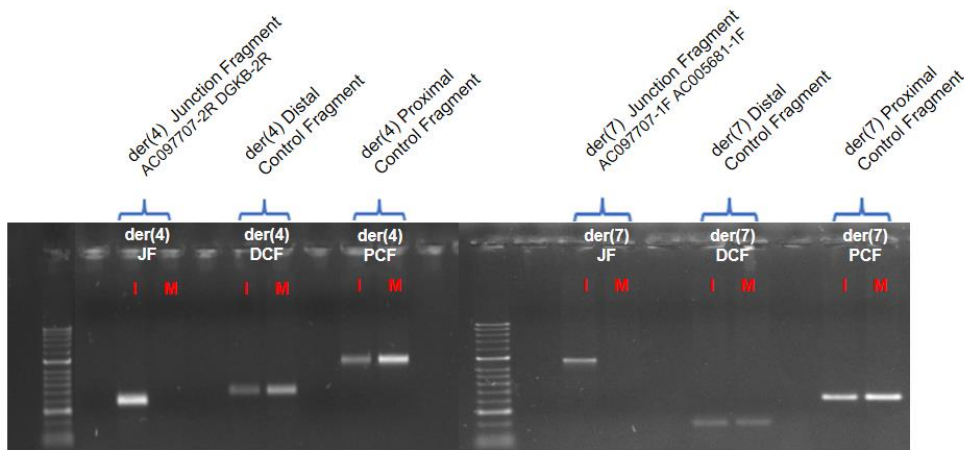
Figure 5.3 – PCR confirmation of the familial study performed about t(4;7)(q13.3;p21,2). PCR of control and junction fragments was carried out from index and his mother DNA samples. (I – fragments amplified from index' DNA sample; M – fragments amplified from mother's DNA sample)

Sanger sequencing of the fragments amplified from index' DNA sample permitted identification t(4;7)(q13.3;p21.2) breakpoints at nucleotide resolution, like is represented on figure 5.4 by the ideograms of the derivate chromosomes. t(4;7)(q13.3;p21.2) constitutes an unbalanced translocation with loss of genetic material on both translocated fragments.

DGRC150012 46,XY,t(4;7)(q21.1;p21.2).seq[GCRh38]t(4;7)(q13.3;p21.2)dn
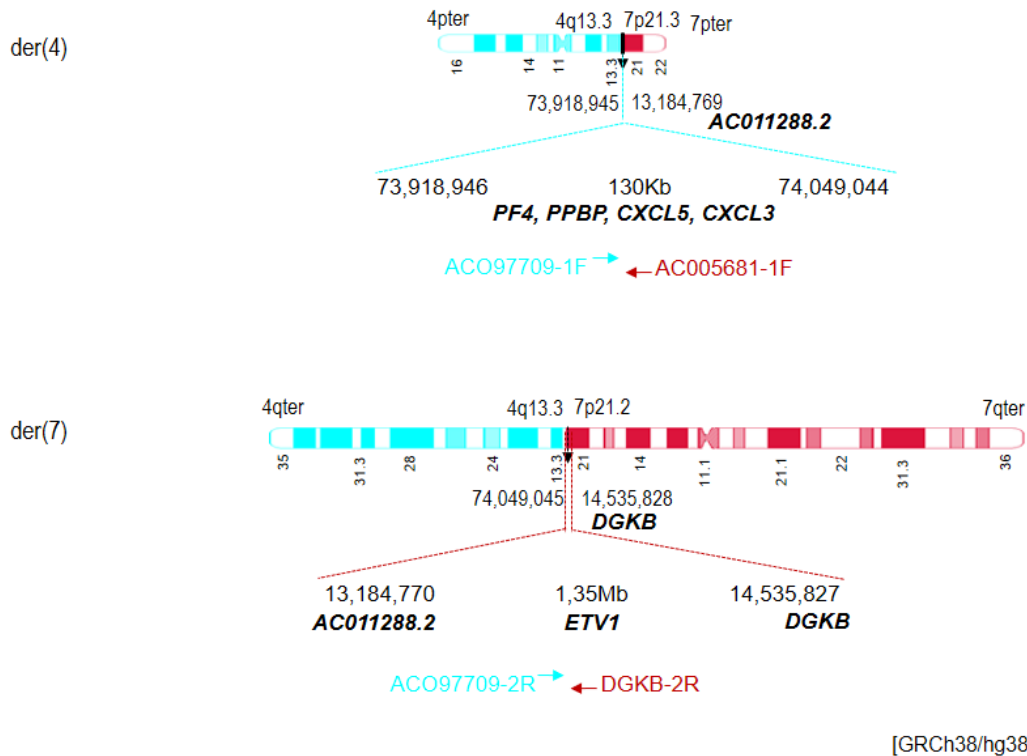
Figure 5.4 – Ideograms illustrating the derivate chromosomes resulted from t(4;7)(q13.3;p21.2). Chromosomal breakpoints are located at chr4:73,918,945, on der(4) and chr7:14,535,828, on der(7) (reference sequence: human genome assembly [GRCh38/hg38]).

On figure 5.5 are represented the Sanger sequencing results from der(4) junction fragment. On der(4) the chromosomal breakpoint is located on 4q13.3 cytoband at chr4:73,918,945, [GRCh38/hg38], having an 130Kb deletion comprised between chr4:73,918,946 and chr4:74,049,044bp, [GRCh38/hg38], within the breakpoint. The translocated fragment begins at 4q13.3 in the position chr4:74,049,045 [GRCh38/hg38]. No SNPs or additional point mutations were found in this sequence. These results were obtained by comparison between the index DNA sequence with human genome reference sequence.

DGRC150012
seq[GCRh38]t(4;7)(q13.3;p21.2)dn

der(4)

cen   4q   4q13.3 ↓ 7p21.3   tel

TTAAAAGGAACTAGAAAAGAAAAAAACTATTCACTGCCTCATTCATGTCTTCTCATGGAA

5′ TTAAAAGGAACTAGAAAAGAAAAAAACTAT CACTGCCTCATTCATGTCTTCTCATGGAA 3′
   |||||||||||||||||||||||||||||
5′ TTAAAAGGAACTAGAAAAGAAAAAAACTATTAGTTTGTGATTATTCTCCTATTAATGTAGG 3′

73,918,916   73,918,945 ▼ 74,049,046   74,049,076

130,101bp del

ATATTACTAA---------------GCTAGATTGT
g.[chr4:pter_cen_73918946_74049045del::chr7:qter_13184769inv]
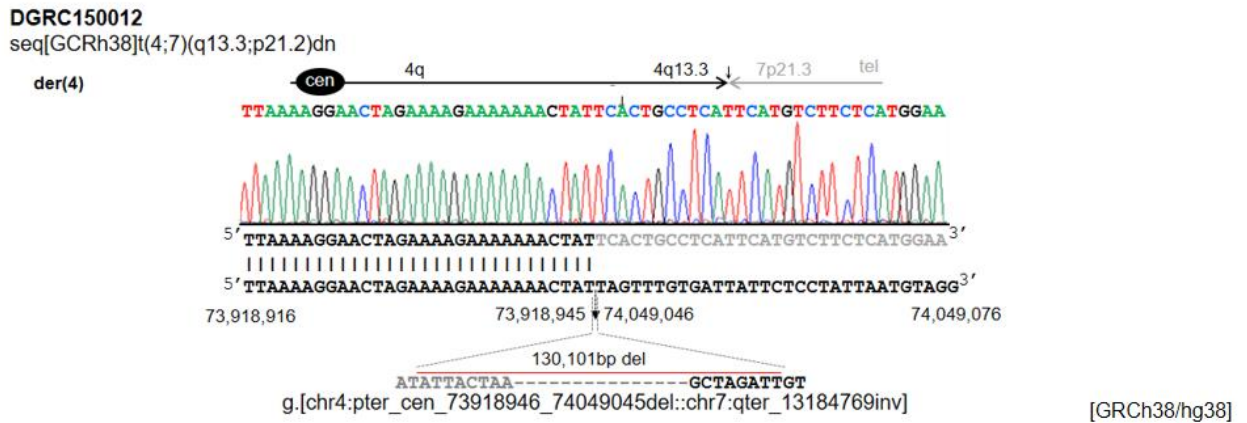
[GRCh38/hg38]

Figure 5.5 – Nucleotide sequences of der(4) breakpoint aligned against the reference sequence. Positions of the breakpoints at DNA sequence level are indicated. (reference sequence: human genome assembly [GRCh38/hg38])

On figure 5.6 are represented the Sanger sequencing results from der(7) junction fragment. On der(7) the chromosomal breakpoint is located on 7p21.2 cytoband at chr7:14,535,828, [GRCh38/hg38], having an 1.35Mb deletion comprised between chr7:13,184,770 and chr7:14,535,827, [GRCh38/hg38], within the breakpoint. The translocated fragment begins at 7p21.3 in the position chr7:13,184,769 [GRCh38/hg38]. No SNPs or additional point mutations were found in this sequence. These results were obtained by comparison between the index DNA sequence with human genome reference sequence.

DGRC150012
seq[GCRh38]t(4;7)(q13.3;p21.2)dn

der(7)

tel   4q13.3   ↓ 7p21.2   7p   cen
                    DGKB

CTACATTAATAGGAGAATAATCACAAACTAACCCCCCTCAGCCTCCCATCGTACTGGGAT

5′ CTACATTAATAGGAGAATAATCACAAACTA ACCCCCCTCAGCCTCCCATCGTACTGGGAT 3′
   |||||||||||||||||||||||||||||||
5′ TTCCATGAGAAGACATGAATGAGGCAGTGA ACCCCCCTCAGCCTCCCATCGTACTGGGAT 3′

13,184,740   13,184,769 ▼ 14,535,828   14,535,858
                                    DGKB

1,351,059bp del

ATGGTTTGGC---------------TCGTGATCCA
g.[chr4:qter_74049045inv::chr7:13184770_14535828del_cen_qter]
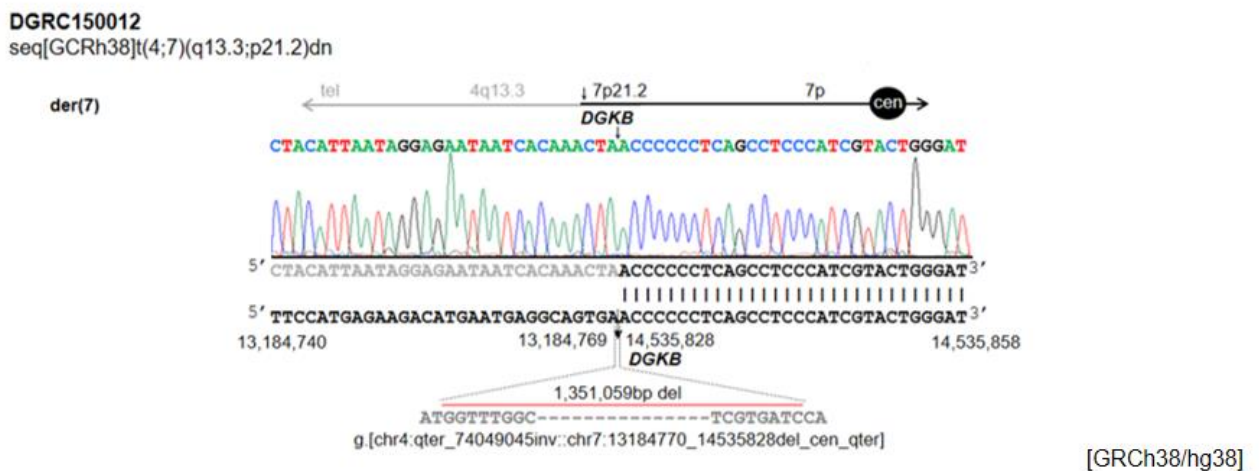
[GRCh38/hg38]

Figure 5.6 – Nucleotide sequences of der(7) breakpoint aligned against the reference sequence. Positions of the breakpoints at DNA sequence level are indicated. (reference sequence: human genome assembly [GRCh38/hg38])

5.4.2. Deletion on 2p16.3

The deletion on 2p16.3 cytoband was detected by array CGH on both index and his mother' genomes, and its breakpoints were redefined by liWGS, performed with a DNA sample from the proband.

This conducted to carry out a familial study about this deletion. So that, control and junction fragments of del2 were amplified by PCR from DNA extracted from peripheral blood of index and of his mother, like previous described in methods section. On figure 5.7 is possible to observe a photograph of the electrophoresis gel were the amplified fragments were ran. Distal control fragment is about 300bp long,

45

proximal control fragment is about 600bp long and the junction fragment is about 800bp long. Since the obtained bands have similar sizes and intensities, is possible to infer that index and his mother share the same deletion on chromosome 2.
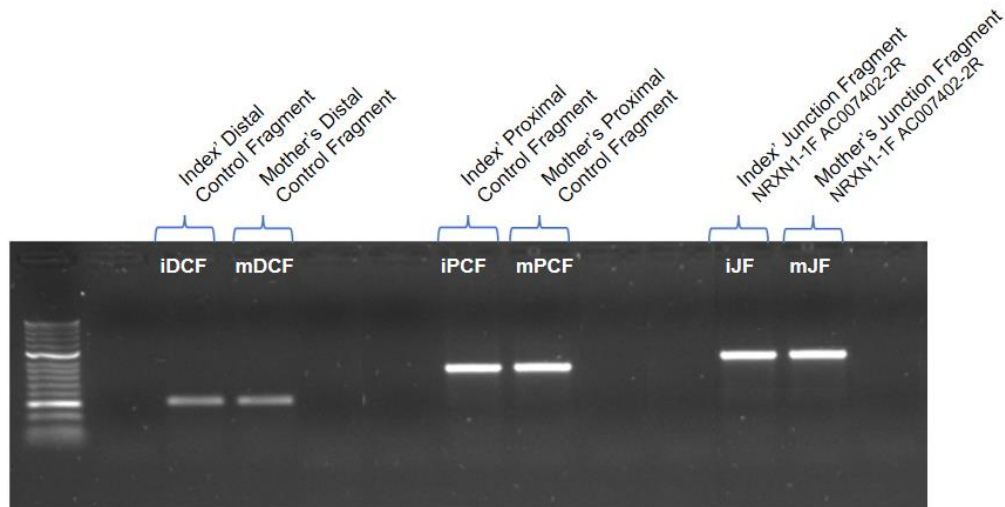
Figure 5.7 – PCR confirmation of the familial study performed about del2p16.3. PCR of control and junction fragments was carried out from index and his mother DNA samples. (I – fragments amplified from index' DNA sample; M – fragments amplified from mother's DNA sample) Comparing the bands obtained form index and his mother' samples is possible to infer that proband and his mother share the same deletion on chromosome 2.

Sanger sequencing of these fragments permitted the refinement of the 2p16.3 deletion breakpoints with nucleotide resolution, like is represented on figure 5.8 by the ideogram of chromosome 2.

Figure 5.8 – Ideogram illustrating chromosome 2 with a deletion on 2p16.3. This deletion of 473Kb, is from maternal inheritance, and is comprised between chr2:50,843,384-51,317,068, disrrupting the gene NRXN1. Primers used to amplify the junction fragment are also represented. (reference sequence: human genome assembly [GRCh38/hg38])

On figure 5.9 are represented the Sanger sequencing results from del 2 junction fragment The chromosomal breakpoint of the 473Kb deletion on chromosome 2 is located at chr2:50,843,383,

[GRCh38/hg38]. These results were obtained by comparison between the index DNA sequence with human genome reference sequence.

DGRC150012
seq[GCRh38]del(2)(p16.3)mat
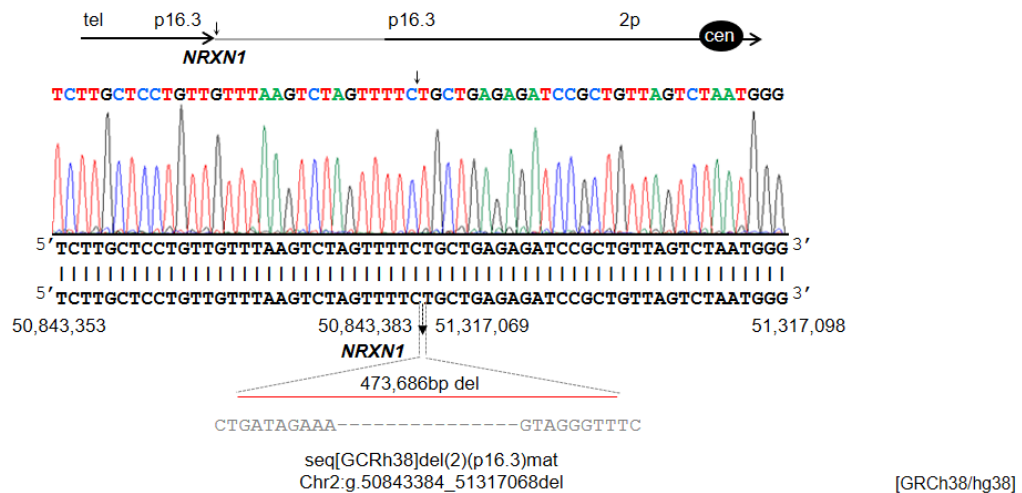
Figure 5.9 – Nucleotide sequences of del2 breakpoint aligned against the reference sequence. Positions of the breakpoints at DNA sequence level are indicated. (reference sequence: human genome assembly [GRCh38/hg38])

Sanger sequencing results were the same for index and his mother junction fragments, sharing the same SNPs and the same breakpoint location. So, this study confirms that the 473Kb deletion on 2p16.3 deletion is from maternal inheritance, having proband inherited the altered allele from his mother.

5.4.3. Deletion on 3p21.1

The approach used for studying the deletion on chromosome 3 was the same used for the previous described. Taking in account the high complexity of the genomic region, which is rich in repetitive elements and homologies between sequences, 6 different combinations of different primers were made in an attempt to amplify the junction fragment of this deletion. Control and junction fragments of del3 were amplified by PCR from DNA extracted from peripheral blood of index subject, like previous described in methods section. On figure 5.10 is possible to observe the photograph of the electrophoresis gel were the amplified fragments were ran. The distal control fragment is about 500bp long. Two proximal control fragments were amplified, the first one is about 300bp long, beginning nearer to the centromere than the second one, which is about 700bp long, and begins nearer to the liWGS cluster that allowed the identification of the deletion. From the observation of amplification pattern, is possible to conclude that the difference between junction fragment 1 and junction fragment 2 is the distance between the reverse primers used for PCR amplification, since the primer forward was common, being the same is verified between junction fragment 4 and junction fragment 5. It is also possible to observe that the reverse primer AC109583-5R don't produce capable amplification when combined with the primers forward that were available. So, only 3 of the 6 junction fragment options

reunite the conditions to be considered for Sanger sequencing, and are indicated by the red boxes in the figure.
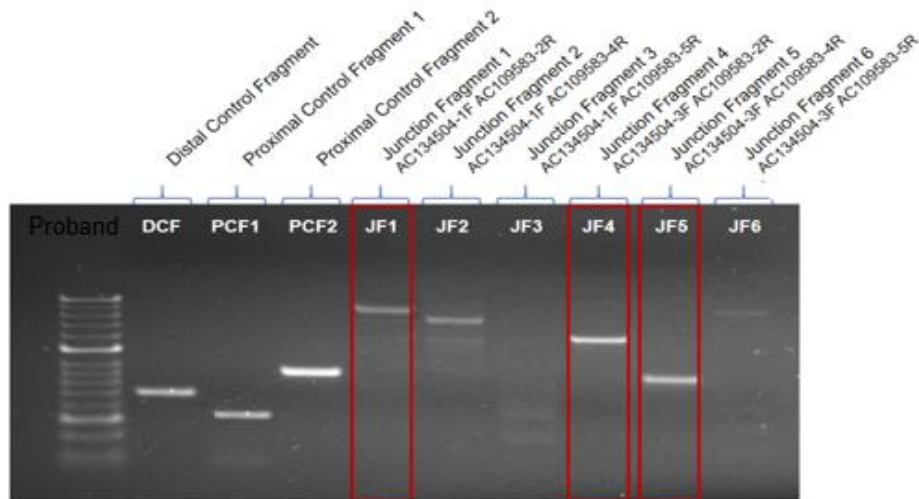
Figure 5.10 – PCR results from the amplification of control and junction fragments of del3p21.1 from a proband sample.

After sanger sequencing was not possible to determine the chromosomal breakpoint of this deletion. All the 3 options of sequenced junction fragments have homologies for 3 non-repetitive but homologous different regions found on the major genomic region where this deletion is inserted, being those chr3:46,752,700-46,756,712, chr3:46,808,419-46,811,877 and chr3:46,829,543-46,832,546 [GRCh38/hg38].

Since the results of Sanger sequencing were not conclusive, a familial study of this deletion was also performed, although there was no evidence of the presence of this deletion on index' mother genome. Surprisingly the mother presented the same pattern of amplification, which could mean that she is also a carrier of this deletion. Junction fragments amplified from mother's DNA sample, were also sequenced, having the same results, 3 homologies at chr3:46,752,700-46,756,712, chr3:46,808,419-46,811,877 and chr3:46,829,543-46,832,546 [GRCh38/hg38].

To mislead the hypothesis of this deletion could be an artefact, the same control and junction fragments were amplified from a control DNA sample which had been submitted do liWGS, and had no alteration found on chromosome 3. Unexpectedly the pattern of amplification was similar to the ones obtained for the proband and for his mother, with unpredicted amplification of junction fragments, like is possible to observe on figure 5.11. The differences observed can be due to the amount of DNA added in each PCR tube, because each PCR tube was made individually, since each one had a different set of primers for each fragment, and the DNA was pipetted from the sample tube, which could not be properly homogenized.
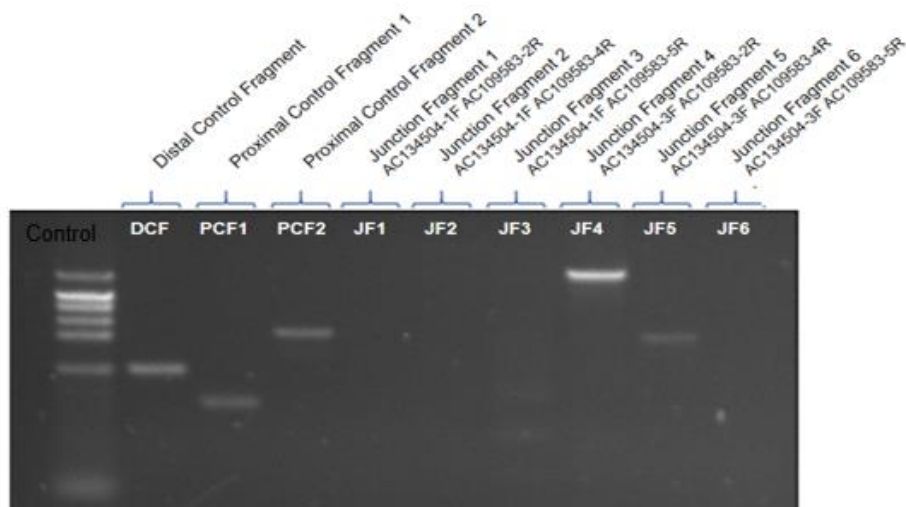
Figure 5.11 – PCR results from the amplification of control and junction fragments of del3p21.1 from a control sample. The pattern of amplification is similar to the ones obtained for amplification of the same fragments from index and his mother' DNA samples.

The results obtained with this control analysis suggests that the primers are not adequate for the amplification of this genomic region, due to its complexity. And other approaches, like for example TA cloning for Sanger sequencing, should be carried out for the determination of the chromosomal breakpoint of this deletion with bp resolution.

5.4.4. Deletions on chromosomes 9 and 11

As previous mentioned liWGS allowed the identification of two deletions one on 9p21.3, at chr9:25,258,355-25,352,944 and other on 11p14.3, at chr11:25,680,453-25,699,601 [GRCh38/hg38].

Analyses of the genomic regions where deletions on chromosomes 9 and 11, chr9:25,258,355-25,352,944 and chr11:25,680,453-25,699,601 [GRCh38/hg38], respectively, allowed to verify that those regions are exempted of genes, constituting gene deserts. Those genomic regions are both highly subjected to genomic structural variation, with non-relevant phenotypic impact in the context of this case study. On figures 5.12 and 5.13 are represented the structural variants identified on chr9:25,258,355-25,352,944 and chr11:25,680,453-25,699,601, respectively.
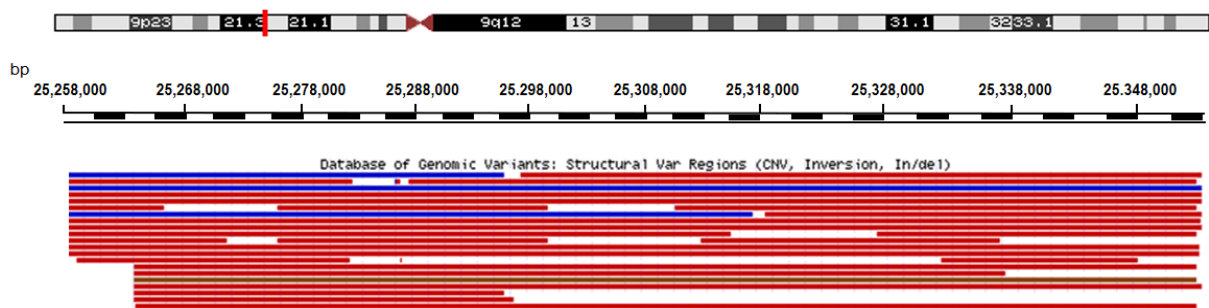
Figure 5.12 – Structural variants present on the genomic region chr9:25,258,355-25,352,944. (Red bars indicate copy number losses, blue bars represent copy number gains, and brown bars indicate Indels) (reference sequence: human genome assembly [GRCh38/hg38])

49
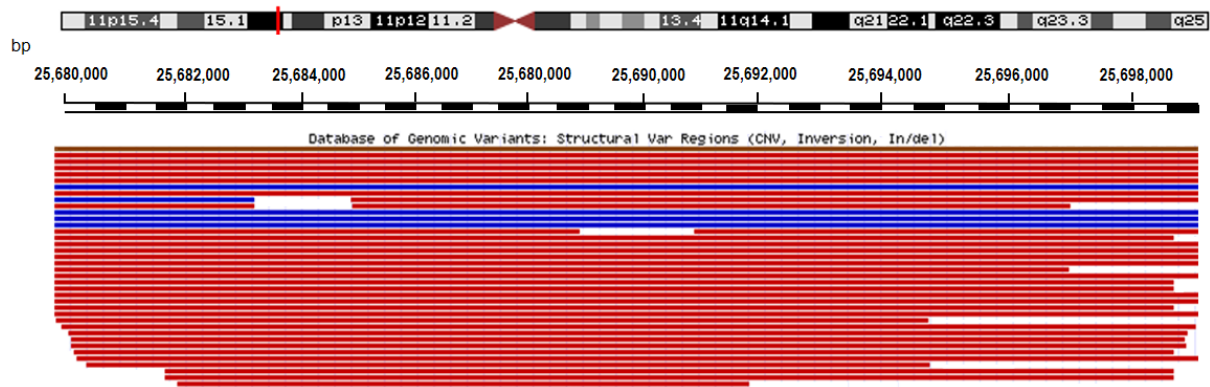
Figure 5.13 – Structural variants present on the genomic region chr11:25,680,453-25,699,601. (Red bars indicate copy number losses, blue bars represent copy number gains, and brown bars indicate Indels) (reference sequence: human genome assembly [GRCh38/hg38])

Sequence specific primers were not designed for PCR amplification of control and junction fragments of these deletions on 9p21.3 and 11p14.3, because taking in account the genomic regions where they are inserted, and the susceptibility of those regions to genomic structural variation, probably they are copy number variations and were considered like not having clinical significance for the phenotype presented by the proband.

5.5. Identification and characterization of the candidate genes

5.5.1. Identification of the candidate genes from 4q13.3 breakpoint region

Like previous referred, on **der(4)** the chromosomal breakpoint is located at chr4:73,918,945, [GRCh38/hg38], having an 130Kb deletion comprised between chr4:73,918,946 and chr4:74,049,044bp, [GRCh38/hg38], within the breakpoint. These breakpoints are in an intergenic region, not disrupting any gene. The genes Platelet Factor 4 (**PF4**) (OMIM #173460, reference gene ENSG00000163737, and transcript NM_002619) Pro-Platelet Basic Protein (**PPBP**) (OMIM #121010, reference gene ENSG00000163736 and transcript NM_002704), C-X-C Motif Chemokin Ligand 5 (**CXCL5**) (OMIM #600324, reference gene ENSG00000163735 and transcript NM_002994), and C-X-C Motif Chemokin Ligand 3 (**CXCL3**) (OMIM #139111, reference gene ENSG00000163734 and transcript NM_002090) are deleted. On figure 5.14 is a schematic representation of the genomic region of chromosome 4 where the deletion is inserted.

"*PF4* is a protein coding gene, which encodes a chemokine of CXC family, synthesized and excreted from the α-granules of activated platelets. This chemokine with homotetramere structure has high affinity to heparin and is chemotactic for other cell types. It is and inhibitor of haematopoiesis and angiogenesis, and is involved in platelet aggregation and in t-cell function, having antimicrobial activity against *Plasmodium Falciparum*. Diseases reported to be associated with *PF4* genetic alterations include Erythromelalgia, Thrombocytopenia, Ischemic stroke and Cold urticarial." (Weizmann Institute of Science 2017)

On figure 5.14 is a represented the genomic region on 4q13.3 affected by the 130Kb deletion on chr4:73,918,946-74,049,044, [GRCh38/hg38].

50
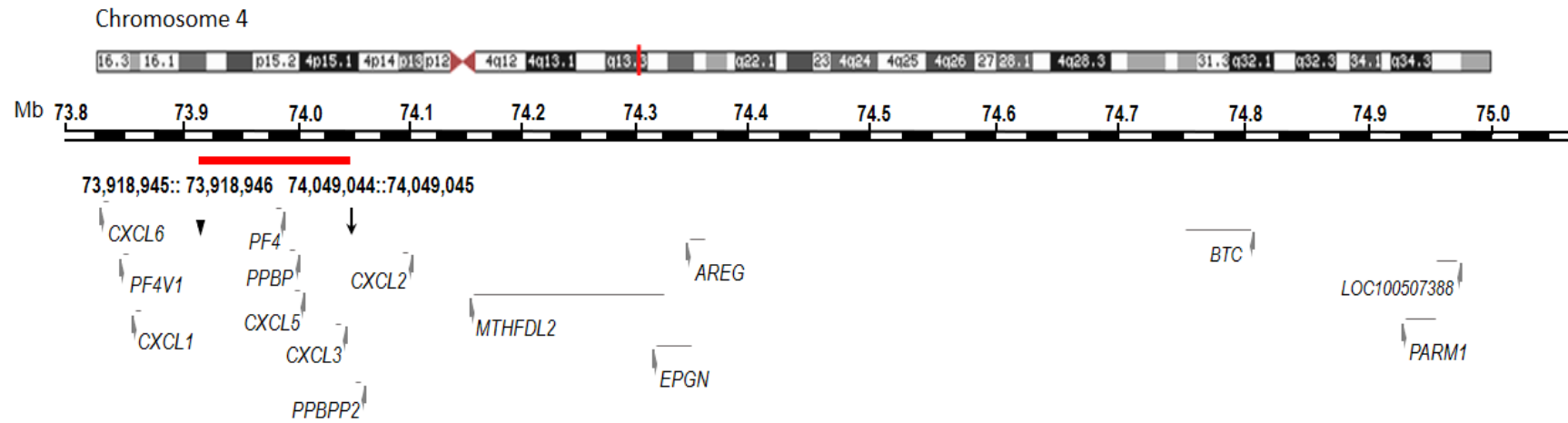
Figure 5.14 – Overview of der(4) breakpoints genomic region. The deleted region, chr4:73,918,946-74,049,044 is highlighted by the red bar, whereas arrows indicate the position of the breakpoints. All the genes present in this region are represented. *PF4, PPBP, CXCL5* and *CXCL3* are comprised within the 130Kb deletion. (reference sequence: human genome assembly [GRCh38/hg38])

"*PPBP* encodes a platelet derived growth factor of CXC chemokine family, which is a potent chemoattractant. Among its functions are neutrophil activation, stimulation of several cellular processes like DNA synthesis, glycolysis, intracellular cAMP accumulation, prostaglandin E2 secretion, synthesis and secretion of hyaluronic acid and sulphated glycosaminoglycan, stimulation of synovial cells to produce plasminogen, and also has antifungal and antimicrobial properties. Erythromelalgia and Essential Thrombocythemia, Transcobalamin II deficiency, Colloid adenoma, asthma, hypertension and vascular diseases are described, on literature, as being associated with PPBP functional mutations." (Weizmann Institute of Science 2017)

"*CXCL5* is a protein coding gene, which encodes a chemokine of CXC chemokine family with CXCR chemokine receptor. *CXCL5* is proposed to bind C-X-C motif of the G protein coupled receptor to recruit neutrophils and promotes angiogenesis and remodulation of connective tissues. It is thought to have action in cell proliferation, migration and invasion in cancer context. CXCL5 has a role on pathologies like paediatric ulcerative colitis, acute cervicitis, pulmonary sarcoidosis, cerebritis and non-specific interstitial pneumonia." (Weizmann Institute of Science 2017)

"*CXCL3* is paralog of *CXCL5*, also encoding a chemokine of CXC chemokine family with CXCR chemokine receptor. It is an antimicrobial secreted protein which participates on inflammatory processes and is a chemoattractant of neutrophils. There are no pathologic conditions described as being connected to *CXCL3* functional mutations." (Weizmann Institute of Science 2017)

According to Database of Genomic Variants, no partial deletions of the genes affected by 4q13.3 130Kb deletion were reported as structural genomic variants, so is uncertain, which the impact of this deletion on the phenotype.

Haploinsufficiency is a phenomenon characteristic of diploid organisms, characterized by the loss of function of a functional copy of a gene, resulting in insufficiency of the synthesized gene product and inability of bring up the wild type state, leading to an abnormal phenotype that is usually pathologic. Inactivation of one functional copy of the gene is frequently due to punctual mutations or chromosomal rearrangements. The haploinsufficiency index (Hi) of a gene varies between 0 to 100%, being the highest probability of a gene showing haploinsufficiency when Hi=0, and the lowest when Hi=100. Since Hi(PF4)%=89.54%, Hi(PPBP)%=90.94%, Hi(CXCL5)%=86.62%, is very unlikely that these genes exhibit haploinsufficiency when a functional copy of each one of these genes is deleted. This suggests that the reduction of functional copies of these genes may not have a major impact on the phenotype presented by proband.

Mice models constructed with several functional mutations of these genes presented complex phenotypes with alterations on haematopoietic system, on morphology and function of blood cells, on immune system and on metabolism and homeostasis. Additionally, homozygotic inactivation of these genes can lead to vascular lesions. (The Jackson Laboratory 2016) This data suggests that the 130Kb deletion on chr4:73,918,945-74,049,045, [GRCh38/hg38], may have phenotypic consequences that overlap the phenotype presented by index subject.

In summary once *PF4, PPBP, CXCL5* and *CXCL3* are deleted, will in principle be candidate genes for this phenotype. Further studies are required to conclude what is the real impact of this deletion on index phenotype.

5.5.2. Identification of the candidate genes from 7p21.2 breakpoint region

Like previous referred, on **der(7)** the chromosomal breakpoint is located at chr7:14,535,828, [GRCh38/hg38], having an 1.35Mb deletion comprised between chr7:13,184,770 and chr7:14,535,827, [GRCh38/hg38]. The breakpoint located at chr7:13,184,770 disrupts *LincRNA AC011288.2* and the breakpoint located at chr7:14,535,828 [GRCh38/hg38]. disrupts Diacylglycerol Kinase Beta (**DGKB)** (OMIM #604070; reference gene ENSG00000136267 and transcript NM_004080) on intron 19, and E Twenty-six Variant 1 (**ETV1**) (OMIM #600541, reference gene ENSG00000006468 and transcript NM_001163147) is deleted. On figure 5.15 is a schematic representation of the genomic region of chromosome 7 where the deletion is inserted.

"*DGKB* (Diacylglycerol Kinase Beta) is a protein coding gene, which encodes a protein member of DGK family. DGK proteins are cell regulators of the intracellular concentration of diacylglycerol (DAG), a second messenger, and have key roles on many cellular processes, since it is involved in Ca^{2+} binding. *DGKB* contains a conserved catalytic cysteine rich domain on C-terminal extremity, and is involved on platelet activation, signalling and activation. At the time any pathological conditions were described as being linked to *DGKB* genetic alterations." (Weizmann Institute of Science 2017)

"*ETV1* encodes a transcription factor of E twenty-six (ETS) family of transcription factors. It plays a role on regulation of many target genes, modulating several biological processes like cell growth angiogenesis, migration, proliferation and differentiation. *ETV1* and other ETS proteins have the property of sequence-specific binding to the DNA consensus sequence 5'-CGGA[AT]-3', through ETS DNA binding domain, located on C-terminal region. On N-terminal region this protein as a short acidic transactivation domain called TAD. This gene is frequently involved in numerous chromosomal translocations that result in multiple fusion proteins expressed in tumours. Among other pathologies Edwin Sarcoma, neuroepithelioma, gastrointestinal stromal tumour, sarcoma, prostate cancer and male reproductive organ cancer were described as being caused by mutations on *ETV1*." (Weizmann Institute of Science 2017)

On figure 5.15 is a represented the genomic region on 7p21.2 affected by the 1.35Mb deletion on chr7:13,184,770-14,535,827, [GRCh38/hg38].

Figure 5.15 – Overview of der(7) breakpoints genomic region. The deleted region, chr7:13,184,770-14,535,827 is represented by the red bar, whereas arrows indicate the position of the breakpoints. All the genes present in this region are represented. *DGKB* and *ETV1* are comprised within the 1.35MB deletion. (reference sequence: human genome assembly [GRCh38/hg38])

According to Database of Genomic Variants, several partial deletions *DGKB* and *ETV1* were reported as structural genomic variants, most of them with pathogenic character overlapping the phenotype presented by the proband, which suggests that this deletion might have clinical impact. And, since Hi(DGKB)%=4.15% and Hi(ETV1)%=3.00%, is highly probable that alterations at the number of functional copies of these genes result in haploinsufficiency with phenotypic consequences.

Mice models constructed with several functional mutations of these genes presented complex neurological phenotypes with alterations at Central Nervous System level, behavioural alterations and alterations on metabolism and homeostasis. (The Jackson Laboratory 2016) Additionally, homozygotic inactivation of *ETV1* may lead to premature death. This data suggests that the 1.35Mb deletion on chr7:13,184,769-14,535,828, [GRCh38/hg38], might have phenotypic consequences.

On the literature is reported, a similar translocation between chromosomes 4 and 7, t(4;7)(q13.3;p15.3), wherein the carriers of this chromosomal rearrangement presented phenotypes characterized by mental retardation, with learning difficulties and hyperactivity, and motor disabilities , which in part is overlapping with proband's phenotype. (Sismani et al. 2008)

It is also described that deletions on 7p21 are associated with Saethre-Chotzen syndrome, an autosomal dominant syndrome, characterized by premature fusion of skull bones along the coronal suture. (Reardon et al. 1993; NHI - U.S. National Library of Medicine 2008) Amongst its symptoms, are included global developmental delay, learning difficulties, hearing loss and heart defects, which overlap with phenotypic traits presented by the proband.

DGKB was reported to regulate spine formation, through lipid regulation, contributing for the maintenance of neuronal networks in synaptic transmission of cognitive processes. So, the decrease of functional copies of this gene might be related to cognitive syndromes. (Shirai et al. 2010)

In summary, *DGKB, LincRNA AC011288.2* and *ETV1* will in principle be candidate genes for this phenotype. Further studies are required to conclude what is the real impact of this deletion on index phenotype.

5.5.3. Identification of the candidate genes from 2p16.3 breakpoint region

The chromosomal breakpoint of the 473Kb deletion on chromosome 2 is located at chr2:50,843,383, [GRCh38/hg38], disrupting the fifth intron of the gene Neurexin 1 (**NRXN1**) (OMIM #600565, reference gene ENSG00000179915 and transcript NM_004801). *LincRNA AC007682.1* is also disrupted by the other chromosomal breakpoint of this deletion located at chr2:51,317,068, [GRCh38/hg38] . .

"*NRXN1* encodes a single-pass type I membrane protein that belongs to neurexin family. This type of proteins bind to neuroligins forming Ca^{2+} dependent neurexin/neuroligin complexes, which are essential for the release of neurotransmitters on the synapses of Central Nervous System. Functional mutations on this gene are allied to molecular dystrophies, Pitt-Hopkins-Like Syndrome 2 and may contribute to susceptibility to schizophrenia, autism and mental retardation." (Weizmann Institute of Science 2017)

On figure 5.16 is a represented the genomic region on 2p16.3 affected by the 473Kb deletion on chr2:50,843,384-51,317,068 [GRCh38/hg38].

According to Database of Genomic Variants, several deletions on *NRXN1* were reported as structural genomic variants, all of them associated with pathogenic phenotypes with features overlapping with the phenotype presented by the index subject. And, since Hi(NRXN1)%=0.81%, is highly probable that alterations at the number of functional copies of these gene result in haploinsufficiency having phenotypic consequences.

Mice models constructed with several functional mutations on *NRXN1* presented complex cardiovascular and neurological phenotypes with alterations at Central Nervous System level, and homozygotic inactivation of *NRXN1* cause alterations of the synapses. (The Jackson Laboratory 2016) Some of the described phenotypic characters overlap with index' phenotype, so, this data suggests that the 473Kb deletion on chr2:50,843,383-51,317,069 is pathological and highly contributes for index' phenotype.

It is described that variation on 2p16.3, with disruption of *NRXN1* confers susceptibility to autism spectrum disorder, schizophrenia, global developmental delay, intellectual disability and dysmorphic features. Additionally, is reported that variation on *NRXN1* gene has incomplete penetrance and variable expressivity, inducing highly variable phenotypes. (Kim et al. 2008; Dabell et al. 2013) So, the 473Kb deletion on *NRXN1* explains the emergence of autism amongst proband's phenotypic traits. The fact of the variation on *NRXN1* has incomplete penetrance may justify the fact of index' mother being reported as a healthy individual. This also suggests that this deletion by itself, might not sufficient to induce the same phenotype presented by the index subject.

In summary *NRXN1* will in principle be a candidate gene for this phenotype. Further studies are required to conclude what is the real impact of this deletion on index phenotype.
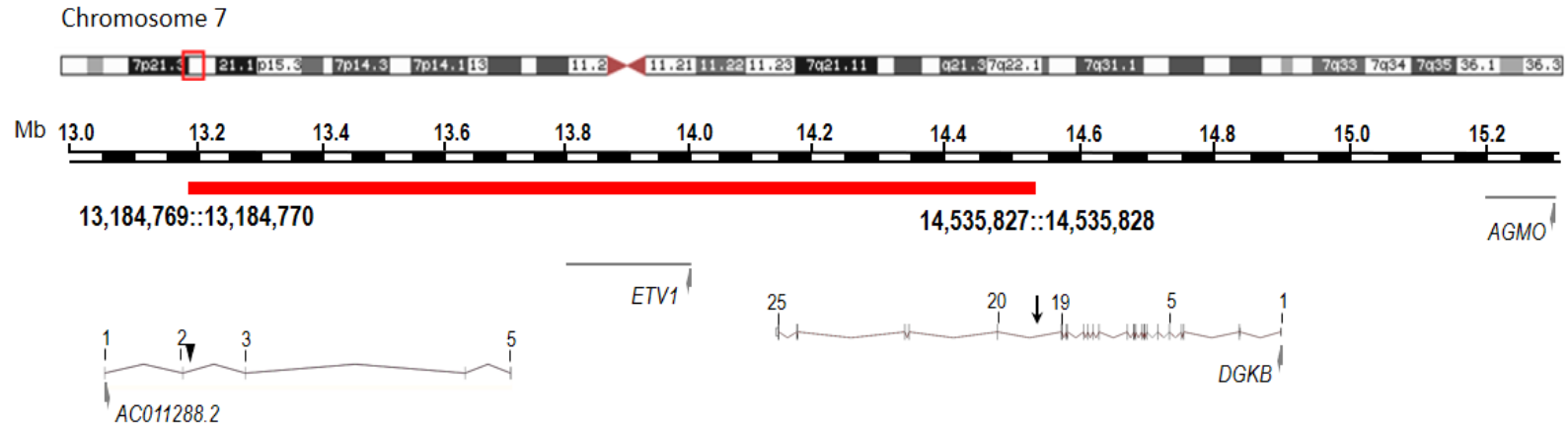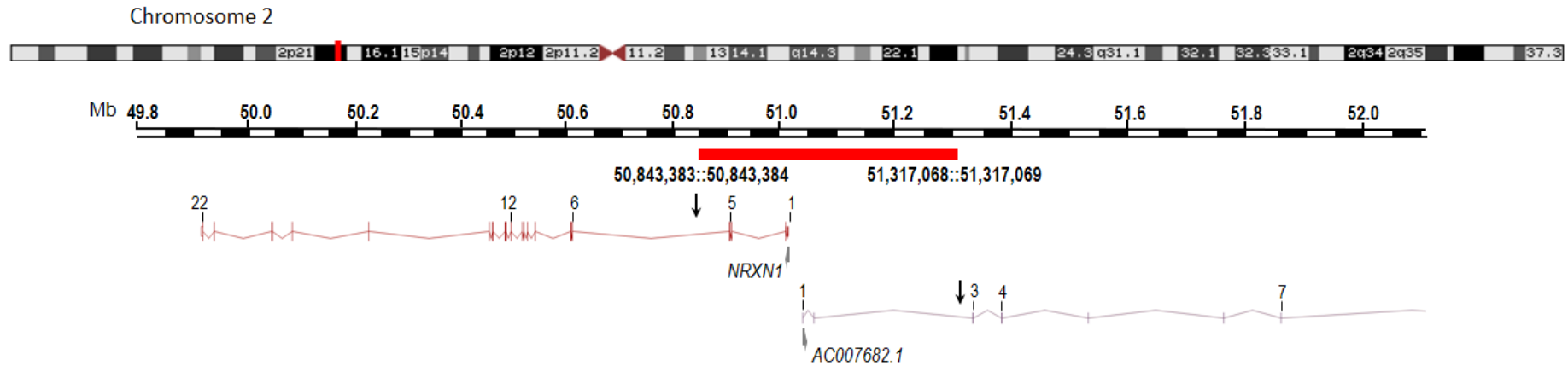
Figure 5.16 – Overview of del2 breakpoints genomic region. The deleted region, chr2:50,843,384-51,317,068, is represented by the red bar, whereas arrows indicate the position of the breakpoints. All the genes present in this region are represented. *NRXN1* is comprised within the 473Kb deletion. (reference sequence: human genome assembly [GRCh38/hg38])

5.5.4. Identification of the candidate genes from 3p21.1 breakpoint region

As previous discussed the analysis of liWGS data allowed to determine that the chromosomal breakpoint of this deletion must be located on the genomic region chr3:46,752,812-46,810,087 [GRCh38/hg38]. It was also estimated that this deletion on 3p21.1 is about 57Kb long, disrupting the gene Protease, Serine 50 (**PRSS50**) (OMIM #607950, reference gene ENSG00000283706 and transcript NM_013270) on intron 1.

"*PRSS50* encodes a protease with both. *serine-type endopeptidase activity* and *threonine-type endopeptidase activity*. Methylation of this gene is associated with its expression in a wide variety of normal tissues, like bladder and kidneys, blood, colon, breast, ovary and placenta, and demethylation is associated to its expression on testis. Hypomethylation of this gene is associated with its expression on breast and ovarian cancers." (Weizmann Institute of Science 2017)

According to Database of Genomic Variants, a deletion of the region chr3:46,743,108-46,984,809 [GRCh38/hg38] is reported as a pathologic structural genomic variant, where the phenotype is characterized by an obsolete malformation of the heart and great blood vessels (Miller et al. 2010), suggesting that this 57Kb deletion may have a role on proband's phenotype. But since Hi(PRSS50)%=91.95%, is very unlikely that this genes exhibit haploinsufficiency when a functional copy of each one of these genes is deleted. So, the probability of this deletion, by itself, has an important role in the phenotype decreases.

Mice models constructed with several functional mutations of these genes presented phenotypes characterized by alterations on haematopoietic and immune systems, with decreased number of T-cells and on skeleton, with bones presenting decreased mineral density and abnormal mineralization. Since those phenotypic characteristics do not overlap with index' phenotype, this data suggests that the 57Kb deletion on 3p21.1, by itself, might not have impact on the phenotype presented by the proband.

In summary once *PRSS50* is disrupted, it will in principle be a candidate gene for this phenotype. Further studies are required to conclude what is the real impact of this deletion on index phenotype.

5.6. Molecular pathogenesis – Protein coding Genes located on TADs interrupted by the Chromosomal Rearrangements and their relation with the clinical phenotype

Is described on the literature that fetal renal pelvic dilatation affects about 1% of the foetuses, and is caused by the underdevelopment of the urinary tract. If it is <15mm usually is temporary with spontaneous resolution, but if it is >15mm is permanent and strongly associated with urinary tract pathologies. (Hothi et al. 2009; Drnasin et al. 2013).

Interatrial communications are reported as the most common of the congenital heart defects. Most cases are the result spontaneous genetic mutations, although, 1/3 of the cases are associated with a hereditary syndrome. Such genes as *ACTC1*, *CITED 2, GATA4, GATA6, MIH6, NX2-5, TBX20* and *TLL1* have been identified as susceptibility genes for this defect. (Orphanet 2017)

Agenesis of the left pulmonary artery is caused by a malformation of the sixth aortic arch during embryonary development, and is frequently associated with other CHD like atrial septal defects,

tetralogy of Fallot and pulmonary atresia. (Reading and Oza 2012; Emren et al. 2015) Usually a consequence of this defect is left pulmonary hypoplasia, which is characterized by incomplete development of the lung tissue, and could be the base of the agenesis of the left lung, since it is caused by the failure of the development of the lung bud during embryogenesis. (Singh et al. 2015; Emren et al. 2015) Bronchial atresia is a developmental abnormality characterized by focal obliteration of the proximal segment of the bronchus associated with hyperinflation of the distal lung. (Gipson et al. 2009)

Is important to note that incidence of autism spectrum disorders is higher in people presenting hearing impairment. Most autists have an rough relation with sound, having serious difficulties in communication. (Network Autism) Language disorders may be associated with hearing impairment or can be associated to neurological disorders. (American Psychiatric Association 2013) The brainstem asymmetry might be related to hearing impairment, non-verbal autism and psychomotor developmental delay, since it plays important roles in communication, because is the processing centre of visual and auditory information, and is one post of control of the motor system. (Inner Body; Luxon 1980; Abrams et al. 2006)

To better understand molecular pathogenesis of the genomic alterations presented by the index subject, and with the purpose of identifying any additional candidate genes for the phenotype, determination of the genes present on the TADs where the chromosomal breakpoints are inserted was made. For t(4;7)(q13.3;p21.2) upstream and downstream TADs were taken in account, since they might have their normal function altered, because of the size of the deletions linked to this translocation, particularly the 1.35Mb deletion on der(7) breakpoint, which erases almost an entire TAD. On table 5.2 are summarized the protein coding genes contained on the TADs considered for t(4;7)(q13.3;p21.2). The deletion on 2p16.3, chr2:50,843,384-51,317,068 [GRCh38/hg38], covers two different TADs, being the disrupted gene *NRXN1* present in both of them, the two disrupted TADs were taken in account for the analysis. On table 5.3 are summarized the protein coding genes contained on the TADs considered for del2(p16.3;p16.3).

Table 5.2 – Protein coding genes located on the two TADs interrupted by t(4;7)(q13.3;p21.2) breakpoints. Genome version [hg38]. TADs are according to Dixon et al. 2012. (**AR** and **AD** means that the phenotype is autosomal recessive or autosomal dominant, respectively. **NA** means that there is not available information on that topic)

GeneCard Genes	Gene name	OMIM Gene	OMIM Phenotype	Inheritance
TAD-1 at position chr4:72275419-73035419 in IMR90 fibroblasts				
ADAMTS3	A Disintegrin-like and metalloproteinase with thrombospondin type 1 motif, 3	605011	NA	NA
Breakpoint spanning TAD at Position chr4:73075419-74155419 in IMR90 fibroblasts				
ANKRD17	Ankyrin repeat domain-containing protein 17	615929	NA	NA
ALB	Albumin	103600	616000	NA
			615999	NA
AFP	Alpha-fetoprotein	104150	615969	AR
			615970	AD
AFM	Afamin	104145	NA	NA
RASSF6	Ras association domain family, member 6	612620	NA	NA
CXCL8	Interleukin-8	146930	NA	NA
CXCL6	Chemokine, C-X-C motif, ligand 6 (granulocyte chemotactic protein-2)	138965	NA	NA
PF4V1	Platelet factor 4, variant 1 (PF4-like)	173461	NA	NA
CXCL1	Chemokine, C-X-C, ligand 1 (GRO1 oncogene; melanoma growth stimulating activity, alpha)	155730	NA	NA
Breakpoint (Chr4:73,918,945)				
PF4	Platelet factor 4	173460	NA	NA
PPBP	Pro-platelet basic protein	121010	NA	NA
CXCL5	Chemokine, C-X-C motif, ligand 5 (epithelial-derived neutrophil-activating peptide 78)	600324	NA	NA
CXCL3	Chemokine, C-X-C motif, ligand 3 (GRO3 oncogene)	139111	NA	NA
Breakpoint (Chr4: 74,049,046)				
PPBPP2	Pro-platelet basic protein-like 2	611591	NA	NA
CXCL2	Chemokine, C-X-C motif, ligand 2 (GRO2 oncogene)	139110	NA	NA
MTHFD2L	Methylenetetrahydrofolate dehydrogenase 2-like, NADP(+)-dependent	614047	NA	NA
TAD+1 at position chr4:74815766-75535766 in IMR90 fibroblasts				
PARM1	prostate androgen-regulated mucin-like protein 1	NA	NA	NA
RCHY1	Ring finger and CHY zinc finger domain containing 1, E3 ubiquitin protein ligase	607680	NA	NA
THAP6	THAP domain-containing protein 6	612535	NA	NA
TAD-1 at position chr7:12513849-13273850 in IMR90 fibroblasts				
SCIN	Scinderin	613416	NA	NA
ARL4A	ADP-ribosylation factor-like 4	604786	NA	NA
Breakpoint (Chr7:13,184,769)				
Breakpoint spanning TAD at Position chr7:13273850-14713850 in IMR90 fibroblasts				
ETV1	ETS variant gene-1	600541	NA	NA
DGKB (disrupted)	Diacylglycerol kinase, beta, 90kD	604070	NA	NA
Breakpoint (Chr7:14,535,828)				
TAD+1 at position chr7:15193850-16073850 in IMR90 fibroblasts				
AGMO	Alkylglycerol monooxygenase	613738	NA	NA
MEOX2	Mesenchyme homeo box 2	600535	NA	NA

Table 5.3 – Protein coding genes located on the two TADs interrupted by del2(p16.3;p16.3) breakpoints. Genome version [hg38]. TADs are according to Dixon et al. 2012. (**AR** and **AD** means that the phenotype is autosomal recessive or autosomal dominant, respectively. **NA** means that there is not available information on that topic)

GeneCard Genes	Gene name	OMIM Gene	OMIM Phenotype	Inheritance
TAD-1 at position chr2:48599357-50479358 in IMR90 fibroblasts				
STON1-GTF2A1L	STON1-GTF2A1L readthrough	NA	NA	NA
GTF2A1L	General transcription factor IIA, 1-like	605358	NA	NA
RP11-310N16		NA	NA	NA
LHCGR	Luteinizing hormone/choriogonadotropin receptor	152790	176410	NA
			238320	AR
			238320	AR
			238320	AR
			176410	AD
FSHR	Follicle stimulating hormone receptor	136435	233300	AR
			608115	AD
			276400	AR
NRXN1 (disrupted)	Neurexin 1	600565	614325	AR
			614332	NA
Breakpoint spanning TAD at Position chr2:50479358-53799359 in IMR90 fibroblasts				
Breakpoint (Chr2:50,843,383)				
NRXN1 (disrupted)	Neurexin 1	600565	614325	AR
			614332	NA
Breakpoint (Chr2:51,317,069)				
ASB3	Ankyrin repeat-containing SOCS box protein 3	605760	NA	NA
GPR75-ASB3	GPR75-ASB3 readthrough	NA	NA	NA
CHAC2	ChaC, cation transport regulator homolog 2 (E. coli)	NA	NA	NA
ERLEC1	Erlectin (endoplasmic reticulum lectin 1)	611229	NA	NA
TAD+1 at position chr2:53799359-55639361 in IMR90 fibroblasts				
GPR75-ASB3	GPR75-ASB3 readthrough	NA	NA	NA
ERLEC1	Erlectin (endoplasmic reticulum lectin 1)	611229	NA	NA
GPR75	G protein-coupled receptor 75	606704	NA	NA
PSME4	Proteasome activator subunit 4	607705	NA	NA
ACYP2	Acylphosphatase 2, muscle type	102595	NA	NA
TSPYL6	TSPY-like 6	NA	NA	NA
C2orf73	chromosome 2 open reading frame 73	NA	NA	NA
SPTBN1	Spectrin, beta, nonerythrocytic-1 (beta-fodrin)	182790	NA	NA
EML6	Echinoderm microtubule associated protein like 6	NA	NA	NA
RTN4	Neurite outgrowth inhibitor (reticulon 4)	604475	NA	NA
CLHC1	clathrin heavy chain linker domain containing 1	NA	NA	NA
RPS27A	Ribosomal protein S27a	191343	NA	NA
MTIF2	Mitochondrial translational initiation factor 2	603766	NA	NA
CCDC88A	Coiled-coil domain-containing protein 88A	609736	617507	NA
CFAP36	cilia and flagella associated protein 36	NA	NA	NA
PPP4R3B	Protein phosphatase 4, regulatory subunit 3, beta	610352	NA	NA
PNPT1	Polyribonucleotide nucleotidyltransferase 1	610316	614932	AR
			614934	AR

After the determination of the protein coding genes present on the considered TADs, establishment of relationships genotype-phenotype was carried out, for that VarElect NGS phenotyper was used.

VarElect NGS phenotyper from Weizman Institute of Science is a platform of extensive phenotype-gene relationships annotation. It identifies and prioritizes genes and variants according to their relevance to diseases and phenotypes of interest.

To carry out this analysis, first all of the deleted and disrupted genes on index' genome were inserted on the platform, and then the phenotypic characteristics were inputted one by one, being performed an analysis of establishment of gene-phenotype relation for each phenotypic feature. The analysis was then extended to all the 50 genes present in the considered TADs in order to identify additional candidate genes for the phenotype. The obtained results are summarized next.

None of the 8 initial candidate genes or either of the 50 integrating the considered TADs appears to be related to the development of **fetal renal pelvic dilatation**.

NRXN1, *PF4*, *PPBP*, *CXCL3*, *CXCL5* and *DGKB* were identified like being indirectly related to the **absent pulmonary artery**, being *NRXN1* the gene with highest relevance score for the development of the feature. This indirect relation is due to interactions, via several signalling pathways, with genes like cartilage associated protein (*CRTAP*), collagen type I alpha 2 chain (*COL1A2*), collagen type I alpha 1 chain (*COL1A1*), CD36 molecule (CD36) and parathyroid hormone 1 receptor (*PTH1R*). These genes are reported to be linked to the development of Osteogenesis Imperfecta type VII, which include the development of this defect among the traits of the disease. So that, this reveals that the combination of the unbalanced t(4;7)(q13.3;p21.2) and the deletion on 2p16.3, contributes additively for the development of this character. Performing the same analysis including the 50 protein coding genes of the considered TADs, did not permit to identify any gene directly associated to this feature, being identified 29, mainly from the TADs of chromosomes 2 and 4, as being indirectly related due to interactions with the same genes linked to Osteogenesis Imperfecta type VII.

All the 8 candidate genes, *NRXN1*, *PF4*, *PPBP*, *CXCL5*, *CXCL3*, *DGKB*, *ETV1* and *PRSS50* were reported to be indirectly related with the development of **atrial septal defects**, being *PF4* and *ETV1* the ones with higher scores of relevance. This association is due to interaction with genes like *MYH6*, *NOTCH2*, *GATA4* that are described on the literature to be linked with the development of congenital heart defects. The combination of all the genomic alterations contribute additively for the development of this feature, being the higher contribution of the deletions on 4q and 7p. Performing the same analysis including the 50 protein coding genes of the considered TADs allowed to identify alpha-fetoprotein (*AFP*), albumin (*ALB*) and neurite outgrowth inhibitor (reticulon 4) clathrin heavy chain linker domain containing I (*RTN4*) like being directly associated with this feature, additionally 42 of the remaining protein coding genes were identified to be indirectly related to the development of this defect due to interactions with genes reported to be linked to development of congenital heart defects. Variation on *AFP* may originate atrial septal defects associated to Patau syndrome and Orofaciodigital Syndrome type VIII, variation on *ALB* may trigger the development of atrial septal defects associated to Hennekam syndrome type 1, was described that *RTN4* might be involved processes relevant to heart development and variation on this gene may be implicated in the development of this defect.

None of the 8 candidate genes were reported as being directly or indirectly related to the development of **unilateral lung agenesis**. Performing the same analysis including the 50 protein coding genes of the considered TADs only allowed the identification of 4 indirectly related genes, *AFP*, *ALB*, ribosomal protein S27A (*RPS27A*) and proteasome activator subunit 4 (*PSME4*), from the TADs of chromosomes 4 and 2 respectively. This association was established due to interaction of this genes with T-box 5 (*TBX5*), a gene reported as being in the basis of development of unilateral lung agenesis. So, the alterations on 4q and on 2p might contribute for the development of this congenital defect.

ETV1 was identified to be indirectly related **with bronchial atresia**, due to an interaction with Wilms tumor 1 (*WT1*), a gene identified to be causative of bronchial atresia, pulmonary malformation and renal agenesis. Performing the same analysis including the 50 protein coding genes of the considered TADs only allowed the identification of 7 genes, from the TADs of chromosomes 4 and 7, indirectly related to the development of the feature, being their contribution due to interactions with *WT1*. This suggests that the unbalanced translocation t(4;7)(q13.3;p21.2) is in the basis of the development of this feature.

All the 8 candidate genes were found to be indirectly related to the development of **pulmonary artery stenosis**, being *ETV1* and *PF4* the genes with highest relevance ratios. This defect is associated to a variety of syndromes, that include vascular malformations among the symptoms, being the relevance of these genes due to interactions with genes like KRAS proto-oncogene, GTPase (*KRAS*), *NOTCH2*, elastin (*ELN*), SOS Ras/Rac Guanine nucleotide exchange factor 1 (*SOS1*), NRAS proto-oncogene, GTPase (*NRAS*), TNF superfamily member 11 (*TNFSF11*). Performing the same analysis including the 50 protein coding genes of the considered TADs allowed the identification of 43 indirectly related genes due to interactions with the previous mentioned genes. Among the indirectly related genes *AFP* is the one with the highest score of relevance, immediately followed by *ETV1*. This suggests once more that the unbalanced rearrangement t(4;7)(q13.3;p21.2) has an important role in the development of the cardiopathy, in this case with higher contribute of the deletion on chromosome 7.

Amongst the 8 candidate genes, like described on the literature, *NRXN1* was identified to be directly related with development of **autistic behavior**. The remaining candidate genes were found to be indirectly related with this feature, due to interactions with genes like *PTEN*, myocyte enhancer factor 2C (*MEF2C*), gamma-aminobutyric acid type A receptor beta 3 subunit (*GABRB3*), described to be causative of a wide variety of syndromes were autistic behavior is developed among the symptoms. *ETV1* has the highest score of relevance among the indirectly related candidate genes. Performing the same analysis including the 50 protein coding genes of the considered TADs allowed the identification of 43 indirectly related genes due to interactions with the previous mentioned genes. These results suggest that the deletion of *NRXN1*, has the main role in the development of this trait, although it has contribution of the remaining genomic alterations.

PPBP, *CXCL3*, *PF4*, *CXCL5* and *DGKB* were identified to be indirectly linked with the development of **moderate global developmental delay,** due to interactions with several genes, like filamin A (*FLNA*), SRY-box 3 (*SOX3*), TEA domain transcription factor 1 (*TEAD1*), apoptosis inducing factor mitochondria associated 1 (*AIFM1*), tripeptidyl peptidase 2 (*TPP2*) described to be causative of a wide variety of

syndromes were this trait is developed among the symptoms. The genes deleted on 4q have the highest relevance scores, which indicates that this deletion might have a major role in the development of moderate global developmental delay. Performing the same analysis including the 50 protein coding genes of the considered TADs allowed the identification of 32 indirectly related genes due to interactions with several genes somehow related to the development of this feature.

All the 8 candidate genes were found to be indirectly related to **hearing impairment**, being *PPBP, ETV1* and *NRXN1* the ones with higher relevance scores. These genes interact with a wide variety of genes, through different signalling pathways, that are linked to development of several syndromes that include hearing impairment among symptoms. Performing the same analysis including the 50 protein coding genes of the considered TADs allowed the identification of 7 genes directly related to the development of this feature. Polyribonucleotide Nucleotidyltransferase 1 (*PNPT1*), is described as being linked to the development of non-syndromic autosomal recessive deafness, variation in Follicle Stimulating Hormone Receptor (*FSHR*) causes 46,xx Testicular Disorder of Sex Development, which includes hearing impairment among the symptoms, *ALB* is associated Hennekam syndrome and Focal Dermal Hypoplasia, both including this trait among the symptoms, mutations on *RPS27A* are causative of a Mitochondrial Myopathy that induce hearing impairment, variation on *AFP* associated to Down syndrome and Patau syndrome may cause this defect, Wegener Granulomatosis caused by mutations on C-X-C Motif Chemokine Ligand 8 (*CXCL8*) originates hearing impairment and variation on Luteinizing Hormone/Choriogonadotropin Receptor (*LHCGR*) is linked to Mccune-Albright Syndrome, being hearing impairment a characteristic symptom of this syndrome.

The 8 candidate genes were found to be indirectly related to the development of **enlarged ventricles** due to interactions with genes like citron rho interacting serine/threonine kinase (*CIT*) and coiled-coil domain containing 88C (*CCDC88C*), genes implicated on neurodevelopmental pathologies. The genes deleted on 4q have the highest relevance scores, which suggest that they might have higher contribution on the formation of this defect. Performing the same analysis including the 50 protein coding genes of the considered TADs allowed the identification of 44 indirectly related genes due to interactions with genes linked to neurodevelopmental malformations.

None of the 8 initial candidate genes or either of the 50 integrating the considered TADs appears to be related to the development of the **sylvian cistern**.

NRXN1, PPBP, CXCL5, PF4, CXCL3 were pointed as being indirectly related with **brain stem asymmetry** due to interactions with tubulin alpha 1 A (*TUBA1A*) and tubulin beta 3 class III (*TUBB3*) genes associated with the development of Tubulinopathy-associated Dysgynia, that include this defect among the characteristic features. Performing the same analysis including the 50 protein coding genes of the considered TADs allowed the identification of 21 indirectly related genes, mostly located on the TADs of chromosomes 2 and 4. The indirect relation of these genes with the defect is due to interactions with the previous mentioned genes.

None of the 8 initial candidate genes or either of the 50 integrating the considered TADs appears to be related to the development of the **dysmorphic temporal horns**.

Through this analysis was possible to identify 8 additional presumably candidate genes for the phenotype, since they are directly related with the development of some phenotypic characters presented by the proband. **ALB** (OMIM #103600, reference gene ENSG00000163631), **AFP** (OMIM #104150, reference gene ENSG00000081051), **PNPT1** (OMIM # 610316, reference gene ENSG00000138035), **FSHR** (OMIM #136435, reference gene ENSG00000170820), **RPS27A** (Ribosomal Protein S27a, OMIM #191343, reference gene ENSG00000143947), **RTN4** (OMIM #604475, reference gene ENSG00000115310), **CXCL8** (OMIM #146930, reference gene ENSG00000169429) and **LHCGR** (OMIM #152790, reference gene ENSG00000138039).

Additively all the chromosomal abnormalities identified on proband's genome contribute for the development of the phenotype. The unbalanced chromosomal translocation is indirectly associated with the cardiopathy and neurodevelopmental alterations, being the 1.35Mb deletion on der(7) breakpoint on the basis of the development of neurodevelopmental alterations, contributing mainly for moderate global developmental delay, hearing impairment and the speech impairment. The deletion of 473Kb on 2p16.3 is in the basis of the development of autism and might also be implicated in the development of the cardiopathy.

Although most of the disrupted and deleted genes only contribute indirectly to the development of the phenotypic traits, the reduction of functional copies of these genes might induce dramatic alterations on a variety of signalling pathways, altering the expression of genes reported to be directly related to the development of the defects.

The phenotypic consequences linked to the 130Kb deletion on der(4), that resulted from t(4;7)(q13.3;p21.2), and linked to the deletion on chromosome 3 are not yet completely clarified. Most of the human genes are poorly annotated, which difficult the task of establishing the molecular pathogenesis. When possible to obtain relevant human biologic samples, in which the candidate genes have expression, determination of gene expression profiles for each candidate gene is required in order to better understand the molecular pathogenesis.

Based on what is known at the time, *NRXN1, DGKB*, *ETV1* and *LincRNA AC011288.2* were identified as being the most likely genes to be responsible for the proband's phenotype mainly characterised by cardiopathy and autism

5.7.   Mechanisms underlying the formation of t(4;7)(q13.3;p21.2)

Through the analysis carried out to determine the chromosomal breakpoints underlying the chromosomal rearrangement t(4;7)(q13.3;p21.2) was identified a homology on the DNA sequence from der(7) between chr7:14,535,894-14,535,911 and chr4:74,048,894-74,048,878, [GRCh38/hg38], being this region of 4q deleted on this rearrangement. The presence of this homology was the trigger to perform an analysis of identification of homologies between the regions of 4q and 7p were the chromosomal breakpoints underlying this chromosomal rearrangement are located. Regions of 20Kb were defined around each breakpoint, and then each region was subdivided in fragments of 1Kb. Homologies were assessed using UCSC tool BLAT. The results of the analysis are summarized on

figure 5.17. Four homologies were found: a homology between chr7:14,532,828-14,533,828 and chr4:74,559,990-74,560,068 with 90.5% of identity; a homology between chr7:14,535,828-14,536,328 and chr4:74,679,473-74,679,547 with 86.7% of identity; a homology between chr7:13,184,769-13,185,769 and chr4:73,781,467-73,781,797 with 89% of identity; and a homology between chr4:74,048,045-74,049,045 and chr7:13,343,378-13,343,629 with 90.4% of identity, [GRCh38/hg38].
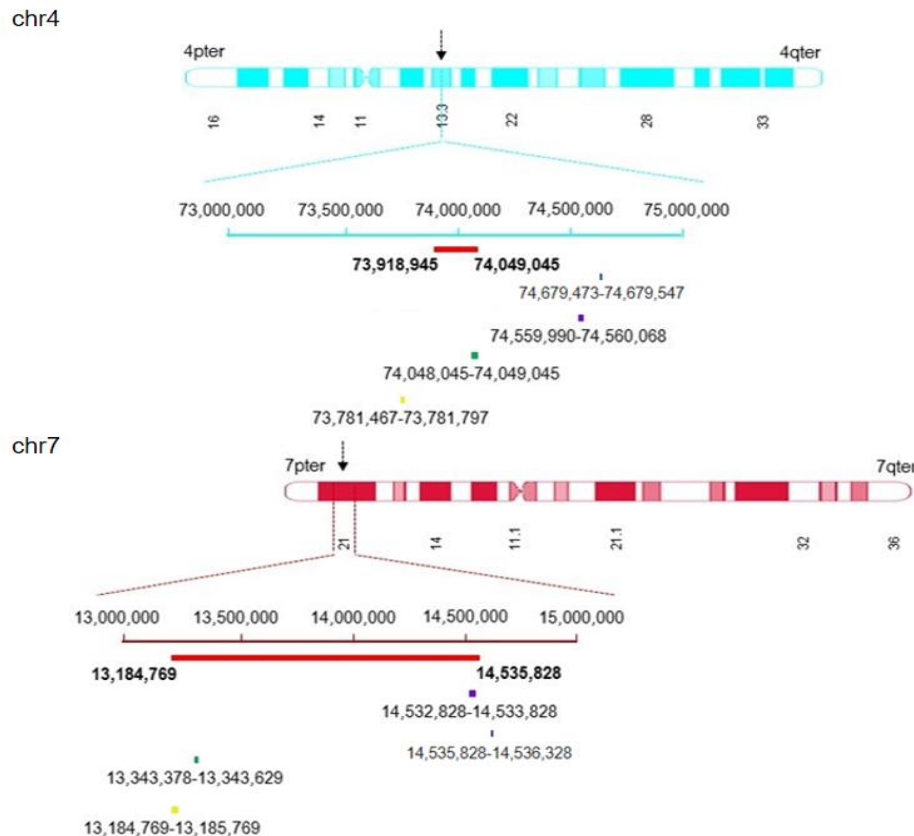
Figure 5.17 – Results of analysis carried out to establish homologies between the genomic regions around the chromosomal breakpoints of the unbalanced t(4;7)(q13.3;p21.2). Is possible to observe chromosomes 4 and 7, the deletions in 4q and 7p, marked in red and 4 homologies found represented with different colours. Homology between chr7:14,532,828-14,533,828 and chr4:74,559,990-74,560,068 (90.5%) is represented on purple; homology between chr7:14,535,828-14,536,328 and chr4:74,679,473-74,679,547 (86.7%)is represented on blue; homology between chr7:13,184,769-13,185,769 and chr4:73,781,467-73,781,797 (89%) is represented on yellow; and homology between chr4:74,048,045-74,049,045 and chr7:13,343,378-13,343,629 (90.4%) is represented on green. (reference sequence: human genome assembly [GRCh38/hg38])

Due to the possibility of the del(7), identified on index genome, be from paternal inheritance, these results suggests that the occurrence of a chromosomal breakage on 4q13.3 near 73,918,945, , [GRCh38/hg38], in the presence of the deletion on chr7:13,184,769-14,535,828, [GRCh38/hg38], may triggered the formation of t(4;7)(q13.3;p21.2). The homologies and the proximity of the chromosomes on chromatin condensed states may lead to incorrect reparation of DNA, promoting the re-joining of fragments from non-homologous chromosomes.

Further studies are necessary to clarify the molecular mechanisms underlying the formation of this chromosomal rearrangement. A Hi-C analysis would help to explain the 3D conformation and organization of the chromatin on index' genome.

5.8. Electroelution assays for isolation of DNA fragments

Literature refers that electroelution is the best methodology for complete purification of DNA, RNA and proteins having a recovery rate up to 80%.

Based on that the electroelution protocol was developed and experimented, following the protocol described in methods section. Isolated high molecular weight bands of *λ/HindIII* ladder were electroeluted using Biometra HSB-elutor. The aim of this assay was to determine the recovery rate for electroelution with the Biometra HSB-elutor, starting with known concentrations of DNA from *λ/HindIII ladder*. On figure 5.18 are represented the photographs of the electrophoresis performed for the isolation of the bands for electroelution, A, and the results obtained after the assay.

6μg of *λ/HindIII* were loaded on the gel and the 23Kb, the 9Kb and the 6Kb bands were isolated. The 23Kb band corresponds to 48% of the total quantity of DNA, having about, 2.9μg of DNA, the 9Kb band corresponds to 19% of the total amount of DNA, having 1.14μg of DNA and the 6Kb band corresponds to 14% of the total quantity of DNA, having 840ng of DNA.
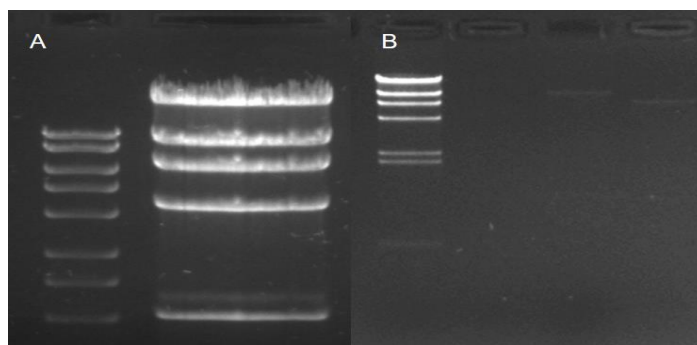
Figure 5.18 – Photographs of the electrophoresis gel performed before the electroelution, for isolation of the interest bands, A, and after electroelution of the isolated fragments, B. The 23Kb, 9Kb and 6Kb were isolated from a total of 6μg of *λ/*HindIII.

After electroelution the fragments were precipitated with ethanol and quantified on NanoDrop ND-1000 and quality was assessed by OD260/280 and OD260/230. The results are summarized on table 5.4.

Table 5.4 – Results of the quantification, on NanoDrop ND-1000, of the electroeluted *λ*/HindIII fragments.

λ/HindIII	Concentration(ng/μL)	OD260/280	OD260/230
23Kb	3.1	1.564	0.391
9Kb	77.5	1.566	0.574
6Kb	104.5	1.560	0.592

After quantification, the isolated fragments were loaded on an 0.8 % (wt/v) agarose gel and the electrophoresis was left running for 2h at 45V. On panel B of figure 5.18 is the photograph of the obtained results. The electrophoresis confirmed the quantification results obtained on NanoDrop ND-1000, absence of the 23Kb band, and both 9 and 6Kb bands present, being the 6Kb band slightly more intense.

Recovery rates were then calculated for 9 and 6Kb bands. A quantification done by comparison of the isolated bands with the bands of the ladder, that was at a total concentration of 50ng/μL, allow to conclude that after electroelution the DNA of the 9Kb band was at a concentration of 1ng/μL and DNA

of the 6Kb band was at a concentration of 2ng/µL. since the DNA pellets were eluted on 50µL of TE-buffer, were recovered 50ng DNA from the 9Kb band and 100ng of DNA from the 6Kb band, being the recovery rates of 4.4% and 11% respectively.

Improvements are required to obtain the recovery ratios mentioned on the literature. The low recovery might be due to defects on the HSB-elutor, since the channels of the equipment are tight and rough, which might be interfering in the whole recovery process. It is also important to ameliorate the ethanol precipitation process, used for DNA purification, in order to achieve cleaner eluates, since OD260/280 and OD260/230 reveal contamination of the samples with salts and glycogen used for DNA precipitation.

6. Conclusion

Having a proband presenting a complex phenotype characterized by a cardiopathy, with atrial septal defect, unilateral agenesis of a lung, unilateral absence of the pulmonary artery, bronchial atresia, and neurodevelopment alterations including autism, moderate global developmental delay, hearing impairment and morphologic alterations of some brain structures, apparently related with a chromosomal rearrangement between 4q and 7p the main goal of this master's thesis was the establishment of the relation phenotype-genotype. DNA samples from the index were submitted to liWGS and the bioinformatic analysis of the data, allowed to determine all the chromosomal alterations and the genomic regions where the chromosomal breakpoints should be located. The translocation between 4q21.1 and 7p21.2 identified by karyotyping was redefined for t(4;7)(q13.3;p21.2), having deletions within the chromosomal breakpoints of the derivate chromosomes, the breakpoints and length of the supposedly maternally inherited deletion on 2p16.3 were also redefined, and additional deletions on chromosomes 3, 9 and 11 were found. Determination of the chromosomal breakpoints at nucleotide resolution was achieved by PCR and Sanger sequencing of the junction fragments of each rearrangement.

Amplification and Sanger sequencing of the junction fragments of derivate chromosome, der(4) and der(7), resulting from the chromosomal translocation t(4;7)(q13.3;p21.2), allowed to determine that der(4) chromosomal breakpoint is located at chr4:73,918,945, [GRCh38/hg38], having an associated deletion of 130Kb on chr4:73,918,046-74,049,044, [GRCh38/hg38], and that der(7) chromosomal breakpoint is located at chr7:14,535,828, [GRCh38/hg38], having an associated deletion of 1.35Mb on chr7:13,184,770-14,535,827, [GRCh38/hg38]. The chromosomal breakpoints identified on der(4), chr4:73,918,945 and chr4:74,049,045, [GRCh38/hg38], respectively, are located in intergenic regions not disrupting any gene, being *PF4*, *PPBP*, *CXCL5* and *CXL3* genes deleted within the 130Kb deletion. The chromosomal breakpoint chr7:14,535,827, [GRCh38/hg38], identified on der(7) disrupts *DGKB* gene, *ETV1* gene is deleted within the 1.35Mb deletion and *LincRNA AC011288.2* is disrupted by the chromosomal breakpoint located at chr7:13,184,770, [GRCh38/hg38].

Amplification and Sanger sequencing of the junction fragment from the deletion on 2p16.3, from DNA samples extracted from index and his mother peripheral blood samples, permitted to determine the existence of a 473Kb maternally inherited deletion with chromosomal breakpoint located at chr2:50,843,383, [GRCh38/hg38], disrupting *NRXN1*, a gene associated with Pitt-Hopkins like syndrome (OMIM #614325), susceptibility to schizophrenia and chromosome 2p16.3 deletion (OMIM #614332) at fifth intron.

It was not possible to determine the location, with nucleotide resolution, of the chromosomal breakpoint of the 57Kb deletion on 3p21.1. estimated to be comprised at chr3:46,752,812-46,810,087, [GRCh38/hg38]. It is estimated that the chromosomal breakpoint of this deletion disrupts the gene *PRSS50*.

All the disrupted and deleted genes were presumably considered candidate genes for the phenotype. The genes involved on chromosomal translocation t(4;7)(q13.3;p21.2) were found to be indirectly

implicated in the development of cardiopathy, and also indirectly implicated in the development of some neurodevelopmental features, like global developmental delay, hearing impairment and morphologic defects on the brain.

PF4, *PPBP*, *CXCL5* and *CXCL3* were mostly found to be indirectly associated with vascular malformations, with the development of atrial septal defect, and with the morphologic alterations in the brain, namely on brain stem asymmetry, and global developmental delay. Although *ETV1* and *DGKB* were described to be linked to cognitive, language and auditory disorders, they were also found to be indirectly related with the traits of the cardiopathy, like the atrial septal defect and bronchial atresia.

The disruption of *NRXN1,* caused by the maternally inherited 473Kb deletion on 2p16.3, was reported to be the cause of the autistic behavior, and was found that may have additional indirect contribution in the development of the cardiopathy, especially in the formation of vascular defects.

Additively all the chromosomal abnormalities identified on proband's genome contribute for the development of the complex phenotype presented.

Although most of the disrupted and deleted genes only contribute indirectly to the development of the phenotypic traits, the reduction of functional copies of these genes might induce dramatic alterations on a variety of signalling pathways, altering the expression of genes reported to be directly related to the development of the defects.

To better understand the molecular pathogenesis, and with the aim of identifying any additional candidate gene, an analysis of all the protein coding genes contained in the TADs disrupted by the chromosomal breakpoints was carried out. It was unveiled that variations on *AFP* and *ALB* were found to be directly related with the development of atrial septal defects and hearing impairment, being also indirectly related with several phenotypic characters presented by the proband. Variations on *RTN4* were also identified to be directly associated with the development of atrial septal defects. Variations on *PNPT1*, *RPS27A*, *CXCL8*, *FSHR* and *LHGCR*, were found to be directly related with the development of hearing impairment. Although any of the genomic alterations present on proband's genome were reported to induce variation on the number of functional copies of *AFP, ALB, PNPT1, RTN4, RPS27A, CXCL8, LHGCR* and *FSHR*, alterations on the TADs where these protein coding genes are inserted, due to alterations on chromatin conformation may induce alterations in the expression levels of this genes having phenotypic effects.

The phenotypic consequences linked to the 130Kb deletion on der(4), that resulted from t(4;7)(q13.3;p21.2), and linked to the deletion on chromosome 3 are not yet completely clarified. Most of the human genes are poorly annotated, which difficult the task of establish the molecular pathogenesis. Further studies are required to clarify what are the real consequences of the genomic alterations in the phenotype presented by the proband in study.

In conclusion, coinheritance of the maternally derived deletion affecting *NRXN1* and the deletion at the breakpoint of the der(7) on 7p21.2, affecting the *DGKB*, *ETV1* and *LincRNA AC011288.2* appear to be the most contributively alterations for the proband's phenotype. Based on what is known at the time,

NRXN1, DGKB, *ETV1* and *LincRNA AC011288.2* were identified as being the most likely genes to be responsible for the proband's phenotype mainly characterised by cardiopathy and autism.

The realization of this master thesis work also confirmed that liWGS is a very powerful tool that will revolutionize the cytogenetic field, since it is the only available methodology that allows the analysis of the whole genome for high-throughput delineation of chromosomal rearrangements, permitting a better phenotype-genotype association.

7. Future work

To have the full perception of the phenotypic consequences associated with the genomic alterations fount on index' genome, it is vital to perform the expression studies and complete Hi-C analysis of index' and clarify the molecular pathogenesis underlying the presented phenotype.

Since the majority of candidate genes don't have significative expression on whole blood or LCLs was not possible to follow the research group routine procedures to assess the expression profile of these genes. Usually, gene expression microarrays performed from RNA extracted from the previous referred samples is the procedure carried out to assess gene expression profiles.

When possible to obtain urine samples from the family, the protocol of immortalization by electroporation with pCI-neo-hEST2 plasmid should be carried out. Plasmid culture and high quality extraction of the plasmidic DNA should be carried out like described on methods section. Urine samples must be processed to isolate the interest primary cell line, and an adequate electroporation protocol must be performed, for immortalization of the cell lines due to reactivation of the expression of hTERT. RNA extraction from the immortalized cells will permit to perform expression studies recurring to reverse transcriptase technology or to expression microarrays.

The establishment of the complete Hi-C protocol and its execution on this context is a priority, since in this case study, liWGS allowed, for example, the identification of a deletion within the breakpoint of der(7) that removes practically an entire TAD. So that, it is important continue investigating what are the implications of the unbalanced rearrangement t(4;7)(q13.3;p21.2) and of the deletions on chromosomes 2, 3, 9 and 11 on 3D conformation and organization of proband's genome. Hopefully, this analysis will allow to elucidate what were the molecular mechanisms underlying the formation of the chromosomal alterations, and how the gene regulation and expression are altered and contribute for the development of the complex phenotype presented by the proband.

8. References

Abrams DA, Nicol T, Zecker SG, Kraus N (2006) Auditory Brainstem Timing Predicts Cerebral Asymmetry for Speech. Journal of Neuroscience 26:11131–11137. doi: 10.1523/JNEUROSCI.2744-06.2006

Fahed AC, Gleb BD, Seidman JG, Seidman CE (2014) Genetics of Congenital Heart Disease: The Glass Half Empty. National Institute of Health 144:724–732. doi: 10.1038/jid.2014.371

American Association on Intellectual and Developmental Disabilities (2017) Definition of Intellectual Disability. In: Definition of Intellectual Disability http://aaidd.org/intellectual-disability/definition#.WQ4-DPnR_IV. Accessed 6 May 2017

American Heart Association (2017) Understand Your Risk for Congenital Heart Defects. In: Understand Your Risk for Congenital Heart Defects http://www.heart.org/HEARTORG/Conditions/CongenitalHeartDefects/UnderstandYourRiskforCongenitalHeartDefects/Understand-Your-Risk-for-Congenital-Heart-Defects_UCM_001219_Article.jsp#.WRcmpfnR_IU. Accessed 13 May 2017

American Psychiatric Association (2013) Diagnostic and Statistical Manual of Mental Disorders, 5th edition, American Psychiatric Publishing

Baht TA, Wani AA (2017) Chromosome Structure and Aberrations, 1st edition, Springer, India

Banerjee S, Riordan M, Bhat MA (2014) Genetic aspects of autism spectrum disorders: insights from animal models. Frontiers in Cellular Neuroscience 8:1–58. doi: 10.3389/fncel.2014.00058

Czepulkowski BH (2001) Analyzing Chromosomes, first edition, Oxford, USA

Bates SE (2011) Classical Cytogenetics: Karyotyping Techniques. In: Schwartz, Philip H.; Wesselschmidt RL (Eds) (ed) Human Pluripotent Stem Cells - Methods and Protocols. Springer Protocols, pp 369–387

Belaghzal H, Dekker J, Gibcus JH (2017) Hi-C 2.0: An Optimized Hi-C Procedure for High Resolution Genome-Wide Mapping of Chromosome Conformation. Elsevier 123:1–19. doi: 10.1016/j.ymeth.2017.04.004

Belton JM, McCord RP, Gibcus J, et al (2013) Hi-C: A comprehensive technique to capture the conformation of genomes Jon-Matthew. National Institute of Health 58:1–16. doi: 10.1016/j.ymeth.2012.05.001.Hi-C

Dabell MP, Rosenfeld JA, Bader P, et al (2013) Investigation of NRXN1 deletions: Clinical and Molecular characterization. American Journal of Medical Genetics 161:717–731. doi: 10.1002/ajmg.a.35780

David D, Marques B, Ferreira C, et al (2013) Co-segregation of trichorhinophalangeal syndrome with a t(8;13)(q23.3;q21.31) familial translocation that appears to increase TRPS1 gene expression. Human Genetics 132:1287–1299. doi: 10.1007/s00439-013-1333-0

Dixon JR, Selvaraj S, Yue F, et al (2012) Topological domains inmammalian genomes identified by

analysis of chromatin interactions. Nature 485:376–380. doi: 10.1038/nature11082

Dolk H, Loane M, Garne E (2011) Congenital heart defects in Europe: Prevalence and perinatal mortality, 2000 to 2005. Circulation 123:841–849. doi: 10.1161/CIRCULATIONAHA.110.958405

Drnasin K, Saraga-Babić M, Saraga M (2013) Clinical importance of pyelocalyceal dilation diagnosed by postnatal ultrasonographic screening of the urinary tract. Medical Science Monitor 19:125–131. doi: 10.12659/MSM.883786

Emren SV, Tülüce SY, Tülüce K (2015) Isolated Congenital Unilateral Agenesis of the Left Pulmonary Artery with Left Lung Hypoplasia in an Asymptomatic Adult Patient. Acta Cardiologica Sinica 31:572–575.

Gardner RJM, Sutherland GR, Shaffer LG (2012) Chromosome Abnormalities and Genetic Counseling, fourth edition. Oxford University Press, New York, p 1–634

Gipson MG, Cummings KW, Hurth KM (2009) Bronchial Atresia. RadioGraphics 29:1531–1535. doi: 10.1148/rg.295085239

Grollmuss O, Limongelli G, Joseph Dabal R, et al (2016) What is New in Genetics of Congenital Heart Defects? Frontiers in Pediatrics 4:1–6. doi: 10.3389/fped.2016.00120

Hemmat M, Yang X, Chan P, et al (2014) Characterization of a complex chromosomal rearrangement using chromosome , FISH , and microarray assays in a girl with multiple congenital abnormalities and developmental delay. Molecular Cytogenetics 7:1–6. doi: 10.1186/1755-8166-7-50

Hothi DK, Wade AS, Gilbert R, Winyard PJD (2009) Mild fetal renal pelvis dilatation: much ado about nothing? Clinical Journal of the American Society of Nephrology 4:168–177. doi: 10.2215/CJN.00810208

Houge G, Liehr T, Schoumans J, et al (2003) Ten years follow up of a boy with a complex chromosomal rearrangement: Going from a >5 to 15-breakpoint CCR. American Journal of Medical Genetics 118A:235–240. doi: 10.1002/ajmg.a.10106

illumina (2017) Large Whole-Genome Sequencing. In: Large Whole-Genome Seq. https://www.illumina.com/techniques/sequencing/dna-sequencing/whole-genome-sequencing/large-genomes.html. Accessed 22 Apr 2017

Inner Body Brain Stem. In: Anatomy. http://www.innerbody.com/image_nerv01/nerv46.html. Accessed 22 Aug 2017

Kim HG, Kishikawa S, Higgins AW, et al (2008) Disruption of Neurexin 1 associated with autism spectrum disorder. The American Journal of Human Genetics 82:199–207 doi: 10.1016/j.ajhg.2007.09.011

Kloosterman WP, Hochstenbach R (2014) Deciphering the pathogenic consequences of chromosomal aberrations in human genetic disease. Molecular Cytogenetics 7:1–12. doi: 10.1186/s13039-014-0100-9

Knight SJL, Regan R, Nicod A, et al (1999) Subtle chromosomal rearrangements in children with unexplained mental retardation. Lancet 354:1676–1681. doi: 10.1016/S0140-6736(99)03070-6

Liang WS, Aldrich J, Tembe W, et al (2014) Long insert whole genome sequencing for copy number variant and translocation detection. Nucleic Acids Research 42:1–12. doi: 10.1093/nar/gkt865

Life Technologies AmpliTaq ® DNA Polymerase PCR Reaction Setup.

Liu J, Bernier F, Lauzon J, et al (2011) Application of Microarray-Based Comparative Genomic Hybridization in Prenatal and Postnatal Settings: Three Case Reports. Genetic Research Insternational 2011:1–9. doi: 10.4061/2011/976398

Lupski JR (2009) Genomic disorders ten years on. Genome Medicine 1:1–42. doi: 10.1186/gm42

Lupski JR, Stankiewicz P (2005) Genomic Disorders : Molecular Mechanisms for rearrangements and Conveyed Phenotypes. PLoS Genetics 1:627–633. doi: 10.1371/journal.pgen.0010049

Luxon LM (1980) Hearing loss in brainstem disorders. Journal of Neurology Neurosurgery Psychiatry 43:510–515.

Millipore (2009) Amicon Ultra-0.5®, First. Millipore Corporation

National Heart Lung and Blood Institute, What Are Congenital Heart Defects? In: What Are Congenit. Hear. Defects? https://www.nhlbi.nih.gov/health/health-topics/topics/chd. Accessed 13 May 2017

National Library of Medicine (2017) What is a genome? In: Genetic Home Reference. https://ghr.nlm.nih.gov/primer/hgp/genome. Accessed 8 May 2017

Nature Education (2014) Histone/ Histones. In: Histone/ Histones. https://www.nature.com/scitable/definition/histone-histones-57. Accessed 8 May 2017

Network Autism (2017) Deafness and autism. http://network.autism.org.uk/good-practice/evidence-base/deafness-and-autistic-spectrum-disorder. Accessed 22 Aug 2017

National Health Institute - U.S. National Library of Medicine (2008) Saethre-Chotzen syndrome. In: Genet. Home Ref. https://ghr.nlm.nih.gov/condition/saethre-chotzen-syndrome. Accessed 20 Sep 2017

Ordulu Z, Wong KE, Currall BB, et al (2014) Describing sequencing results of structural chromosome rearrangements with a suggested next-generation cytogenetic nomenclature. American Journal of Human Genetics 94:695–709. doi: 10.1016/j.ajhg.2014.03.020

Orphanet (2017) Interatrial communication. In: Interatrial Communication. http://www.orpha.net/consor/cgi-bin/OC_Exp.php?lng=EN&Expert=1478. Accessed 21 Aug 2017

Pan Q, Hu H, Han L, et al (2016) Mapping Breakpoints of Complex Chromosome Rearrangements Involving a Partial Trisomy 15q23 . 1-q26 . 2 Revealed by Next Generation Sequencing and Conventional Techniques. PLoS One 11:1–14. doi: 10.1371/journal.pone.0154574

Pierpont ME, Basson CT, Benson DW, et al (2007) Genetic Basis for Congenital Heart Defects: Current

Knowledge. Circulation 115:3015–3038. doi: 10.1161/CIRCULATIONAHA.106.183056

Pinkel D, LI M (2005) Clinical Cytogenetics and molecular cytogenetics. Journal of Zhejian University 37:162–163. doi: 10.1038/ng1569

PreAnalytiX (2015) PAXgene® Blood RNA Kit Handbook, Second. PreAnalytiX by QIAGEN, Feldbachstrasse

QIAGEN (2015) QIAamp® DNA Blood Midi/Maxi Handbook, Fourth. QIAGEN, Austin

QIAGEN (2010) QIAamp® RNA Blood Mini Handbook, Second. QIAGEN, Austin

QIAGEN (2012) HiSpeed® Plasmid Purification Handbook, First. QIAGEN

Rafati M, Ghadirzadeh MR, Heshmati Y, et al (2012) Familial versus sporadic intellectual disability: contribution of subtelomeric rearrangements. Molecular Cytogenetics 5:1–4. doi: 10.1186/1755-8166-5-4

Ramani V, Cusanovich DA, Hause RJ, et al (2016) Mapping 3D genome architecture through in situ DNase Hi-C. Nature Protocols 11:2104–2121. doi: 10.1038/nprot.2016.126

Rao SSP, Huntley MH, Durand NC, et al (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159:1665–1680. doi: 10.1016/j.cell.2014.11.021

Reading DW, Oza U (2012) Unilateral absence of a pulmonary artery: a rare disorder with variable presentation. Baylor University Medical Center Proceedings 25:115–118.

Reardon W, McManus SP, Summers D, Winter RM (1993) Cytogenetic evidence that the Saethre-Chotzen gene maps to 7p21.2. American Journal of Medical Genetics 47:633–636. doi: 10.1002/ajmg.1320470510

Richards AA, Garg V (2010) Genetics of congenital heart disease. Current Cardiology Reviews 6:91–97. doi: 10.2174/157340310791162703

Ropers HH (2010) Genetics of Early Onset Cognitive Impairment. Annual Reviews on Genomics and Human Genetics 11:161–187. doi: 10.1146/annurev-genom-082509-141640

Samsam M, Ahangari R, Naser SA (2014) Pathophysiology of autism spectrum disorders: revisiting gastrointestinal involvement and immune imbalance. World Journal of Gastroenterology 20:9942–9951. doi: 10.3748/wjg.v20.i29.9942

Schmitt AD, Hu M, Ren B (2016) Genome-wide mapping and analysis of chromosome architecture. Nature Reviews on Molecular Cell Biology 17:743–755. doi: 10.1038/nrm.2016.104

Shirai Y, Kouzuki T, Kakefuda K, et al (2010) Essential Role of Neuron-Enriched Diacylglycerol Kinase (DGK), DGKβ in Neurite Spine Formation, Contributing to Cognitive Function. PLoS One 5:e11602. doi: 10.1371/journal.pone.0011602

Singh U, Jhim D, Kumar S, et al (2015) Unilateral agenesis of the lung: a rare entity. American Journal of Case Reports 16:69–72. doi: 10.12659/AJCR.892385

Sismani C, Kitsiou-Tzeli S, Ioannides M, et al (2008) Cryptic genomic imbalances in patients with de novo or familial apparently balanced translocations and abnormal phenotype. Molecular Cytogenetics 1:1–15. doi: 10.1186/1755-8166-1-9

Stankiewicz P, Lupski JR (2010) Structural Variation in the Human Genome and its Role in Disease. Annual Reviews of Medicine 61:437–455. doi: 10.1146/annurev-med-100708-204735

Talkowski ME, Ernst C, Heilbut A, et al (2011) Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. American Journal of Human Genetics 88:469–481. doi: 10.1016/j.ajhg.2011.03.013

The Jackson Laboratory (2016) MGI - The international database resource for the laboratory mouse. In: Mouse Genome Informatics. http://www.informatics.jax.org/. Accessed 31 Aug 2017

Utami KH, Hillmer AM, Aksoy I, et al (2014) Detection of chromosomal breakpoints in patients with developmental delay and speech disorders. PLoS One 9:1–10. doi: 10.1371/journal.pone.0090852

Vaidutis, Eglė P, Laima A, et al (2016) Identification of genetic causes of congenital neurodevelopmental disorders using genome wide molecular technologies. Acta Medica Lituanica 23:73–85. doi: 10.6001/actamedica.v23i2.3324

Vissers LELM, Gilissen C, Veltman JA (2015) Genetic studies in intellectual disability and related disorders. Nature Reviews Genetics 17:9–18. doi: 10.1038/nrg3999

Wan TS-K (2010) Molecular cytogenetics: techniques, developments and applications. Journal of Hong Kong Institute of Medical Laboratory Sciences 12:1–12.

Weizmann Institute of Science (2017) GeneCards - Human Genes; Gene Database; Gene Search. In: GeneCards. http://www.genecards.org/. Accessed 27 Aug 2017

World Health Organization (2017) Congenital anomalies. In: Congenital Anomalies (Kyoto). http://www.who.int/mediacentre/factsheets/fs370/en/. Accessed 26 Apr 2017

Zarzosa-Alvarez AL, Sandoval-Cabrera A, Torres-Huerta AL, Bermudez-Cruz RM (2010) Electroeluting DNA fragments. Journal of Visualized Experiments 43:1–2. doi: 10.3791/2136

Zhang F, Carvalho CMB, Lupski JR (2009) Complex human chromosomal and genomic rearrangements. Cell Press 25:298–307. doi: 10.1016/j.tig.2009.05.005

Zoghbi HY, Bear MF (2012) Synaptic dysfunction in neurodevelopmental disorders associated with autism and intellectual disabilities. Cold Spring Harborb Perspectives in Biology 4:1–22. doi: 10.1101/cshperspect.a009886

Zymo Research Corporation (2013) ZR Plasmid Miniprep™ - Classic, First. Zymo Research Corporation

Appendixes

Appendix 1 – Establishment of permanent growing lymphoblastoid cell lines (Adapted from Neitzel, 1986)

1. Peripheral blood collection

- Collect 3-15mL of peripheral blood by venipuncture into an appropriate vessel containing sodium heparin preservative.

Blood samples should be stored and shipped at RT.

EDTA or sodium citrate tubes can be used as well with similar results

It is better to use samples that are not older than 48h. However, it is possible to establish LCL's from older samples

2. Isolation of PBMCs via density gradient

All subsequent steps should be carried out in a class II biological safety cabinet.

- Dilute the peripheral blood (1:1) with sterile complete RPMI-DM without FBS-HI in an upright, appropriately sized sterile tube.

- Prepare tubes with ficcoll-paque plus according to the amount of blood:

Option 1: Blood volume of 3mL + 3mL of medium (6mL total)

Prepare one leucosep tube (12 mL), with membrane, with 3mL of ficoll-paque plus and centrifuge 1000g for 30seconds;

Option 2: Blood volume 3 to 14mL + 3 to 14mL of medium (6-28mL total)

Prepare falcon tubes with ficoll-paque respecting the ratio 1:1:1 between ficoll-paque, blood and dilution medium and centrifuge 1000xg for 30s.

Option 3: Blood volume of 15mL + 15mL of medium (30mL total)

Prepare one tube leucosep, with membrane, with 15mL of ficoll-paque plus and centrifuge 1000xg for 30s;

- Carefully and slowly layer diluted blood on Ficoll-Paque Plus. Allow the blood to run down the side of the tube, on an angle of 45° to avoid mixing of the layers, using a disposable plastic Pasteur pipette.

- Centrifuge the tubes, with the centrifuge brake off, since it allows the proper separation:

Option 1: 12mL leucocep tube or 15mL falcon tubes: - Centrifuge Heraeus Multifuge 3s: 800xg for 15 min; - 581DR Eppendorf Centrifuge: 800xg for 15 min;

Option 2: 50mL leucocep of falcon tubes: - Centrifuge Heraeus Multifuge 3s: 20-22 min at 635g; - 581DR Eppendorf Centrifuge 20-22 min at 635g.

- Carefully transfer each cloudy white interphase layer containing MNCs to a new 15mL conical tube with a plastic Pasteur pipette and bring the volume up to the maximum capacity (14mL) of the tube with complete RPMI-DM without serum.

- Centrifuge the tubes, with the centrifuge brake on:

Option 1: Centrifuge Heraeus Multifuge 3s: Centrifuge the cells at 300xg for 10 min.

Option 2: 581DR Eppendorf Centrifuge: Centrifuge the cells at 300xg for 10 min.

- Discard the supernatant, ressuspend the pellet in residual liquid and bring the volume up to 10mL with complete RPMI-DM without serum.
- Centrifuge the cells with the centrifuge brake on:

Option 1: Centrifuge Heraeus Multifuge 3s: Centrifuge the cells at 230xg for 10 min.

Option 2: 581DR Eppendorf Centrifuge: Centrifuge the cells at 230xg for 10 min.

- Discard the supernatant and ressuspend the cells according to their finality. Count the cells if needed.
- Aliquot the cells as needed for expansion.

PBMCs can be frozen for later use. Combine 5 million ($5x\ 10^6$) or more cells per mL of cell freezing medium. Aliquot 1 mL per vial and freeze the vials in a controlled-rate freezing container. Cells can be stored in a liquid nitrogen storage container (-150°C) for a prolonged time. Recovery of ≥80% is expected after proper thawing. Apart from freezing, excess cells can be pelleted for future DNA and RNA extraction. Cell pellets should be stored at -80°C.

3. Transformation of blood lymphocytes with EBV

- Pre-warm two leyton tubes per sample in a 37°C incubator and 1.6mL of filtered EBV medium per sample in 37°C water bath.
- Ressuspend the PBMC pellet in 1.6mL of filtered warmed EBV medium
- Subdivide in 2 Leyton tubes and incubate both for 1-2h at 37ºC in water bath. For thawed PBMC'S increase incubation time for 2.5h.
- When incubation time is ended, quickly observe at the optical microscope to evaluate the formation of cell clumps. If there is not a good amount, increase the incubation time
- Add to each tube 800µL of complete RPMI-DM medium (30% FBS-HI Performance plus) with cyclosporine A (2X), previously heated 37ºC.

Cyclosporine A is a immunosupressor drug that facilitates EBV infected-B cell outgrowth over T lymphocytes, increasing transformation efficiency. This effect is probably due to the inhibition of immune T-Lymphocytes that would destroy EBV infected cells.

- Incubate 72h (37 ºC). Replace approximately half the medium with complete fresh RPMI with cyclosporine A (1X), previously heated 37ºC. Gently ressuspend the aggregates.
- Replace nearly half the medium every 5-7 days with fresh medium with 30% serum. If cell aggregates are not observed keep adding cyclosporine A (1X) until fourteenth day of culture. After that, stop adding cyclosporine. Alternatively, continue adding cyclosporine until the first sub culture.
- Evaluate if the cell density is enough for culture division into 2 leyton tubes or 1 tissue culture flask. Only one leyton should be observed. The other is manipulated only for medium changes.

Reagents:

Complete RPMI medium (LCL infection)

	Final concentration	For 10mL complete medium
RPMI 1640 Dutch Modified		7 mL
Glutamin 200mM	2mM	1µL
Pen/Strep (Gibco, 15140-148; lot#125675)	1.5%	150µL
FBS-Heat inactivated (Performance plus Gibco 10082-147; lot#1366479)	30%	3mL

Complete RPMI medium (LCL maintenance)

To 500mL of RPMI-DM add:

- 5mL of Glutamine 200Mm;
- 7.5mL of PEN/STREP.
- Aliquot 100mL and store it.

To the remaining 400mL of medium add 80mL of FBS-Heat inactivated for a final concentration of 20% serum.

Freezing Medium

- 90% FBS-HI;
- 10% DMSO (Cells should be in contact with DMSO at room temperature the least amount of time possible).

Cyclosporine-A preparation

- From Cyclosporine A - Sandimmun Novartis Farma 50mg/mL
- Stock Solution (50mg/mL): Carefully open the vial and transfer the content (1mL) to an identified 1.5mL eppendorf tube. Protect from light;
- Working Solution (1mg/mL ⇔1µg/µL): Transfer 100µL of Stock solution to an identified vial with 5mL absolute ethanol. Protect from light;
- Complete medium 1µg/µL – Prepare at time of use: Add 5µL of Cyc-A Working Solution to 5mL of complete medium RPMI-DM with 30% Fbs-Hi Performance Plus.
- Complete medium 0.5µg/µL – Prepare at time of use: Add 2.5µL of Cyc-A Working Solution to 5mL of complete medium RPMI-DM with 30% Fbs-Hi Performance Plus.

(Adapted from PAXgene® Blood RNA Kit Handbook (PreAnalytiX 2015))

Blood

Mix

Centrifuge 15min at 4600rpm

Wash pellet
Rnase-Free water (RNFW)
Vortex until pellet is visibly dissolved

Centrifuge 20min at 4600rpm

Ressuspend
Buffer BR1
Vortex until pellet is visibly dissolved

Transfer to microcentrifuge tube (MCT)

Add proteinase K (PK) and binding buffer (BR2) Vortex 5s

Incubate 15min at 55ºC using a shaker (900rpm)

Tranfer to PAXgene Shredder Spin Column (PSC)

Centrifuge 3min at max speed

Tranfer the flow-through to microdentrifuge tube

Add etanol 100% Vortex until mix

Bind total RNA
Load on PAXgene RNA spin column (PRC)
Centrifuge 1min at 10000*g (2x)

Wash
Buffer BR3
Centrifuge 1min at 10000*g

Digest DNA
Dnase I
Incubate 30min at room temperature

Wash 3X
1: Buffer BR3
Centrifuge 1min at 10000*g
2: Buffer BR4
Centrifuge 1min at 10000*g
3: Buffer BR4
Centrifuge 3min at 10000*g

Centrifuge 1min at 10000*g

Elute
Buffer BR5
Incubate 30s at room temperature
Centrifuge 1min at 10000*g

Ready to use RNA

Sample
whole bood or LCL pellets

Lyse
Proteinase K + AL buffer
(65ºC; 30 min)

Bind
Ethanol 100%
Centrifuge 3min at 3000rmp

DNA

Wash
AW1
Centrifuge 1min at 4600rmp

Wash
AW2
Centrifuge 15min at 4600rmp

Elute
Low TE, 70ºC
Centrifuge 2 min at 4600rpm

Ready to use DNA

Appendix 4 – "QIAamp® RNA Blood Mini" RNA extraction from LCL Workflow (Adapted from QIAamp® RNA Blood Mini Handbook (QIAGEN 2010))

Sample
0,8*10^7 cells per LCL pellet

Lyse
Buffer RLT
Pipet until "single cell"

Homogenization with "QIAshredder"
Centrifuge 2min at max speed

Ethanol 70%
Pipet al the lysate to
"QIAamp spin column"

Bind
Centrifuge 15s at 8000*g

DNaseI treatment
Buffer RW1→Centrifuge 15s at 8000*g→DNaseI→Incubate 30min at room temperature→ Centrifuge 15s at 8000*g

Wash 3X
1: Buffer RPE
Centrifuge 15s at 8000*g

2: Buffer RPE
Centrifuge 3min at max speed

3: Centrifuge 1min at max speed

Elute 2X
1: Rnase Free Water
Centrifuge 1min at 8000*g
2: Rnase Free Water
Centrifuge 1min at 8000*g

Ready to use RNA

Appendix 5 – "Amicon® Ultra-0.5 Centrifugal Filter Devices" PCR product purification Workflow (Adapted from Amicon® Ultra-0.5 Centrifugal Filter Devices user guide (Millipore 2009))

Load the Sample

Add 500µl of PCR product + Low TE buffer to the spin column matrix

Centrifuge 17min at max speed

Discard the flow through

Add 300µl of Low TE buffer to the spin column matrix

Centrifuge 12min at max speed

Discard the flow through

Invert the spin column, and place it on a fresh tube

Elute

Centrifuge 2min at 1000xg
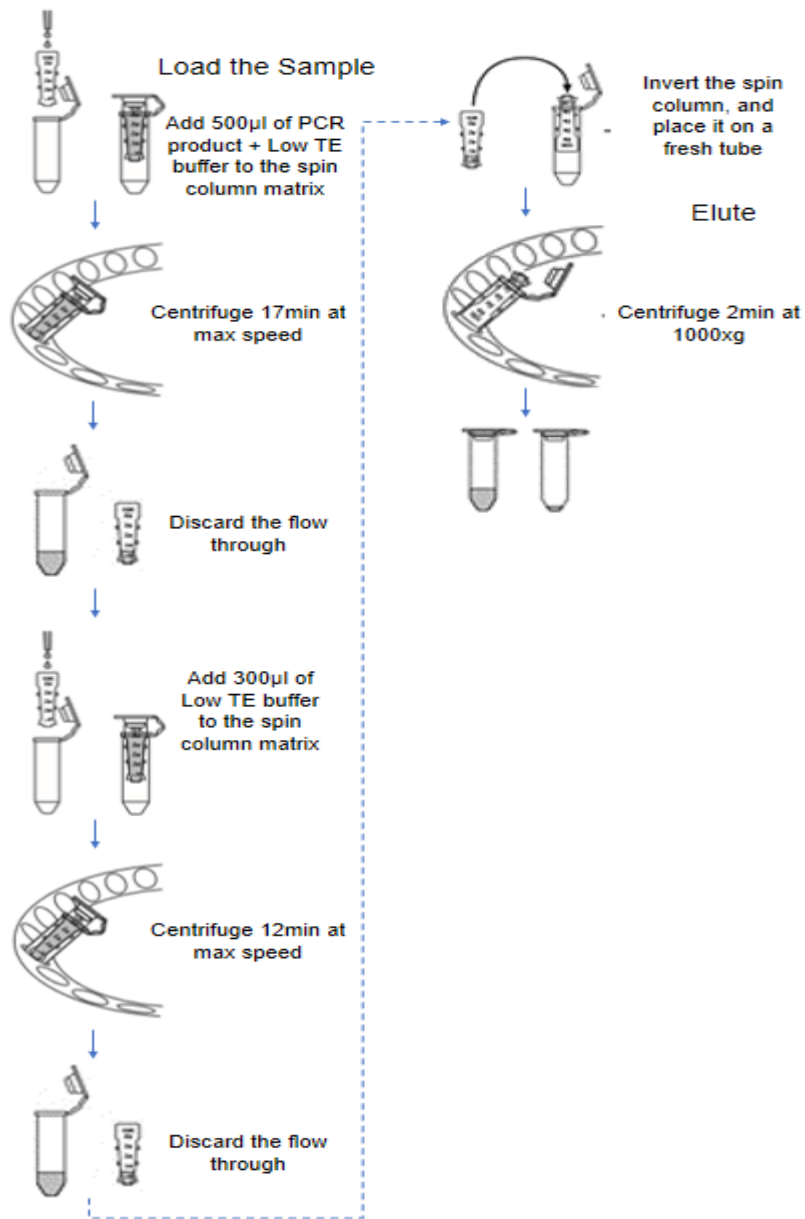
Appendix 6 – "ZR Plasmid MiniprepTM - Classic" Plasmid extraction from Miniprep Cultured Bacteria Workflow (Adapted from ZR Plasmid Miniprep™-Classic (Zymo Research Corporation 2013))

0.5 - 5ml of Bacterial Culture

Centrifuge 20s at max speed

Ressuspend
P1 Buffer
Pipet or Vortex until pellet is visibly dissolved

Lyse
P2 Buffer
Mix by inverting 2-4 times

Cells are completely lysed when the solution appears clear, purple, and viscous; Incubate 1-2 minutes at room temperature

Neutralization
P3 Buffer
Mix gently but thoroughly
Incubate 1-2 minutes at room temperature

Centrifuge 2min at max speed

Transfer the supernatant into the "Zymo-Spin™ IIN column"

Bind
Centrifuge 30s at max speed

Wash 2X
1: Endo-Wash Buffer
Centrifuge 30s at max speed
2: Plasmid Wash Buffer
Centrifuge 1min at max speed

Elute
Low TE
Centrifuge 30s at max speed

Ready to use Plasmid DNA

50 - 150ml of Bacterial Culture

Centrifuge 15min at 6000*g, at 4ºC

Alkaline Lysate

Lyse

1: Buffer P1 + Buffer P2
Incubate at room temperature for 5min

2: Buffer P3

Fill in the "QIAfilter Cartrige" and incubate
at room temperature for 10 min

Clear lysate by filtration using
QIAfilter Cartridge

Equilibrate a "HiSpeed Midi Tip" with
Buffer QBT

Bind DNA
Allow the clear lysate to enter the resin by
gravity flow

Wash DNA
Buffer QC; Gravity flow

Elute DNA
Buffer QF; Gravity flow

Percipitate DNA

Isopropanol; Incubate at room
temperature for 5min

Collect DNA using QIAprecipitator

Transfer the eluate/isopropanol mix into a
20ml serynge and filter it through the
QIAprecipitator

Wash DNA
70% etanol

Dry QIAprecipitator membrane

Elute DNA
TE or Low TE Buffer

Utrapure ready to use Plasmid DNA

Making LB agar plates

1. In a water bath, liquefy the LB agar media;
2. Wipe the bench with 70% ethanol with a clean paper towel and prepare the number of plates you will need;
3. Add the selection antibiotic to the LB agar media at the appropriate concentration;
4. Near a Bunsen burner, pour the LC agar media into the plate in order to have around 1cm of media;
5. Leave your plates at room temperature inside a tissue culture hood overnight to dry;
6. After drying, store the plates upside down at 4ºC inside plastic bags to prevent condensation.

Streaking and Isolating Bacteria on an LB Agar Plate

1. Obtain an LB agar plate with appropriate antibiotic and label the bottom of the plate with the plasmid name, the date, the antibiotic resistance and your initials.
2. Keep your lab bench area sterile by working near a flame or Bunsen burner.
3. Using a sterile toothpick or pipette tip, touch the bacteria growing within the punctured area of the stab culture.
4. Divide the petri plate in 4 quadrants. Run the toothpick lightly over the 2 upper quadrants, making 4 horizontal lines. Sterilize the toothpick on the Bunsen burner and make 4 vertical lines on second and third quadrants, passing through the previous done horizontal lines to spread the bacteria over streak the plate. Sterilize the toothpick once more on the Bunsen burner and make 4 horizontal lines on third and fourth quadrants. Isolated colonies should appear on the fourth quadrant.
5. Incubate the plate overnight (12-18h) at the designated growth temperature.
6. Check for single colonies. A single colony should appear as a white dot growing on the solid medium. See the red spot above in the figure.

Miniprep Inoculation

1. Once you have single colonies, proceed to inoculate miniprep cultures.
2. Prepare the culture tubes by adding 2-5mL of liquid LB with the appropriate antibiotic.
3. Using a sterile toothpick, select a single colony from your streaked LB agar plate, and merge the toothpick into the liquid LB + antibiotic and swirl.
4. Loosely cover the culture tube with the cap that, and incubate the bacterial culture at 37°C for 12-18h in a shacking incubator at 300rpm.
5. After incubation, check for growth, which is characterized by a cloudy haze in the media.

Note: A good negative control is LB media + antibiotic without any bacteria inoculated. You should see no growth in this culture after overnight incubation.

6. You can now isolate your plasmid DNA from the bacterial culture.

Plasmid extraction

Plasmidic DNA was extracted from miniprep cultures using ZR Plasmid Miniprep™-Classic kit, following the manufacturer's instructions like is schematized on Appendix 6.

Starter culture Inoculation

1. Prepare the culture flasks by adding 9mL of liquid LB with the appropriate antibiotic.

2. Add 1mL of the cultured midipreps to the culture flasks previously prepared.

3. Loosely cover the culture flasks with the cap that, and incubate the bacterial culture at 37°C for 12-18h in a shacking incubator at 300rpm.

4. After incubation, check for growth.

Midiprep Inoculation

1. Prepare the culture flasks by adding 60mL of liquid LB with the appropriate antibiotic.

2. Dilute the starter cultures 1/500 to 1/1000 into liquid LB with the appropriate antibiotic, and inoculate the culture flasks

3. Loosely cover the culture flasks with the cap that, and incubate the bacterial culture at 37°C for 12-18h in a shacking incubator at 300rpm.

4. After incubation, check for growth.

Isolation of Plasmid DNA with the HiSpeed Plasmid Midi Kit (Qiagen) - Midiprep

Before starting:

- Add RNAse A to Buffer P1 (skip this step if already added);
- Check Buffer P2 for SDS precipitation, if needed warm the buffer at 37ºC until no precipitate is visible;
- Pre-chill Buffer P3 at 4ºC;
- Optional: Add LyseBlue to the volume of Buffer P1 you will use in during the protocol.

Procedure:

1. Harvest the bacterial cells by centrifugation at 6000xg for 15 min at 4ºC;
2. Remove the supernatant completely!
   Note: If needed it is possible to stop the protocol and continue later by freezing the dry pellet at -20ºC.
3. Ressuspend the pellet in 6mL of Buffer P1;
4. Add 6mL of Buffer P2, mix thoroughly by vigorously inverting the tube 4-6 times and incubate at room temperature for 5 min. DO NOT VORTEX!
   During the incubation prepare the QIAfilter Cartridge by screwing the cap onto the outlet nozzle.

5. Add 6mL of chilled Buffer P3 to the lysate and mix immediately by vigorously inverting 4-6times;

6. Pour the lysate into the barrel of the QIAfilter Cartridge and incubate a RT for 10 min. DO NOT INSERT THE PLUNGER!

    During the incubation a precipitate will float and form a layer on top of the solution, if after the incubation this precipitate has not floated to the top of the solution carefully run a sterile pipette tip around the walls to dislodge it.

7. Equilibrate a HiSpeed Midi Tip by applying 4mL of Buffer QBT and allow the column to empty by gravity flow;

8. Remove the cap from the QIAfilter outlet nozzle. Gently insert the plunger into the QIAfilter Midi and filter the cell lysate into the HiSpeed tip (app. 15mL are recovered);

    Remove a 300uL sample of the filtered lysate and save for analytical gel to determine whether growth and lysis conditions were optimal.

9. Allow the clear lysate to enter the resin by gravity flow;

    Remove a 300uL sample of the flow-through and save for analytical gel to determine the efficiency of DNA binding to the QIAGEN resin.

10. Wash the HiSpeed Midi Tip with Buffer QC by gravity flow;

    Remove a 400uL sample of the wash fraction and save for analytical gel.

11. Elute DNA with 5mL of Buffer QF. Collect the eluate in a tube with a minimum capacity of 10mL;

    Remove a 100uL sample of the eluate and save for analytical gel.

12. Precipitate DNA by adding 3.5mL of room-temperature isopropanol to the eluted DNA;

13. Mix and incubate a RT for 5 min.

14. During the incubation remove the plunger from a 20mL syringe and attach the QIAprecipitator Midi Module onto the outlet nozzle. DO NOT USE EXCESSIVE FORCE, BENDING OR TWISTING to attach the QIAprecipitator;

    Always remove the QIAprecipitator from the syringe before pulling up the plunger.

15. Place que QIAprecipitator over a waste bottle, transfer the eluate/isopropanol mixture into the 20mL syringe and insert the plunger. Filter the eluate/isopropanol mixture through the QIAprecipitator using constant pressure;

    The following steps (19-29) must be performed within 10 minutes in order to prevent DNA loss.

16. Remove the QIAprecipitator from the 20mL syringe and pull out the plunger. Re-attach the QIAprecipitator and add 2mL of 70% ethanol to the syringe;

17. Wash the DNA by inserting the plunger and pressing the ethanol through the QIAprecipitator using constant pressure;

18. Remove the QIAprecipitator from the 20mL syringe and pull out the plunger;

19. Attach the QIAprecipitator to the 20mL syringe, insert the plunger and dry the membrane by pressing air through the QIAprecipitator quickly and forcefully;

20. Repeat the last step;

21. Dry the outlet nozzle of the QIAprecipitator with absorbent paper to prevent ethanol carryover;

22. Remove the plunger from a new 5mL syringe and attach the QIAprecipitator onto the outlet nozzle;

23. Hold the outlet of the QIAprecipitator over a 1.5mL collection tube. Add 1mL of Buffer TE to the 5mL syringe, insert the plunger and elute the DNA to the collection tube using constant pressure.
BE CAREFUL the residual elution buffer in the QIAprecipitator tends to foam!
Do not use less than 500uL to elute the plasmid, since incomplete wetting of the QIAprecipitator will lead to reduced DNA yields.
Water or other buffers may also be used to elute the DNA. Buffer TE contains EDTA which may inhibit downstream enzymatic or sequencing reactions.

24. Remove the QIAprecipitator from the 5mL syringe, pull out the plunger and reattach the QIAprecipitator to the 5mL syringe;

25. Transfer the eluate from step 26 to the 5mL syringe and elute for a second time into the same 1.5mL collection tube.

26. Quantify the DNA in Nanovue or Nanodrop;

27. Store the DNA at -20ºC.

Glycerol Stock Inoculation

1. Prepare 2mL cryovials with 500µL of 50% (wt/v) glycerol solutiom

2. Add 500 µL of the midipreps to the previous prepared cryovials.

3. Snap freeze the glycerol stocks by storing it at -80ºC. Glycerol stocks are important for long-term storage of plasmids.

Appendix 9 – *in situ* Hi-C Nuclei Isolation and Staining protocol for light microscopy

(based on Belaghzal et al. 2016; Rao et al. 2014 and Ramani et al. 2016 protocols)

PROTOCOL FOR MICROSCOPE OBSERVATION W/ CELL STAINING

4 ALIQUOTS:

C = control (without buffer exchange from PBS to Hi-C Lysis Buffer)

A-20 = 20 Potter passages

A-40 = 40 Potter passages

P = 40 Pasteur passages

1- Count 2.5 million cells per aliquot (total 10 million cells)

2- Centrifugation: 300xg, 10 min, RT

3- Ressuspend in PBS to a density of 0.5 million per mL → 50mL, divide in 2 tubes of 25mL

4- Crosslink: add formaldehyde to cf=1% → 712.5µL formaldehyde 35% to each tube

5- Incubation exactly 10 min, RT, tube inversion every 2 min

6- Quenching: add glycine to cf=0.15M → 1.5mL glycine 2.5M to each tube

7- Incubate 5 min RT + over 15 min ice

8- Centrifugation: 300xg, 10 min, 4ºC

9- Ressuspend each tube to 1mL cold PBS (0.5 mL suspension per aliquot), divide suspension in 4 aliquots of 0.5mL

10- Repeat centrifugation

[prepare 2mL Hi-C Lysis Buffer (w/ 20µL PIC)]

11- Ressuspend one aliquot in 0.5mL PBS (= sample C). Ressuspend three others with 0.5mL Hi-C Lysis Buffer. Leave all on ice.

[homogenization must take less than 15min to make]

12- Hi-C aliquots: after 5min on ice (with gentle inversions), start homogenization:

sample A-20 = 2 × 20 Potter

sample A-40 = 2 × 40 Potter

sample P-40 = 2 × 40 Pasteur

[between each set, leave 1min on ice]

[leave on ice, with gentle inversions until time reaches 15min]

13- Wash nuclei with 2 centrifugations of 2500*g 5min, ressuspend first time in 0.5mL PBS, ressuspend second time in 250µL PBS.

14- Sample staining, through 1:1 mixing of 10µL of each sample with 10µL Tripan Blue or Light Green 1% (wt/v) or Giemsa solution 8% (v/v).

15- Smear 5µL of stained sample onto glass slide, place coverslip above and add nail polish to margins of coverslip.

16- Observe at optical microscope with oil immersion.

Appendix 10 – *in situ* Hi-C Protocol Developed for the First partial Experimental Assay

(based on Belaghzal et al. 2016; Rao et al. 2014 and Ramani et al. 2016 protocols)

DAY 1 - cell crosslink and nuclei isolation

1. Prepare cell suspension of $5×10^6$ cells per aliquot (total of cells can be in one suspension, and after crosslink they are divided into aliquots)

TAKE 36µL ALIQUOT FOR CONTROL (stored at 4ºC in culture medium)

2. Centrifuge at 300×g, 10min, RT

3. Ressuspend cells in PBS, to a density of $0.5×10^6$ cells/mL

4. Add formaldehyde to cf=1% - equivalent of 270µL formaldehyde 37% or 285µL formaldehyde 35%, to each 10mL of cell suspension.
    Quickly mix tube several times by tube inversion.

5. Incubate for exactly 10 min, at RT. Gentle inversion of tub every 2min.

6. Quench formaldehyde with glycine addition for a cf = 0.15M – equivalent of 600µL glycine 2.5M to each 10mL suspension. Mix well by inversion.

7. Incubate 5 min at RT, then at least 15 min on ice.

8. Centrifuge cell suspension: 300×g, 5 min, 4ºC

9. Discard supernatant, ressuspend in 1mL ice-cold PBS.

10. Repeat centrifugation at step 8, discard supernatant.

11. Add to cell pellet 1mL ice-cold Hi-C Lysis Buffer.

12. Leave incubating 15 min on ice. In the middle of the incubation, perform Potter and Pasteur pipette homogenization: perform 2 sets, between each set leave sample to rest on ice for 1min.

13. Centrifuge at 2500×g, 5 min, 4ºC.

14. Discard supernatant, add 500µL NEB 2.1 buffer. Leave sample equilibrating on ice (or 4ºC) for 1h.

15. Repeat step 13 and 14.

16. Centrifuge again for 2500×g, 5 min, 4ºC. After discarding supernatant, add 388µL ice-cold NEB 2.1

TAKE 46µL ALIQUOT FOR CONTROL:
10µL for nuclei count, 36µL for DNA extraction (stored at 4ºC w/ EDTA cf=0.5mM)

17. Add 38µL SDS 1% (to a cf=0.1%, vf= 380µL)

18. Incubate exactly 10min at 65ºC, then put sample on ice.

19. Add 43µL of 10% Triton X-100 (to a cf=1%, vf= 423µL)

20. Add 8.5µL of 10× NEB2.1 (equivalent to added SDS and Triton X-100, and volume of restriction enzyme to be added for overnight incubation) (vf= 431.5µL)

21. Add 4µL of HindIII (equivalent to 400U), mix gently

22. Overnight digestion at 37ºC, with 300rpm incubation (program for 16h)

DAY 2 – restriction fragment fill-in and ligation

1. 14 hours after beginning overnight digestion, add 1µL of HindIII, leave incubating in equal conditions for 2 more hours. (vf= 432.5µL)

2. Perform enzyme inactivation – for HindIII, 20 min incubation at 80ºC. In the end, put sample on ice and prepare Fill-in Mix.

TAKE 46µL ALIQUOT FOR CONTROL (vf= 386.5µL)

3. Add 60µL Fill-in Mix to sample, mix gently. (vf= 446.5µL)

4. Incubate at 37ºC, for 1h, with 300rpm agitation.

5. Prepare Ligation Mix (753.5µL) and add to sample (vf= 1200µL)

6. Incubate for 4h, at 16ºC, with 300rpm agitation.

7. Add 25µL of 20mg/mL Proteinase K. Incubate over 2h at 65ºC.

8. Add another 25µL of 20mg/mL Proteinase K. Incubate overnight (16h) at 65ºC.

TAKE 46µL ALIQUOT FOR CONTROL

(vf= 1204µL)

Hi-C Lysis Buffer	NEB2.1	Fill-in Mix	Ligation Mix
10mM Tris (pH 8.0)	10mM Tris (pH 7.9)	1.5µL of 10mM dATP	120µL of 10× NEB T4 DNA Ligase Buffer
10mM NaCl	50mM NaCl	1.5µL of 10mM dTTP	120µL 10% Triton X-100
0.2% NP-40	10mM MgCl$_2$	1.5µL of 10mM dGTP	6µL of 20g/mL BSA
1× Protease Inhibitor Cocktail (freshly added)	100µg/mL BSA	37.5µL of 0.4mM biotin-14-dCTP	5µL of 400U/µL T4 DNA Ligase
		6µL of 10× NEB	502.5µL Braun water
		10µL of 5U/µL DNA Polymerase I (Klenow)	
		2µL Braun water	

Addendum to Masters dissertation: "Characterization of a chromosome rearrangement associated   with cardiopathy and autism", by Sara Melo Dias

This addendum was formulated as consequence of the alteration of some assumptions and conclusions of the referred masters thesis, completely unrelated with the researchers involved in the study.

Cytogenetic analysis of the index patient, carried out prior to the study, revealed the existence of a seemingly *de novo* chromosomal translocation t(4;7)(q21.1;p21.2)dn. Furthermore, it was proven by array CGH that the index patient inherited a maternal 490 kb deletion on 2p16.3 and a paternal 742 kb deletion on 7p21.2.

Based on whole genome sequencing data the translocation was interpreted as unbalanced with a 130 kb deletion at the 4q13.3 breakpoint and a 1351 kb deletion at the 7p21.2 breakpoint.

Until the masters thesis submission deadline, the unavailability of father's DNA sample, allied with the fact that the father do no present phenotype, made it possible to reckon the unbalanced chromosomal translocation t(4;7)(q13.3;p21.2) as *de novo* translocation with possible originated by the paternal 742 kb deletion on 7p21.2.

Once the father's DNA was available for testing, we realised that he is also carrier of the rearrangement and therefore the translocation is not a de novo chromosomal rearrangement.

Reinterpretation of the sequencing data revealed that the apparently unbalanced translocation is in fact a "chromothripsis-like" complex chromosomal rearrangement (CCR), with seven breakpoints within the same 4q13.3 and 7p21.2, chromosome regions. The results of the reinterpreted sequencing data are summarized in table 1.

Table 1 - Summary of the reinterpreted sequencing data (reference sequence: human genome assembly [GRCh38/hg38])

Cytogenetic karyotype	Type of alteration		Cytoband	Position [hg38]	Cytoband	Position [hg38]	Break point delimited region [hg38]	Alteration size (bp)
46,XY,t(4;7)(q21.1;p21.2)	**A**	**Translocation**	**4q13.3**	**74,049,528**	**7p21.2**	**14,536,000**	**chr7:13,184,731-14,536,001**	**---**
	B	**Translocation**	**4q13.3**	**73,918,923**	**7p21.3**	**13,184,730**	**chr4:73,918,924-74,049,529**	**---**
		Excision/Insertion	**7p21.3**	**9,347,909**	**7p21.3**	**12,969,278**	**chr7:9,347,909-9,353,608**	**---**
		Excision/Insertion	**7p21.3**	**9,252,251**	**7p21.2**	**13,834,044**	**chr7:13,834,044-14,536,000**	**---**
		Excision/Insertion	**7p21.3**	**12,792,486**	**7p21.3**	**13,185,229**	**chr7:12,792,486-12,969,278**	**---**
		Excision/Insertion	**4q13.3**	**73,920,144**	**7p21.3**	**9,353,608**	**chr4:73,918,923-73,920,144**	**---**
		Excision/Insertion	**4q13.3**	**74,048,971**	**7p21.3**	**9,251,571**	**chr4:74,048,971-74,049,528**	**---**
		Deletion	**---**	**---**	**---**	**---**	**chr7:12,792,486-12,969,278**	**176,793**
	C	**Deletion**	**---**	**---**	**---**	**---**	**chr7:13,834,044-14,536,000**	**701,957**
	D	Deletion	2p16.3	50,843,321	2p16.3	51,318,853	chr2:50,843,321-51,318,853	475,532
		Deletion	3p21.1	46,752,812	3p21.1	46,810,087	chr3:46,752,812-46,810,087	57,275
		Deletion	9p21.3	25,258,355	9p21.3	25,352,944	chr9:25,258,355-25,352,944	94,589
		Deletion	11p14.3	25,680,453	11p14.3	25,699,601	chr11:25,680,453-25,699,601	19,148

In bold are highlighted the structural alterations determined to be linked to the CCR highlighted in bold. **A** and **B** correspond to the breakpoints of der(7) and der(4) described on the masters dissertation. der(7) breakpoint is located at chr7:14,535,828 and der(4) breakpoint is located at chr4:73,918,945. These previously determined breakpoints correspond to two of the seven CCR breakpoints. **C** is the paternally inherited 742 kb deletion on 7p21.2 and **D** is the maternally inherited 490 kb deletion on 2p16.3 which have its breakpoint placed at chr2:50,843,383.

Amplification and Sanger sequencing of the junction fragments allowed identification and comparisons of breakpoints between the father and son at nucleotide resolution. A schematic representation of the CCR is shown below.
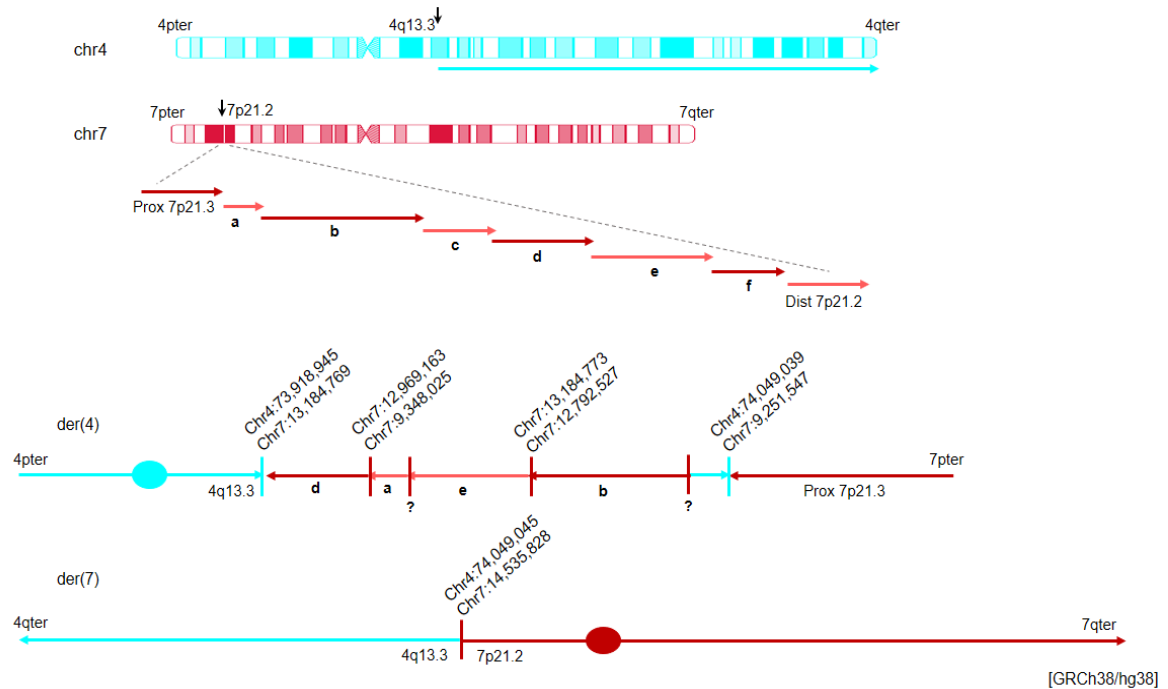

Figure.1 – Schematic representation of the "chromothripsis like" CCR with indication of the chromosomal breakpoints that were possible to determine with nucleotide resolution. Summarily, on der(4), the breakpoints between 4q13.3 and fragment **d** are positioned at chr4:73,918,945 and chr7:13,184,769. The breakpoints resulting from the junction of **d** and **a** fragments of 7p21.2, are situated at chr7:12,969,163 and chr7:9,348,025. The breakpoints amid fragments **e** and **b** are located at chr7:13,184,773 and chr7:12,792,527. The breakpoints resultant from the joint between a 4q13.3 fragment, inserted on the rearranged 7p21.2 region, and **Prox7p21.3** fragment are placed at chr4:74,049,039 and chr7:9,251,547. On der(7) the breakpoints of the translocated portion of 4q and 7p21.2 were determined to be at chr4:74,049,045 and chr7:14,535,828. [GRCh38/hg38].

All breakpoints determined with nucleotide resolution are identical between the index patient and his father, confirming the paternal inheritance of this "chromothripsis-like" CCR. None of the identified additional breakpoints were found to have direct implication on the blooming of the phenotypic features presented by the index patient.

The disrupted/deleted genes by the CCR include *DGKB, lncRNA AC01128.2* and *ETV1*, all in chromosome 7. The analysis of the CCR breakpoints allowed to note that the 130 kb on 4q13.3 breakpoint, described on the masters thesis, in fact does not exist. As so, there is no reduction of functional copies of *PF4*, *PPBP*, *CXCL5* and *CXCL3* genes.

DGKB, lncRNA AC011288.2 and *ETV1* were reported to be indirectly involved in neurodevelopmental alterations and autism spectrum disorder, due to implications in several signalling pathways that might alter the patterns of expression of genes reported to be directly related with the origination of the phenotypic traits presented by the index patient. Therefore, the reduction of the number of functional copies of these genes only appear to have indirect contribution on the blossoming of the phenotype.

The maternally inherited 473 kb deletion on 2p16.3 disrupts the gene *NRXN1*. Partial deletions of this gene have an incidence ratio of 1/2500 to 1/4000 in schizophrenic individuals and/or in individuals with global developmental delay and an incidence ratio of 1/5000 in individuals with no phenotypic features. Deletions on 2p16.3 are also related with neurodevelopmental alterations such as autism spectrum

disorder and cognitive alterations. Most of healthy adult individuals only discovered they were carriers of deletions on 2p16.3 due to their participations on studies, where were considered healthy controls, or when their offspring presented the phenotypic features linked to disruption of *NRXN1*, like seizures, autistic behaviour, learning difficulties and/or speech and language delay. As the result of the incomplete penetrance of variation on *NRXN1*, many cases of families were reported to have several members of the family, carrying the same deletion on 2.16.3, with highly variable phenotypes and different degrees of severity.

Considering the complexity of the phenotype presented by the index patient, 2 theories arise in an attempt to explain it. On the one hand, emerges the hypothesis that the deletion on 2p16.3 is the main cause of index patient's phenotype, since there were reported cases where the disruption of NRXN1 besides causing neurodevelopmental features also causes cardiopathies as well as non-verbal autism accompanied by hearing impairment. On the other hand, the second hypothesis is that in addition to the maternally inherited 473 kb deletion on 2p16.3, the paternally inherited CCR, contribute additively to the development of the complex phenotype presented by the index patient.

Both theories do not explain totally the complexity of the phenotype presented, so it is possible the existence of point mutations in genes reported to be directly related to the emergence of the phenotypic traits present by the index patient.

At the time, *NRXN1*, *DGKB, lncRNA AC011288.2* and *ETV1* are pointed as the candidate genes for the phenotype presented by the index patient, due to their implication in the development of phenotypic traits like autism spectrum disorders and global developmental delay. Preforming gene expression studies and mutation screening assays, is essential to clarify the molecular pathogenesis inherent to this case study.