UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL

Characterization of the molecular pathogenesis of a malformation syndrome associated with a complex double chromosome translocation

Mariana Sofia Carvalho Marques

Mestrado em Biologia Humana e Ambiente

Dissertação orientada por:
Doutor Dezső David
Professora Doutora Deodália Dias

2017

ACKNOWLEDGEMENTS

Antes de mais quero agradecer ao Instituto Nacional de Saúde Doutor Ricardo Jorge e, em especial, ao Departamento de Genética Humana por me terem acolhido durante o período de desenvolvimento desta Dissertação.

Agradeço ao Doutor Dezső David pela orientação e acompanhamento dados ao longo deste percurso, pela disponibilidade e pela oportunidade de aprendizagem e integração no Grupo de Investigação em Doenças Genómicas.

Agradeço igualmente à Professora Deodália Dias pelo apoio que me deu em todas as alturas, por estar sempre disponível e por me ter tirado as dúvidas mais importantes nas alturas mais inusitadas.

Gostaria também de agradecer a todos os parceiros da Harvard Medical School pela assistência prestada, especialmente à Doutora Cynthia Morton e ao Doutor Michael Talkowski.

Agradeço também ao Dr. João Freixo do serviço de Genética Médica do Hospital Dona Estefânia, sem quem não teria sido possível a revisão fenotípica do caso apresentado nesta dissertação, bem como ao Dr. Rui Gonçalves por ter indicado pela primeira vez este caso para estudo em projeto de investigação.

Um grande obrigado ao Doutor José Manuel Furtado por me ter ensinado técnica asséptica e mostrado a beleza da cultura de células.

Quero agradecer também aos meus (passados e presentes) colegas e amigos de trabalho. À Raquel pela passagem do conhecimento e pela amizade com que me acolheu, à Carina pela ajuda a perceber os meandros da secretaria e pelo constante carinho, à Patrícia que sempre foi uma querida, ao José Ferrão por se ter tornado o meu apoio nas alturas em que já não sabia para que lado me virar, à Mariana que torna os momentos mais tristes em sorrisos, à Joana pela ajuda com a minha falta de paciência para tecnologia e pela alegria e diversão que trouxe ao trabalho, à Inês que mesmo triste nunca dispensa um sorriso e à Neuza que se tornou uma grande amiga.

Mais, quero agradecer à Manuela por me ter acompanhado em todos os passos do caminho, nos piores e nos melhores dias.

Aos meus amigos, à Natacha por nunca ter desistido de me chatear mesmo quando eu passo dias sem dizer nada e ao meu grupo de licenciatura, a Raquel e o Renato, por fazerem questão de manter uma amizade à distância que é tão importante para mim.

Ao Zé que nunca me deixou durante o caminho, que me deu a mão mesmo quando eu queria fugir e que fez questão de me manter à tona dos nossos sonhos.

Agradeço muito à minha família, aos meus pais que nunca deixaram de me apoiar mesmo nos piores momentos e por me abanarem até eu voltar a mim. Obrigada Pai e Mãe por TUDO! À Susana que com a sua música conseguiu simultaneamente chatear-me e fazer-me rir que nem uma perdida. Aos meus Avós que nunca pararam de me dar carinho. Avó, eu sei que sabes e que quando tenho dúvidas só tenho de perguntar que tu explicas!

Um grande obrigado a todos os que passaram este tempo comigo!

MSCM, Fevereiro de 2017

ABSTRACT

Congenital anomalies are devastating conditions responsible for high neonatal mortality, as well as high morbidity of the surviving individuals. Chromosomal rearrangements are a leading cause of severe congenital malformations and are associated with about 25% of perinatal deaths due to congenital anomalies.

The aim of this study is the identification of candidate genes responsible for the phenotype characterized by intrauterine growth retardation, severe developmental delay, brain malformations and refractory epilepsy identified in an individual with an apparently balanced *de* novo double chromosomal translocation t(2;7)(q23;q32),t(5;6)(q23;q26)dn.

Identification and mapping of the structural chromosomal aberrations were performed by whole-genome array analysis, array painting with genomic amplicons of the derivative chromosomes and by whole genome sequencing of large-insert jumping libraries (liWGS). Subsequently all junction fragments were amplified and the breakpoints were identified at nucleotide resolution by Sanger sequencing.

Genome array analysis identified a 651.76 kb deletion at 14q24.3 (g.76,673,181-77,324,937 [GRCh37/hg19]). Transforming growth factor beta 3 (*TGFB3*), a gene associated with autosomal dominant arrythmogenic right ventricular dysplasia and Loeys-Dietz syndrome (OMIM #107970 and #615582), is situated 224 kb upstream from the proximal deletion breakpoint..

Translocation breakpoints were identified both by array painting and liWGS. The 2q23.3 breakpoint of the t(2;7)(q23.3;q32.1), disrupts IVS5 of pre-mRNA processing factor 40 homolog A (*PRPF40A*), a protein coding gene related to Huntington's disease (OMIM#143100). The calcium channel, voltage-dependent, beta-4 subunit (*CACNB4*) gene, localized 600 kb upstream of this breakpoint, is associated with three epilepsy related autosomal dominant disorders (OMIM #613855, 607682 and 607682). The Staphylococcal nuclease and tudor domain containing 1 (*SND1*) gene disrupted by the 7q32.1 breakpoint, is not presently associated with any known phenotype. However, the RNA binding motif protein 28 coding gene (*RBM28*), situated 300 kb downstream of the 7q32.1 breakpoint, has been associated with progressive neurological defects (OMIM #612079).

Concerning the t(5;6)(q23.2;q26) translocation, the 5q23.2 breakpoint is situated in an intergenic region whereas the 6q26 breakpoint disrupts IVS3 of PARK2 co-regulated gene (*PACRG*). This gene shares a bidirectional promoter with parkin RBR E3 ubiquitin protein ligase (*PARK2*), which is associated with early onset Parkinson disease. About 300kb downstream of this breakpoint is the homolog of quaking mouse (*QKI*) gene that also plays a role in brain development.

The application of liWGS unveiled the presence of two additional cryptic alterations on der(6), an excision/insertion and an inversion.

The cryptic excision at 6q22.33 disrupts protein tyrosine phosphatase receptor type K (*PTPRK*), a gene from the protein tyrosine phosphatase family which is associated with tumor suppression. As a result of the excision/insertion, the excised 48 kb fragment containing *PTPRK* exon 7 and flanking intronic sequences is inserted 36 Mb further distal at 6q26. Located 70kb from the *PTPRK* gene, the laminin 2 (*LAMA2)* gene was reported has being involved in brain malformations, including polymicrogyria. The inversion breakpoint at 5q23.2 is located within an intergenic region.

In conclusion, these findings suggest that disruption of *PRPF40A* and *PACRG* genes, in association with misregulation of *CACNB4*, *RBM28, PARK2*, *QKI* and *LAMA2* genes from the breakpoint regions are the most likely candidate genes responsible for this complex malformation phenotype. Additionally the modulating effect of *TGFB3* gene cannot be excluded.

Comparative analysis of this complex chromosome rearrangement by array painting and liWGS demonstrates that currently only liWGS is able to identify the full spectrum of balanced, otherwise cryptic, structural alterations. In this way, liWGS allows high-throughput delineation of chromosomal rearrangements, allowing a better phenotype-genotype association.

A major drawback of studying chromosome anomalies is the unavailability of relevant human biological material or of data from such samples. Theoretically, to overcome this issue, animal or induced pluripotent stem cells models can be used.

During this study, the obtainment of a proband-specific iPSC model was attempted. Unfortunately, the complexity of the pluripotency induction process, the associated costs and the requisites of using non-viral vectors hinder the development of such cellular models for the study of the molecular pathogenesis of congenital anomalies.

Proband derived lymphoblastoid cell line (LCL), non-integrative episomal plasmids containing the four Yamanaka factors – *OCT3/4*, *c-MYC*, *SOX2* and *KLF4* and an electroporation platform were used for the pluripotency induction experiments. Electroporated cells were maintained on a human foreskin fibroblasts (HFF) feeder-layer.

While performing the reprogramming experiments, several technical difficulties were identified. A major difficulty is achieving high transfection efficiency of LCL with episomal plasmids without high cell mortality rates. Although no LCL derived iPSC colonies were obtained, the identification of the critical steps in the induction protocol of LCL derived cells will certainly contribute for further development of such cellular models.

Furthermore, the availability of individual-derived iPSCs will definitely lead to a robust cellular model for the study of the molecular pathogenesis of chromosome rearrangements associated with congenital anomalies.

Keywords: Congenital anomalies; Complex chromosome rearrangement; large-insert Whole Genome Sequencing; induced Pluripotent Stem Cells.

This study was supported by FCT project HMSP-ICT/0016/2013.

RESUMO

As anomalias congénitas constituem uma das principais causas da mortalidade fetal, neonatal e infantil na Europa. Adicionalmente, devido à gravidade dos fenótipos apresentados pelos indivíduos que sobrevivem, este tipo de anomalias contribui em grande medida para a morbilidade, sobrecarregando em muito o sistema público de saúde.

Os rearranjos cromossómicos constituem uma das principais causas no desenvolvimento de malformações congénitas graves e estão associados a cerca de 25% das mortes devidas a anomalias congénitas no período perinatal. Por seu lado, os rearranjos cromossómicos estruturais, nomeadamente deleções, duplicações, inversões e translocações cromossómicas, têm vindo a ser recorrentemente associados a diversos fenótipos deletérios, nomeadamente síndromes malformativos caracterizados por atraso global do desenvolvimento psicomotor e anomalias cerebrais graves.

O principal objetivo deste estudo é a identificação de genes candidatos responsáveis pelo fenótipo caracterizado por atraso do crescimento intrauterino, atraso grave do desenvolvimento psicomotor, malformações cerebrais e epilepsia refratária identificado num indivíduo portador de uma translocação cromossómica dupla *de novo*, aparentemente equilibrada - t(2;7)(q23;q32),t(5;6)(q23;q26)dn.

A identificação e mapeamento das alterações cromossómicas estruturais foi realizada através da utilização de *array* genómico, de *array painting* com amplicões dos cromossomas derivados e ainda por sequenciação pangenómica de grandes insertos (do inglês *large-insert Whole Genome Sequencing*). Seguidamente, todos os fragmentos de junção das diversas alterações estruturais foram amplificados e os respetivos pontos de quebra foram identificados com resolução nucleotídica por sequenciação de Sanger.

Através da análise do *array* genómico foi possível a identificação de uma deleção de 651.76 kb na banda 24.3 do cromossoma 14, nas posições g.76,673,181- 77,324,937 (Genoma de Referência [GRCh37/hg19]). A montante do ponto de quebra proximal da deleção, a uma distância de 224 kb, o gene codificante para o fator de transformação do crescimento beta 3 (*TGFB3*) encontra-se associado a duas doenças autossómicas dominantes, displasia arritmogénica do ventrículo direito e síndrome de Loeys-Dietz (OMIM#107970 e #615582).

Os pontos de quebra da translocação dupla foram identificados por *array painting* e por sequenciação pangenómica de grandes insertos. O ponto de quebra do derivado do cromossoma 2 da t(2;7)(q23.3;q32.1) interrompe o homólogo para o fator 40 de processamento de pré-mRNA (*PRPF40A*), um gene codificante para uma proteína associada com a doença de Huntington (OMIM#143100). O gene codificante para a subunidade beta do canal de cálcio dependente de voltagem (*CACNB4*), localizado 600 kb a montante deste ponto de quebra está associado com três

condições autossómicas dominantes que envolvem diversos variantes de epilepsia (OMIM #613855, #607682 e #607682). O gene codificante para a nuclease estafilocócica e domínio *tudor* 1 (*SND1*), que se encontra interrompido pelo ponto de quebra do derivado do cromossoma 7, não está, presentemente, associado a qualquer fenótipo conhecido. Por outro lado, o gene codificante para a proteína 28 de ligação de RNA (*RBM28*), localizado 300 kb a jusante do ponto de quebra deste derivado, tem vindo a ser associado a defeitos neurológicos progressivos (OMIM #612079).

Relativamente à translocação t(5;6)(q23.2;q26), o ponto de quebra do derivado do cromossoma 5 está localizado numa região intergénica, enquanto que o ponto de quebra da translocação no derivado do cromossoma 6 interrompe o terceiro intrão do gene co-regulador da Parkina (*PACRG*). Este gene partilha um promotor bidirecional com o gene codificante para a proteína ubiquitina ligase da Parkina RBR E3 (*PARK2*), o qual se encontra associado com o aparecimento precoce da doença de Parkinson. A jusante deste ponto de quebra, a cerca de 300 kb, encontra-se o gene homólogo do murganho *quaking* (*QKI*), o qual apresenta igualmente um papel no desenvolvimento cerebral.

A aplicação da sequenciação pangenómica de grandes insertos revelou a presença de duas novas alterações crípticas no derivado do cromossoma 6, uma excisão/inserção e uma inversão.

A excisão críptica na banda q22.33 no derivado do cromossoma 6 interrompe o gene codificante para o recetor da proteína tirosina fosfatase tipo K (*PTPRK*). Este gene é um membro da família das proteínas fosfatases de tirosina que se encontram maioritariamente associadas à supressão de tumores. Como resultado da excisão/inserção o fragmento excisado, que tem uma extensão de 48 kb, contém o exão 7 do gene *PTPRK* e flanqueia sequências intrónicas, é inserido a montante no ponto de quebra da banda 6q26, a uma distância de 36 Mb. Localizado 70 kb a jusante do gene *PTPRK*, o gene codificante para a laminina 2 (*LAMA2*) foi reportado como estando envolvido em malformações cerebrais, incluindo polimicrogiria. Por outro lado, o ponto de quebra da inversão no 5q23.2 está localizado numa região intergénica.

Em suma, os dados encontrados sugerem que a interrupção dos genes *PRPF40A* e *PACRG*, em associação com a desregulação dos genes *CACNB4*, *RBM28*, *PARK2*, *QKI* e *LAMA2* que flanqueiam as regiões dos pontos de quebra, serão os genes candidatos mais prováveis para a explicação do fenótipo reportado de malformação complexa. Adicionalmente, o efeito modulador do gene *TGFB3*, que se encontra a montante do ponto de quebra proximal da deleção no cromossoma 14, não poderá para já ser excluído.

A análise comparativa deste rearranjo cromossómico complexo por *array painting* e por sequenciação pangenómica de grandes insertos, permitiu demonstrar que, neste momento, apenas a última abordagem tem a capacidade de identificar o espectro completo dos rearranjos cromossómicos estruturais aparentemente equilibrados, que de outra forma continuariam desconhecidos. Desta forma,

a sequenciação pangenómica de grandes insertos permite a delineação dos rearranjos cromossómicos estruturais com uma elevada resolução, permitindo uma associação mais fiável entre o genótipo e o fenótipo reportados.

Um dos principais obstáculos no estudo de anomalias cromossómicas é a indisponibilidade de material biológico humano relevante, bem como de dados acerca desse mesmo material biológico. Teoricamente, de forma a ultrapassar este problema poderão se utilizados modelos animais ou modelos celulares, incluindo células pluripotentes induzidas.

Durante o decorrer deste estudo foram iniciadas as experiências para a obtenção de um modelo de células pluripotentes induzidas específicas do indivíduo portador da translocação t(2;7)(q23;q32),t(5;6)(q23;q26)dn. Contudo, a complexidade do processo de indução de pluripotência, os custos associados e a necessidade da não utilização de vetores virais, dificultou grandemente o desenvolvimento deste modelo celular para o estudo da patogénese molecular das anomalias congénitas apresentadas.

De forma a realizar as experiências de indução de pluripotência foram utilizadas linhas linfoblastóides derivadas do indivíduo índex, plasmídeos episssomais não integrativos contendo os quatro fatores de Yamanaka - *OCT3*/4, *c-MYC*, *SOX2* e *KLF4* – e uma plataforma de eletroporação celular. As células eletroporadas e em processo de indução foram mantidas numa camada de suporte constituída por fibroblastos de prepúcio humano inativados por radiação gama.

Ao longo do desenvolvimento destas experiências foram encontradas diversas dificuldades técnicas. A principal dificuldade encontrada deteve-se na obtenção de uma elevada eficiência de transfecção da linha linfoblastóide com os plasmídeos episssomais sem obter, simultaneamente, taxas de mortalidade excessivamente elevadas. Apesar de não ter sido possível a obtenção de colónias de células pluripotentes derivadas da linha linfoblastóide, foram identificados os passos críticos do protocolo de indução, o que contribuirá certamente para o futuro desenvolvimento destes modelos celulares.

Adicionalmente, a disponibilidade de células pluripotentes induzidas específicas para cada indivíduo portador de um rearranjo cromossómico irá definitivamente conduzir a um modelo celular robusto para o estudo da patogénese molecular dos rearranjos cromossómicos associados com anomalias congénitas e síndromes malformativos. Da mesma forma, a possibilidade de diferenciar células pluripotentes induzidas específicas de um indivíduo índex em quaisquer células dos três folhetos embrionários é, sem dúvida, uma vantagem no estudo dos rearranjos cromossómicos.

Palavras-Chave: Anomalias congénitas; Rearranjos cromossómicos complexos; sequenciação pangenómica de grandes insertos; células pluripotentes induzidas

Este estudo foi financiado pelo projeto FCT HMSP-ICT/0016/2013.

TABLE OF CONTENTS

TABLES LIST

FIGURES LIST

ABBREVIATIONS

ARVD1	arrhythmogenic right ventricular dysplasia, familial, 1
BACs	bacterial artificial chromosomes
bp	base pairs
chr	chromosome
CCR	complex chromosomal rearrangement
CGH	comparative genomic hybridization
CN	copy number
DGAP	Developmental Genome Anatomy Project
DNA	deoxyribonucleic acid
der	derivative chromosome
DSBs	double strand breaks
EDTA	ethylenediamine tetraacetic acid
FISH	fluorescence *in situ* hybridization
FBS	fetal bovine serum
GRCh37	Genome Reference Consortium Human genome build 37
Gy	Gray
HFF	human foreskin fibroblasts
IMDM	Iscove's Modified Dulbecco's Medium
iPSC	induced pluripotent stem cells
kb	kilobases
LB	Luria-Bertani
LDS5	Loeys-Dietz syndrome 5
liWGS	Large-insert Whole Genome Sequencing
Mb	megabases
MDC1A	muscular dystrophy, congenital merosin-deficient, 1A
NGS	Next-Generation Sequencing
OMIM	Online Mendelian Inheritance in Man
PACs	P1-derived artificial chromosomes
PCR	Polymerase Chain Reaction
PFGE	Pulsed-Field Gel Electrophoresis
RPMI	Roswell Park Memorial Institute
RNA	ribonucleic acid
TBE	Tris-borate Ethylenediamine tetraacetic acid
T_m	melting temperature
YACs	yeast artificial chromosomes

1. INTRODUCTION

1.1. CHROMOSOME STRUCTURE

The word chromosome comes from the greek *khrôma*, meaning colored, and *sôma*, meaning body (Gardner et al. 2012).

Chromosomes were first observed by light microscopy in the late 19th century in dividing eukaryotic cells. Later, it was found that these structures corresponded to the most condensed form of chromatin, a fiber that consists in equal parts of deoxyribonucleic acid (DNA) and proteins, mostly histones (Alberts et al. 2008).

In the 1950's, with the discovery of DNA's double helix polymer structure by Franklin, Watson and Crick (Franklin and Gosling 1953; Watson and Crick 1953), it was found that DNA is not a structural molecule, as initially thought, but in fact it contains hereditary information. This double helix polymer consists in two antiparallel strands of complementary nucleotides - adenine and thymine, cytosine and guanine – which are organized to form genes that code the working proteins in the organism, and are in that way responsible for our hereditary information (Nelson and Cox 2008).

Most DNA is found in the cell nucleus. During interphase, it is possible to observe two conformations of the chromatin fiber, euchromatin which is more distended and is usually under active transcription processes, and heterochromatin which is highly condensed and is mainly responsible for gene regulation. During mitosis, each DNA molecule reaches its most condensed form, showing a two arms structure divided by a centromere with both arms terminating in telomeres (Nelson and Cox 2008)

Chromosome nomenclature is based on mitotic chromosomes. The short arm of the chromosome is called *p*, from the French word *petit*, and the long arm is called *q*. With the application of G-banding, it is also possible to identify several bands delimiting regions in each chromosome (Nelson and Cox 2008). Morphologically, chromosomes may be divided into metacentric, submetacentric, acrocentric or telocentric, according to where the centromere divides the *p* and *q* arms.

1.2. STRUCTURAL CHROMOSOMAL REARRANGEMENTS

Structural chromosomal rearrangements are defined as changes in the chromosomal structure that require chromosome breakage followed by abnormal end-joining (Griffiths et al. 2007). Structural chromosomal rearrangements may be balanced or unbalanced. Balanced rearrangements refer to alterations in the chromosomal gene order and may appear as inversions or translocations. On the

other hand, unbalanced rearrangements refer to alterations that involve gain or loss of genetic material in the form of duplications and deletions, respectively (Luthardt and Keitges 2001; Griffiths et al. 2007).

The first evidence of a chromosome inversion was published in 1921 by Alfred Sturtevant. This chromosome anomaly occurs when there are two double strand breaks (DSBs) in the same chromosome and the broken fragment rejoins the same chromosome in the reverse orientation (Kirkpatrick 2010). Inversions may be pericentric, in which the inverted region includes the centromere, or paracentric, in which the centromere is not involved in the anomalous region (Griffiths et al. 2007; Kirkpatrick 2010).

A chromosome translocation is defined as a chromosomal anomaly in which a portion of a chromosome breaks and reattaches to a different chromosome (Agarwal et al. 2006; Griffiths et al. 2007; Potter 2008; Roukos and Misteli 2014). This anomaly may originate direct gene disruption or disruption of regulatory elements, namely *cis* regulatory elements which are key in gene expression control. Besides this, when the previously broken sequence is rejoined a fusion gene may be formed, and its expression often leads to deleterious phenotypes (Kleinjan and Coutinho 2009; Roukos and Misteli 2014).

Deletions consist in loss of material in a single chromosome. There may be interstitial deletions, where there are two DSBs in the same chromosome and a portion of a chromosomal arm is lost, or terminal deletions, when a single break occurs near the telomere and the terminal portion of the chromosome is lost (Luthardt and Keitges 2001). Occasionally, when two terminal deletions occur the resulting non-corrected extremities fuse forming a ring chromosome, which is often highly unstable during cell division (Luthardt and Keitges 2001; Bershteyn et al. 2014).

Chromosome duplications usually occur by unequal crossing-over between homologous chromosomes (Luthardt and Keitges 2001). These anomalies may affect phenotype by altering gene dosage and, consequently, altering gene expression (Clancy et al. 2016).

In 1980, a chromosomal rearrangement involving more than two breakpoints in two or more chromosomes was defined as being a complex chromosomal rearrangement (CCR) (Pai et al. 1980). However, with the development and worldwide spreading of new techniques that allowed for a more detailed vision over chromosomal structural rearrangements, the definition of CCR has been revised. Nowadays, a structural rearrangement is considered complex not only when there are more than two breakpoints in two or more chromosomes, but also when it is possible to find more than three breakpoints independently of the number of chromosomes involved (Houge et al. 2003). In this way, CCRs may present combinations of translocations, insertions and transpositions (Pellestor et al. 2011; Liao et al. 2014).

1.3. CONGENITAL MALFORMATIONS ASSOCIATED WITH CHROMOSOMAL REARRANGEMENTS

According to the World Health Organization, congenital malformations are defined as structural or functional anomalies that occur during the intrauterine period and can be identified prenatally, at birth or during infancy (WHO 2016).

Congenital malformations are a leading cause of fetal and infant mortality in Europe (Dolk et al. 2010). Between 2003 and 2007, the European Surveillance of Congenital Anomalies Network (EUROCAT) recorded a total prevalence of major congenital anomalies of 23.9 in every 1,000 births. Of those, 2.5% died in the first week after birth (Dolk et al. 2010). Furthermore, congenital anomalies were responsible for 2.0% of fetal deaths and stillbirths (Dolk et al. 2010).

For the surviving individuals, congenital anomalies usually have severe health consequences, highly contributing to long term morbidity and, consequently, to a major burden to the European public health system (Dolk et al. 2010; Corsello and Giuffrè 2012).

Congenital anomalies may arise both from environmental or genetic factors (Corsello and Giuffrè 2012). Chromosomal rearrangements are responsible for 15% of all major congenital anomalies diagnosed before one year of age and are associated with about 25% of perinatal deaths due to congenital anomalies (Wellesley et al. 2012).

Chromosomal structural rearrangements, including CCRs, have been associated with multiple congenital anomalies, including malformation syndromes and global developmental delay (Houge et al. 2003; Kloosterman and Hochstenbach 2014). Usually, there is a correlation between the number of breakpoints and the severity of the phenotype (Houge et al. 2003). However, the genomic localization of the chromosomal breakpoints accounts for most of the phenotypical consequences (Kloosterman and Hochstenbach 2014).

Between 2011 and 2014, a copy number variation morbidity map of developmental delay was developed by analyzing the karyotypes and phenotypes from 15,767 young individuals. The individuals presented a number of phenotypes associated with developmental delay, as for instance congenital malformation, hypotonia and feeding difficulties, speech and motor deficits, growth retardation, cardiovascular and renal defects, epilepsy, hearing impairment, craniofacial and skeletal abnormal features and behavioral issues. This analyses made it possible to further annotate candidate genes affected by unbalanced chromosomal rearrangements, including Parkin 2 (*PARK2*) gene, which was found to contain a fragile site prone to deletions and overall breaks (Cooper et al. 2011; Coe et al. 2014).

Balanced genomic alterations have also been reported as being associated with developmental delay. Several authors have described numerous translocations and inversions associated with

developmental delay, intellectual disability and congenital anomalies (Higgins et al. 2008; Talkowski et al. 2012; Schluth-Bolard et al. 2013; Utami et al. 2014).

The Developmental Genome Anatomy Project (DGAP) has described several apparently balanced translocations and inversions associated with both developmental delay and congenital malformations. In fact, of the 40 chromosomal rearrangements studied, at least 25 presented developmental delay or mental retardation (Higgins et al. 2008).

Besides developmental delay, brain abnormalities associated with neurological disorders have been described in individuals presenting either balanced or unbalanced chromosomal rearrangements.

Several balanced translocations have been reported has being associated with brain malformations. In 2007, 27 balanced translocations were described with association to cognitive disorders and brain malformations, including encephalopathies leading to epileptic seizures (De Gregori et al. 2007). Furthermore, DGAP described 6 balanced translocations associated with both cortical malformations, including microcephaly, and epileptic seizures (Higgins et al. 2008).

In 2012, a 6q26 terminal deletion, which disrupted the Parkin Co-regulated (*PACRG*) gene, was found to be associated with polymicrogyria, a cortical malformation characterized by the presence of small and partially fused gyri at the surface of the brain. The individual presenting this deletion also presented a phenotype of microcephaly and progressive medication refractory epilepsy that lead to non-reversible left hemiparesis (Quelin et al. 2012).

1.4. METHODOLOGY EVOLUTION ON THE STUDY OF CHROMOSOMAL STRUCTURAL ANOMALIES

1.4.1. Classic Cytogenetic

After the discovery, in 1956, that the correct number of chromosomes in humans is 46 (Tjio and Levan 1956), several techniques of fixation and coloration were developed to analyze these newly found structures. The fixation of metaphase chromosomes from cultured peripheral blood leukocytes and skin fibroblasts exposed to colchicine was the first technique that allowed the observation of chromosomal anomalies (Moorhead et al. 1960). In fact, this technique allowed for the detection of the chromosomal numerical anomalies present in Down syndrome (Lejeune et al. 1959), Turner syndrome (Ford 1958) and Klinefelter syndrome (Jacobs and Strong 1959).

In 1971, G-banding was developed. This technique, based on the application of trypsin and Giemsa staining, allowed for a higher resolution chromosome staining when compared to the previous used Q-banding technique (Seabright 1971). In fact, G-banding allows for the observation of structural chromosomal rearrangements with a size higher than 5Mb (Riegel 2014). In 1973, through the use of

this methodology, it was possible to identify the first chromosomal translocation associated with a deleterious phenotype. It was t(9;22), the translocation responsible for the Philadelphia chromosome associated with chronic myeloid leukemia (Rowley 1973).

In 1976, the protocol for synchronization of lymphocyte cultures was developed and allowed for the obtainment of a significantly high number cell the metaphase stage for observation with G-banding (Yunis 1976). With this technique it was possible to visualize chromosomal deletions, namely the deletion in chromosome 5 associated with Cry-du-Chat syndrome (Yunis 1976).

1.4.2. Fluorescence *in situ* hybridization

Even though G-banding represented an enormous advancement in the observation of chromosomal structural anomalies, the resolution of this technique was tremendously limited.

In 1982, a new method for detecting specific sequences in chromosomes using fluorescence biotinylated polynucleotides was developed (Langer-Safer et al. 1982). This methodology, later called fluorescence *in situ* hybridization (FISH) became one of the reference techniques in detection of chromosomal anomalies.

Throughout the years, many FISH probes were developed for specific DNA sequences, making it possible to paint whole chromosomes, chromosome arms, centromeric and subtelomeric regions and it is even possible to design locus-specific probes (Durmaz et al. 2015). Furthermore, with the Human Genome Project it was possible to design more and more probes targeting specific sequences. In fact, the use of cosmids, bacterial artificial chromosomes (BACs), P1-derived artificial chromosomes (PACs) and yeast artificial chromosomes (YACs) were made available for diagnostic purposes due to the Human Genome Project (Cheung et al. 2001; Riegel 2014; Durmaz et al. 2015). The use of artificial chromosomes has since became the most used technique for mapping chromosome breakpoints and to visualize chromosome imbalances. However, the highest resolution achieved by FISH was only of 2.3kb (Vorsanova et al. 2010).

FISH increased the resolution of the identification of chromosomal structural rearrangements to a sub-microscopic level. However, the complexity of this technique and the fact that it is highly expensive and time consuming, especially when using BACs or YACs to evaluate chromosomal rearrangements in the whole genome led to the development of new techniques, namely array-based Comparative Genomic Hybridization (CGH) (Vorsanova et al. 2010; Durmaz et al. 2015).

1.4.3. Array-based Comparative Genomic Hybridization and Array painting

Chromosomal CGH was first developed in 1992 in order to detect genomic imbalances on solid tumor cells (Kallioniemi et al. 1992). This technique is based on quantitative two-color FISH. Total genomic DNA obtained from both test samples and controls is differentially labelled with two fluorescent dyes,

co-precipitated and co-hybridized onto normal metaphase chromosomes, which appear differentially marked according to the gain or loss of DNA content (Kallioniemi et al. 1992; Riegel 2014).

Array-based CGH has brought the advantage of not being necessary to use metaphase chromosomes to hybridize the genomic DNA of interest. In fact, DNA is hybridized to reference DNA sequences adherent to a glass slide (Solinas-Toldo et al. 1997; Pinkel et al. 1998). This technique allows for the detection of whole genome copy number variation - duplications and deletions. The resolution of array-based CGH has been improved overtime. Shorter sequences are being used as targets and now it is even possible to perform single nucleotide polymorphism arrays that allow for a resolution of 5-10kb (Scouarnec and Gribble 2011).

Even though regular array-based CGH does not allow for the detection of balanced genomic alterations, new protocols have been developed in order to solve the issue.

In 2003, a new methodology termed array painting was described. Even though this technique is not sensitive to unbalanced genomic alterations, it allows for high resolution analysis of breakpoints in balanced chromosomal rearrangements. The methodology of array painting involves the separation, by flow sorting, of the derivative (der) chromosomes from a balanced translocation. DNA of the sorted chromosomes is then amplified using degenerate oligonucleotide-primed PCR, differentially labelled and hybridized onto the array. Since der chromosomes were sorted and amplified, the fluorescence ratio appears differentiated according to which der chromosomes hybridized to which clone. When the array reads the breakpoint region where both der chromosomes hybridize, the measured fluorescence ratio is intermediate making it possible to map the breakpoint region (Fiegler et al. 2003).

1.4.4. Next-Generation Sequencing

Next-generation sequencing (NGS) methodologies were first popularized 2005. Nowadays, these technologies allow for the sequencing of a whole genome with a much lower cost and much quicker than using the traditional Sanger sequencing (Metzker 2010).

Due to a previous library preparation of DNA fragments made in order to produce sequence reads, NGS allows for the reading of millions of DNA molecules simultaneously. These reads are aligned to the a reference genome and the results are analyzed using bioinformatic tools, making it possible to analyze a whole genome and detect insertions, deletions and structural variants with a much higher resolution than it would be possible with the previously described techniques (Metzker 2010).

In 2008, the Illumina/Solexa platform, based on massive parallel sequencing, allowed for the mapping of chromosomal breakpoint regions for the first time. After chromosome flow sorting and amplification, the amplified der chromosomes were randomly fragmented and subjected to 27 to 36

sequencing cycles, making it possible to map the breakpoint regions with an accuracy of less than 1kb (Chen et al. 2008).

Between 2007 and 2011, genomic DNA started to be used to produce libraries, allowing to skip both the flow sorting and PCR amplification steps. This allowed for breakpoint mapping using whole genome paired-end sequencing. Breakpoint mapping was possible by identifying paired-end reads for which the respective pair (mate pair) aligned to a different region on the reference genome (Korbel et al. 2007; Chen et al. 2010; Talkowski et al. 2011).

This read-pair method was popularized due to its ability to quickly indicate the possibility of a chromosomal structural rearrangement by clustering discordant pairs. It is expected that the aligned read pairs map at a certain distance and with a determined orientation. When two read pairs map with a significantly longer distance it is indicative of a possible deletion. On the other hand, when two read pairs map in different chromosomes it is indicative of a translocation breakpoint (Scouarnec and Gribble 2011).

Between 2011 and 2014, a new method for whole genome sequencing using large-insert jumping libraries was developed. This methodology, based on the previously used mate pair protocol from SOLiD 4 System (Applied Biosystems) which includes cap adaptors for circularization of the DNA of interest and a posterior step of enzyme restriction, was adapted for 25 cycle Illumina sequencing (Talkowski et al. 2011; Hanscom and Talkowski 2014).

The resulting libraries consist on short DNA fragments that represent junctions of much larger circularized genomic fragments. The use of such short reads allow for effective genomic coverage, with only 2,74% of non-covered genomic regions, and also minimizes the cost of performing Whole Genome Sequencing (Talkowski et al. 2011; Hanscom and Talkowski 2014).

This protocol has since been used to identify several structural chromosomal rearrangements, identifying cryptic rearrangements that would not have been otherwise unraveled. The newly described chromothripsis phenomena, in which a chromosome becomes completely shattered and is abnormally rejoined, is one of the examples of a set of cryptic rearrangements that could only be found using this high resolution technology (Kloosterman et al. 2011; Talkowski et al. 2011). Furthermore, this methodology has been applied to characterize structural rearrangements in pre-natal diagnosis with a much higher resolution than currently available diagnosis techniques (Ordulu et al. 2016).

1.5. MODELS FOR THE STUDY OF STRUCTURAL CHROMOSOMAL ANOMALIES PHENOTYPIC CONSEQUENCES

As referred previously, structural chromosomal anomalies most often lead to deleterious phenotypes. However, the relationship between breakpoint regions and the individual's phenotype is usually not obvious and most of the time it is not feasible to obtain samples from the affected tissues, nor is it possible to follow the affected individuals throughout an extensive amount of time in order to observe the evolution of the phenotype.

In 2003, a 10 year study of an individual with a CCR was published. During this 10 year period, by accompanying the development of structural chromosomal rearrangements analyses techniques, it was possible to continuously improve the defined karyotype and, in that way, better understand the presented phenotype (Houge et al. 2003). However, prolonged studies are most often not feasible.

In order to overcome this difficulty, animal and cellular models have been developed in order to understand the genotype-phenotype association when it comes to structural chromosomal rearrangements.

When it comes to animal models, mice are most often used to model chromosomal rearrangements, especially because of their biological and genetic similarity to humans (Weyden and Bradley 2008). Indeed, several mouse models were created to mimic the phenotypic consequences of a number structural chromosomal disorders (Brault et al. 2006; Weyden and Bradley 2008). However, creating a functional mouse model requires specific working conditions and it is proven to be a long and extremely expensive process.

The second option on modelling structural chromosomal rearrangements is the use of cellular models. However, as in mice modelling, inducing the exact DNA double strand breaks that happen in a structural chromosomal rearrangement, namely in a translocation or an inversion, has been proven to be incredibly difficult. The new CRISPR/Cas9 methodology opened doors on modelling chromosomal DNA. In fact, this technology already allowed for the modelling of a murine embryonic stem cell line that carries a t(5;7) balanced translocation (Jiang et al. 2016).

Induced pluripotent stem cells (iPSCs) provide another possible model for the study of structural chromosomal rearrangements. As it is described below, inducing pluripotency allows for the possibility of having a unique line for each individual specific chromosomal rearrangement.

1.5.1. Induced Pluripotent Stem Cells

The concept of pluripotency is generally applied to a cell that is able to differentiate into cells that originate any of the three germ layers - endoderm, mesoderm and ectoderm. Yet, pluripotency

definition requires a cell with this property to be able to originate a whole new organism with all his constituent parts (Han et al. 2010; Robinton and Daley 2012).

This property started to be explored in 1960's (Robinton and Daley 2012) and the first publication regarding cell reprogramming was only made in 1986, when Lassar and his colleagues reported differentiation of the mouse fibroblast line 10T1/2 through the use of 5-azacitidine. These scientists found that if 10T1/2 line was transfected with DNA from 5-azacitidine induced myoblasts it was possible to observe myogenic conversion (Lassar et al. 1988).

About seven years before the first reprogramming experience being published, native mouse embryonic stem cells were isolated and cultured, beginning the era of pluripotent stem cell study (Robinton and Daley 2012). Later, in the 1990's, Thomson et al. were able to isolate human embryonic stem cells from human embryos remaining from in vitro fertilization procedures, and found that these cells were in fact able to differentiate in a number of tissue specific cells, confirming its pluripotent capacity.

Due to severe ethical issues on the fact that in order to isolate human embryonic stem cells it is necessary to use human embryonic tissue, scientists quickly searched for new ways to produce stem cells. Thus, in 2006, Takahashi and Yamanaka and Thompson and its group independently found that it was possible to reprogram mouse embryonic and adult fibroblasts by introducing four transcription factors through retroviral transduction (Takahashi and Yamanaka 2006; Han et al. 2010; Robinton and Daley 2012). One year later, in 2007, the same authors were able to reprogram human adult fibroblasts for the first time using the same four reprogramming factors and a set of lentiviral plasmids (Takahashi et al. 2007).

The factors used in Doctor Yamanaka's laboratory became known as the Yamanaka factors once reprogramming and inducing somatic cells became more and more popular in science. These four factors, octamer-binding transcription factor 4 (*OCT3/4*), SRY-BOX 2 (*SOX2*), proto-oncogene c-Myc (*c-MYC*) and kruppel like factor 4 (*KLF4*), were able to reprogram completely both mouse adult fibroblasts (Takahashi and Yamanaka 2006) and human adult fibroblasts (Takahashi et al. 2007), introducing the concept of induced pluripotent stem cells (iPSC). Also, Takahashi found that the induced cells expressed high levels of telomerase activity and had the ability to form subcutaneous teratomas when injected into immunodeficient mice. Additionally, the group was able to differentiate these cells into cardiac and neural cells, confirming once again its pluripotent ability (Takahashi et al. 2007).

Since Takahashi first induced pluripotency in adult human fibroblasts with lentiviral vectors, many other starting cells were used and other methodologies were developed.

iPSCs have already been developed from cord blood cells, namely CD34+ cells (Ramos-Mejía et al. 2012; Okita et al. 2013), peripheral blood cells , of which it has been mostly used PBMCs (Merling et al. 2016) and CD34+ cells (Mack et al. 2011; Meng et al. 2013), dental pulp cells (Lizier et al. 2013) and even squamous cells present in urine (Zhou et al. 2011; Wang et al. 2013).

1.5.1.1 Methodologies and Applications of iPSC

When it comes to methodologies, iPSC have been developed through the use of integrative vectors, namely retrovirus and lentivirus, but the tendency has been to start using integration-free vectors. Retrovirus and lentivirus may leave a residual viral activity throughout the reprogramming process, interfering with the reprogramming potential of the cells. Also, especially when it comes to retroviral transfection, it is possible to obtain only partially induced cell lines that will always depend on the viral vector to continue the reprogramming process. Lentivirus have some advantage over retrovirus since they offer the possibility of transfecting polycistronic expression cassettes that offer higher efficiencies than retrovirus (Stadtfeld and Hochedlinger 2010). A very present concern when using integrative viral vectors is its association with insertional mutagenesis and the probability of tumor induction if these cells were actually applied to a patient (Zhang 2013).

Lately, the most used vectors for pluripotency induction have been non-integrative vectors, namely transfection with Sendai virus and with episomal plasmids (Stadtfeld and Hochedlinger 2010; Manzini et al. 2015). Sendai virus is a RNA based virus which shows no integration in its host genome. Of the non-integrative vectors available, Sendai virus has been proven to have the highest efficiency when it comes to induce pluripotency, at around 0,01% (Fusaki et al. 2009; Fujie. et al. 2014).

Episomal plasmid vectors are used in combination with electroporation or pore-making reagents (e.g. Lipofectamine) in order to be able to enter the cells. Episomal plasmids are also non-integrative vectors and do not represent, in the face of today's knowledge, any danger towards the receiver. Even though this vectors have been widely used towards de development of iPSC the reprogramming efficiency rarely surpassed 0,001% (Mack et al. 2011; Dowey et al. 2012; Zhang 2013; Okita et al. 2013; Barrett et al. 2014; Thomas et al. 2015).

The main objectives of iPSC development were both to broaden the knowledge of cell reprogramming mechanisms and to develop a way to have access to all types of individual specific cells in order to proceed to individual specific therapies in regenerative medicine (Yamanaka 2012).

At the moment, cells differentiated from iPSC were already used in the treatment of platelet diseases, spinal cord damages and in brain degenerative diseases such as Parkinson, Alzheimer and schizophrenia (Yamanaka 2012). Also, iPSC generated from $CD34^+$ cells and differentiated into

retinal cells are being studied as a way to restore vision to patients with retinitis pigmentosa (Zhu et al. 2016).

Nonetheless, the main application of human iPSC has been in the understanding of human pathogenesis (Han et al. 2010). In fact, iPSC derived from human fibroblasts and human blood cells are being differentiated into several cell types and used to understand and monitor disease progression, which will allow for a greater knowledge on pathogenesis development mechanisms (Han et al. 2010; Robinton and Daley 2012).

2. OBJECTIVES

The main objective of the present study is to identify the molecular alterations responsible for the malformation phenotype present in a subject with an apparently balanced *de novo* double chromosome translocation t(2;7)(q23;q32),t(5;6)(q23;q26)dn. In this way, it is intended to fulfil the following goals:

1. Enlightenment of the breakpoints localizations at a nucleotide level by amplifying the breakpoint regions by PCR and Sanger sequencing the amplified fragments;
2. Characterization of the breakpoint regions associated with the translocation regarding the disrupted genes and the genes flanking these regions;
3. Proposition of an association of the interrupted genes and the genes flanking the breakpoint regions with the phenotype presented by the subject.

In addition to the main goal, it is intended to assess the feasibility of establishing an iPSC based disease model for this congenital anomaly using subject's derived LCL. For this purpose it is expected to:

1. Establish ideal irradiation conditions for inactivating human foreskin fibroblasts;
2. Establish an electroporation methodology that will allow efficient transfection of the episomal vectors to the LCL.

3. MATERIAL AND METHODS

3.1. DETERMINING CHROMOSOME BREAKPOINTS

3.1.1. Sample collection

Blood samples from the proband were collected by venepuncture. Informed consent was obtained by signing the project's Declaration of Consent, previously approved by the Ethics Committee of National Institute of Health Dr. Ricardo Jorge. The present study was carried out according to the *Declaration of Helsinki* from the World Medical Association (Helsinki 1964; Tokyo 1975; Venice 1983; Hong Kong 1989; Somerset West 1996 e Edinburgh 2000).

3.1.2. Establishment and maintenance of lymphoblastoid cell lines

EBV-rich medium was obtained by culturing the B95-8 cell line in Roswell Park Memorial Institute (RPMI) 1640 medium (Gibco, New York, USA), supplemented with 10% fetal bovine serum (FBS) (Gibco, New York, USA), 1% L-glutamine (Gibco, New York, USA) and 1.5% Penicillin/Streptomycin (Gibco, New York, USA) and continuously increasing the cell culture density (Miller 1982; Neitzel 1986).

Cells were maintained in a 37ºC incubator with 5% CO2 and the medium was changed every other day for 12 days, until it turned lemon yellow overnight. The following week, medium was changed daily in order to accelerate cell culture growth. Cell culture density was adjusted to $1x10^7$ cells/mL in 15mL of medium and the culture was incubated for 12 days, until the medium was saturated with viral particles. At the end of the incubation period cell suspension was centrifuged and the viral supernatant was filtered twice through a 0,45μm syringe filter. EBV medium can be aliquoted and stored at 4ºC for at least one year without losing the ability of infecting and immortalizing peripheral blood mononuclear cells (PBMCs).

PBMCs were isolated from peripheral blood preserved in sodium heparin. Peripheral blood was diluted in equal parts with RPMI medium and carefully transferred to either a 12mL or 30mL LeucoSep centrifuge tube (Greiner Bio-One, Kremsmünster, Austria) pre-prepared with one volume of Ficoll-Paque solution (GE Healthcare Life Sciences, Washington, USA).

The tubes were centrifuged at 800xg and the cloudy white interphase containing PBMCs was transferred to a new centrifuge tube with a plastic Pasteur pipette. Two washes with phosphate-buffered saline (PBS) (Gibco, New York, USA) were made and PBMCs were counted with a hemocytometer. About $5x10^6$ cells are needed in order to proceed with the immortalization. The remaining cells were suspended in freezing medium (90% FBS + 10% DMSO) and frozen in a liquid nitrogen storage.

Isolated PBMCs were then exposed to EBV-rich medium and incubated at 37ºC for two hours. This incubation period should end when it is possible to observe small cell clusters. At the end of the incubation period, one volume of RPMI supplemented with 10% FBS, 1% L-glutamin, 1,5% Penicillin/Streptomycin and 1μg/μL of cyclosporine A (Novartis, Basel , Switzerland) was added to the culture.

Cell medium was changed 72 hours after starting the culture. For the following 21 days, 50% of the medium, without cyclosporine A, was changed once a week until the observed cell clusters became larger and presented the typical rosette morphology (Hussain and Mulherkar 2012) (Figure 3.1).

The established culture was then changed to a T25cm^2 flask and half of the medium was changed every other day. Once the flask was confluent the cells were divided at a reason of 1:2 or 1:3, depending on the proliferation rate. After freezing six aliquots of each culture, the cells were counted with a hemocytometer and cell density was maintained for at least seven days in order to proceed to nucleic acid extraction.

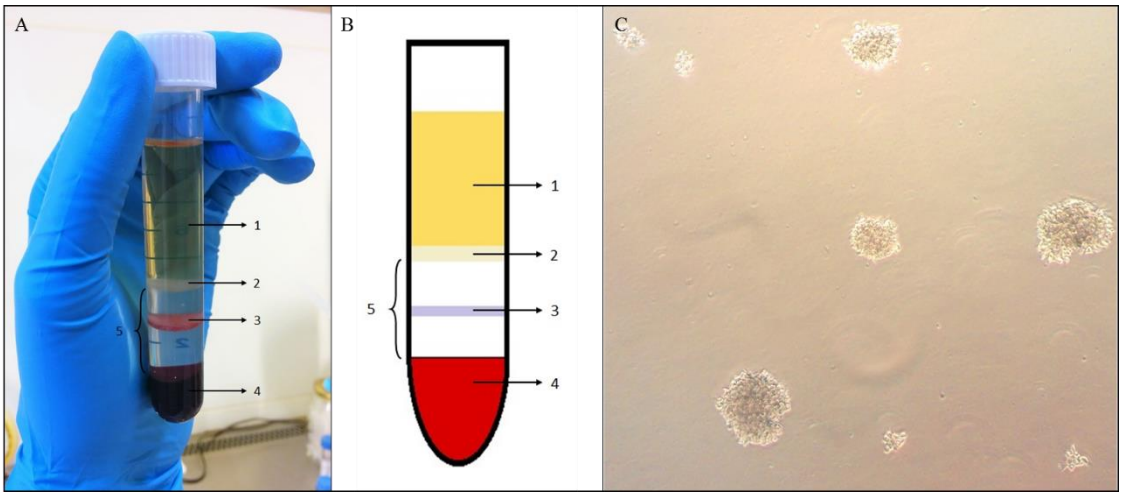

Figure 3.1: PBMCs isolation and mature LCL culture.
A Leucosep tube after separation by ficoll density gradient.
B Schematic representation of phase separation in PBMCs isolation. Legend of both A and B 1. Plasma; 2. Interphase containing PBMCs; 3. Leucosep tube membrane; 4. Erythrocytes; 5. Ficoll layer.
C Mature LCL culture showing the characteristic cell clusters in suspension.

3.1.3. DNA extraction from peripheral blood and LCL

DNA extraction from both peripheral blood preserved in EDTA and from LCL was made with the QIAamp DNA Blood Midi Kit (Qiagen, Hilden, Germany). This is a column based kit in which the membrane has high affinity to genomic DNA. The initial sample of either peripheral blood or LCL was treated with Protease K and a lysis buffer. DNA was precipitated with ethanol and transferred to the purification column. Then two washing buffers were used to purify the DNA. Elution was made using low Tris-EDTA buffer.

3.1.4. High Resolution Genomic Array

In order to screen for unbalanced genomic alterations, previously isolated genomic DNA from the proband was analysed by high-resolution genomic array, using the CytoScan HD array according to the manufacturer's instructions (https://tools.thermofisher.com/content/sfs/manuals/cytoscan_assay_user_manual.pdf, Affymetrix, California, USA). The data retrieved from the array was analysed using ChAS software from Affymetrix (California, USA).

3.1.5. Flow sorting of derivative chromosomes, Genomic amplification and Array painting

Flow sorting of derivative chromosomes, genomic amplification and array painting were performed as previously described (David et al. 2013).

To summarize, metaphase chromosomes of the proband derived LCL were flow sorted by fluorescence-activated cell sorting with a FACS Vantage (Becton–Dickinson, New Jersey, USA), according to standard protocols, by a service provider at CHROMBIOS Molecular Cytogenetics (Raubling, Germany). Genomic amplification was carried out using the REPLI-g Midi Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Purification of the genomic amplicons was made with the QIAamp Mini Spin Columns (Qiagen, Hilden, Germany) according to manufacturer's instruction.

Genomic amplicons of isolated derivative chromosomes were analysed by high-resolution oligonucleotide array painting using the Cytogenetics Whole-Genome 2.7 array from Affymetrix (Santa Clara, CA, USA) by a service provider at the Gulbenkian Institute of Sciences (Oeiras, Portugal), according to manufacturer's instructions. Data retrieved from the arrays was analysed with the Chromosome Analysis Suite (ChAS) software from Affymetrix (Santa Clara, CA, USA).

3.1.6. Large-insert Whole Genome sequencing

Large-insert Whole Genome sequencing was performed at the Centre for Human Genetic Research (Massachusetts General Hospital, Boston), under the cooperation protocol established by the Harvard Medical School-Portugal Program, according to the large-insert jumping libraries protocol (Talkowski et al. 2011).

Bioinformatic analysis of the sequencing raw data was also performed at the Centre for Human Genetic Research (Massachusetts General Hospital, Boston), using the pipeline described by Talkowski et al. (2011). An independent bioinformatic analysis was performed by the Genomic Diseases Research Group (National Institute of Health Doctor Ricardo Jorge, Lisbon), using a Python based pipeline adapted from the one described by Talkowski et al. (2011).

3.1.7. Amplification of junction fragments and Sanger sequencing

Refinement of the breakpoint regions determined by both array painting and liWGS was made by sequence-specific polymerase chain reaction (PCR) followed by Sanger sequencing.

Sequence-specific oligonucleotides were designed using either OLIGO software or NCBI Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Generally, oligonucleotides should have around 10 to 25bp in order to avoid both non-specific amplification and the formation of secondary structures. Primer self-complementarity and dimerization ability should be avoided in order to maximize the efficiency of the amplification reaction (Roux 2009). To maximize the efficiency of the reaction, primers were designed with a maximum of 25bp and RepeatMasker tool on UCSC Genome Browser (https://genome.ucsc.edu/; Kent et al. 2002) was used in order to avoid primer annealing on highly repetitive genomic regions flanking the breakpoints. Designed primer sequences and amplification conditions are shown in table 3.1.

Table 3.1: Sequence of the primers used in the amplification of the junction fragments of the double translocation, respective primers coordinates, annealing temperatures and size of the amplified fragment (JF – Junction Fragment; CF – Control Fragment) (Reference Human Genome Assembly 37/hg19).

Primer Designation	Sequence (5' → 3')	Primer Coordinates	Ann. Temp. (ºC)	Size (bp)	
				JF	CF
Junction Fragment der(2)					
2q23.3 translocation					
PRPF40A_IVS5-1F	ATCTAACATCTGTCGGGCACTG	2:153546821-153546842	62	340	405
SND1_IVS16-2R	AAATAGTGCTTGGGAGTTGGAA	7:127652671-127652692			
Junction Fragment der(7)					
7q32.1 translocation					
SND1_IVS16-1F	AATCTGGACTCTTGGTAGGTGG	7:127652241-127652262	62	476	452
PRPF40A_IVS5-2R	GAAGGGAAGCAGAACTTGTAGG	2:153547204-153547225			
Junction Fragment der(5)					
5q23.2 translocation					
AC109464.2-1F	ATGAGGGTTGGAAATGAAAATC	5:124174933-124174954	61	665	1,404
PACRG_IVS3-2R	TCACCAAAAGGAAGACTCAACA	6:163471742-163471764			
Junction Fragments der(6)					
6q22.33 translocation					
PTPRK_IVS6-1F	GGGAAAAACAGAGATGATGAAA	6:128510715-128510736	61	803	754
AC10235.7-1F	ACCTATCAAAATCCTGGCTGTC	5:125625845-125625866			
5q23.2 inversion					
AC109464.2-2R	AACTGATTTCCACAAGCCACAC	5:124176315-124176336	63	1,184	1,357
AC10235.7-2R	CGCAACTCGTCTCTAAGCATTT	5:125625966-125625988			
6q26 insertion					
PACRG_IVS3-3F	TTTTACCCTTTTCTCTGACCTCT	6:163471742-163471764	57	705	596
PTPRK_IVS7-5R	CTATACGGTGCTTCCAATGTTT	6:128463590-128463611			
6q22.33 excision					
PTPRK_IVS7-1F	CCTGTTGAGATTTGGAGTATGG	6:128462690-128462711	63	533	961
PTPRK_IVS6-2R	TATGATGAGTTGTTTGGCTTCC	6:128511447-128511469			

PCR conditions were optimized using 100ng of genomic DNA isolated from LCL. AmpliTaq DNA Polymerase with buffer I by Applied Biosystems was used. Gradient PCR, with temperatures ranging from 58ºC to 64ºC, was performed in order to find the optimal annealing temperature. When needed, Stratagene's Opti-Prime PCR Optimization Kit, which contains twelve buffers with varying pH and concentrations of $MgCl_2$ and KCl, was also used.

After PCR optimization, breakpoint regions were amplified using genomic DNA extracted from the proband's blood. The amplified fragments were purified using Amicon Ultra 0,5mL columns (Merck Millipore, Darmstadt, Germany). Low TE buffer was added to the PCR product to make a volume of 500μL and the solution was transferred to the purification column. The column was centrifuged at 14,000xg for seventeen minutes and the eluate was dismissed. Then, the column was filled with 300μL of low TE and was centrifuged at 14,000xg for twelve minutes. Finally, the column was transferred upside down to a new tube and centrifuged for two minutes at 1,000xg in order to collect the purified DNA.

The purified fragments were directly sequenced using the BigDye terminator cycle sequencing kit (Applied Biosystems, California, USA). The annealing temperature of the sequencing reaction was adjusted according to the optimal annealing temperature of the used primer. Reaction products were separated on the Applied Biosystems 377 PRISM automated sequencer, which is based on a four-color fluorescent labeling technology, according to manufacturer's instructions. Resulting sequences were aligned with the human genome assembly GRCh37/hg19, using reference sequences retrieved from UCSC Genome Browser (https://genome.ucsc.edu/; Kent 2002; Kent et al. 2002) .

Online Mendelian Inheritance in Man (OMIM) database (https://www.omim.org/; Hamosh et al. 2005) and GeneCards (http://www.genecards.org/; Stelzer et al. 2016) database were used as a foundation for retrieving information about the disrupted genes and the genes flanking the breakpoints.

3.2. PLURIPOTENCY INDUCTION

3.2.1. Plasmid isolation and characterization

E. coli containing episomal plasmids expressing pluripotency induction factors *OCT3/4*, *Sox2*, *SV40LT*, *KLF4*, *shRNA-P53*, *L-MYC* and *LIN28* (Addgene references 27082, 20927, 27077, 27080 and 27078) (Table 2) were cultured overnight in Luria-Bertani (LB) agar medium, at 37ºC. Three well isolated colonies of each culture were picked and inoculated in a 15mL centrifuge tube with 10mL of LB medium and cultured overnight at 37ºC with continuous agitation. 5mL of each of the starter cultures were inoculated in 50mL of LB medium and cultured overnight at 37ºC with continuous agitation.

Plasmid DNA was isolated from 25mL of bacterial culture using the PureLink HiPure Plasmid Filter Midiprep kit (Invitrogen, California, EUA), an endonuclease free kit able to purify up to 350µg of high copy number plasmid DNA from 25mL of bacterial culture. The purification columns of this plasmid DNA isolation kit combine the filtration of the bacterial lysate with anion-exchange resin purification, granting plasmids with a high purification grade.

Plasmid DNA was eluted in DNase RNase free water (Invitrogen, California, EUA) and stored at -20ºC.

Restriction enzyme digestion following agarose gel electrophoresis was made in order to confirm the extracted plasmids. Samples of the different plasmid DNAs were incubated overnight with 20U of the appropriate restriction enzyme (Table 3.2). After the overnight incubation, 10U of enzyme were added to the reaction and left to incubate for three hours. Non-digested and linearized plasmids were applied in an agarose gel electrophoresis.

Table 3.2: Plasmids used for pluripotency induction, respective sizes, reprogramming factors and restriction enzymes used in the characterization.

Plasmid	Size (bp)	Reprogramming Factor	Restriction Enzymes	
pCXLE-EGFP[1]	10,912	-	-	-
pEP4-E02S-ET2K[1]	17,522	*OCT4* *SOX2* *SV40LT* *KLF4*	-	-
pCXLE-hOCT3/4-shp53-F	11,681	*OCT3/4* shRNA against p53	EcoRV	SpeI
pCXLE-hUL	12,051	*L-MYC* *LiN28*	EcoRV	SpeI
pCXLE-hSK	12,693	*SOX2* *KLF4*	EcoRV	SpeI

A few months after plasmid DNA extraction and elution in DNase RNase free water (Invitrogen, California, EUA), the quality of the frozen plasmid DNA was evaluated by Pulsed-Field Gel Electrophoresis (PFGE) (Bio-Rad, California, USA). This technique consists on the alternate application of two separate electric fields which oblige the molecules to assume different orientations according to which field is active, allowing for a better molecule separation with a higher resolution (Reed et al. 2007).

[1] pCXLE-eGFP and pEP4-E02S-ET2K were previously extracted and characterized.

The conditions for plasmid DNA were based on the 5kb ladder separation proposed by the Instruction Manual and Applications Guide of the PFGE apparatus. The agarose gel was made at 1% (wt/vol) in 0,5x Tris/borate/EDTA buffer (TBE) and it was ran at 6V/cm with a 2 second switch time for 6 hours.

3.2.2. Feeder Layer Inactivation

The cell line CRL-2429 of Human Foreskin Fibroblasts (HFF) was flash thawed and cultured with Iscove's Modified Dulbecco's Medium (IMDM) (Gibco, New York, USA), completed with 10% FBS (Gibco, New York, USA), in a T75cm^2 flask (Sarstedt, Nümbrecht, Germany). Medium changes were performed twice a week until the cells reached about 80-90% confluence, at which point cells were detached with Trypsin-EDTA (Gibco, New York, USA) and subcultured to another flask.

This process was repeated until a sufficient amount of cells was obtained and then the cells were detached and resuspended in basal IMDM. The resuspended cells were inactivated with gamma (γ) radiation in the Gammacell 3000 Elan irradiator (Theratronics, Ontario, Canada). A process of radiation dosage optimization was made, irradiating the same cell culture from 25 to 65Gy, and it was found that the optimum irradiation was of 40,61Gy.

After the irradiation, inactivated HFF were frozen in liquid nitrogen with standard freezing medium. A small amount of cells was always plated in order to perform a quality control of the irradiation process. It was considered that the cells were successfully inactivated when after plating they were capable of only having one more mitotic division.

3.2.3. LCL Electroporation

Transfection of plasmid DNA was made through cell electroporation, in which electric impulses are given to the cells in order to open small pores on their membrane that will allow for plasmid DNA entrance.

The chosen platform for electroporation was the 4D-Nucleofector System (Lonza, Basel, Switzerland). This system can be used with either 20µL strips or 100µL cuvettes which allows for the optimization of the procedure with a minimum amount of cells (minimum 0,2x10^6 cells) and then multiply all the reagents by five and still have the same results in the 100µL cuvettes (4D-Nucleofector™ System Manual). Furthermore, the system allows for the performance of cell electroporation in sterile conditions.

LCL were flash thawed and maintained in a 37ºC incubator. Complete RPMI medium was changed every other day for one week, including on the day before the electroporation. At the day of the electroporation cells were always counted with a hemocytometer.

This system offers a variety of kits, as well as countless electroporation programs, depending on the target cell type. Kit SF (Lonza, Basel, Switzerland) was used throughout this project. The electroporation kit contains a Nucleofector solution and a Supplement solution that must be mixed and added to the cells prior to electroporation. The volume of cell suspension correspondent to the number of cells needed for the essay was centrifuged and the supernatant was completely discarded. The remaining pellet was resuspended in the eletctroporation mix and plasmid DNA was added. DNA suspension contained all plasmids in the same proportion (Barrett et al. 2014; Thomas et al. 2015).

Cells were then placed in strips or cuvettes and electroporated with the appropriate program. After electroporation cells were incubated for 48 hours in a 24- or 12-well plate and half of the reprogramming medium was changed. From this point on the reprogramming medium was changed every other day until the appearance of colonies (Annex 8.1).

4. RESULTS

4.1. CLINICAL DESCRIPTION

The proband presented intrauterine growth retardation. His fetal growth and weight were below the tenth percentile.

Apgar scores were 9 and 10 at one and five minutes, respectively. At birth, his weight was 2130g.

In the neonatal period, the proband presented congenital microcephaly.

At one year of age, the proband was diagnosed with severe developmental delay and left hemiparesis of brachial predominance.

Magnetic resonance imaging scan (MRI) revealed brain malformations, including right perisylvian cortical displasy with an associated schizencephalic cleft and signs of polymicrogyria in the posterior part of the right frontal lobe.

At two years of age, the proband presented the first epileptic episode. Subsequently, he was diagnosed with complex partial refractory epilepsy. At four years of age the proband was submitted to a cortical disconnection of the dysplastic area with few clinical improvements.

His growth development was stable and his weight is now within the tenth percentile and his height within the fifth percentile. The cephalic perimeter is still below the fifth percentile, confirming the previously observed microcephaly.

4.2. CYTOGENETIC STUDIES

Analysis of GLT-banded metaphase chromosomes of the proband (II:3, Figure 4.1), a child of a non-consanguineous Portuguese couple, revealed a *de novo* apparently balanced double translocation between the long arms of chromosomes 2 and 7 and the long arms of chromosomes 5 and 6 - 46,XY,t(2;7)(q23;q32),t(5;6)(q23;q26)dn (Figure 4.2). Karyotypes of the proband's parents were normal (I:1 and I:2, Figure 4.1), therefore the translocation is *de novo*. His siblings (II:1 and II:2, Figure 4.1) are healthy.

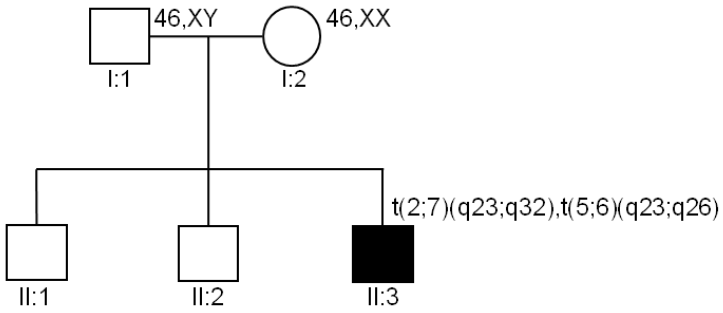

Figure 4.1: Pedigree of the proband's family with the *de novo* double translocation. The proband with the t(2:7)(q23;q32),t(5;6)(q23,q26)dn associated with intrauterine growth retardation, severe developmental delay, brain malformations and epilepsy is depicted by a black square.

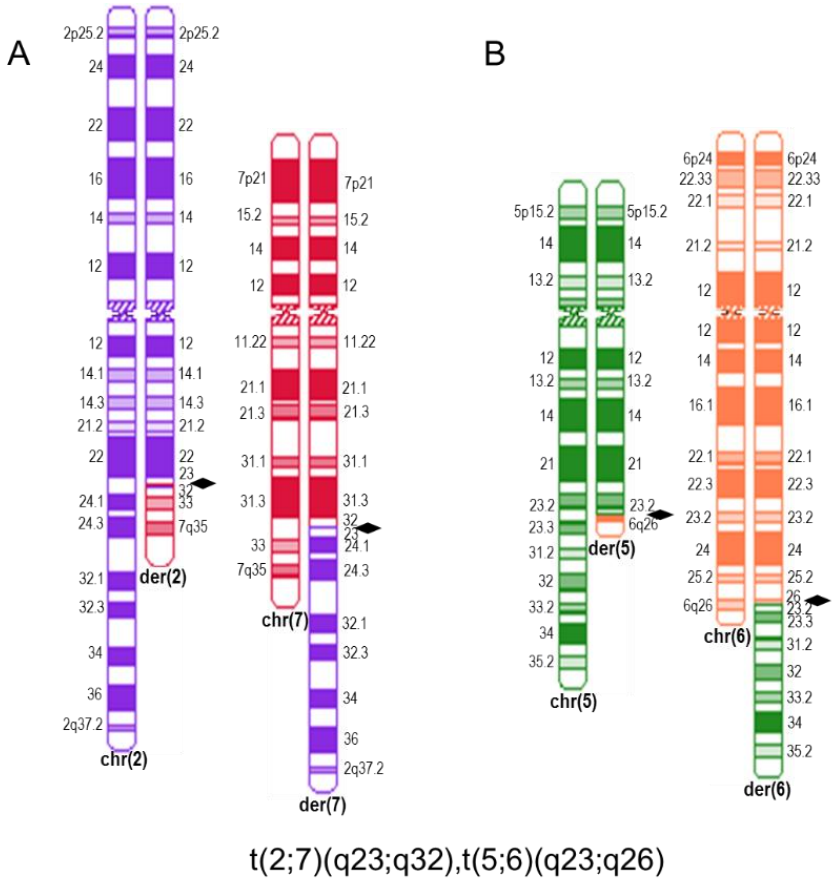

Figure 4.2:Ideograms of t(2;7)(q23;q32) and t(5;6)(q23;q26) with the breakpoints depicted by black diamonds.
A. chr(2) and der(2) with the breakpoint at 2q23 and chr(7) and der(7) with the breakpoint at 7q32.
B. chr(5) and der(5) with the breakpoint at 5q23 and chr(6) and der(6) with the breakpoint at 6q26

4.3. UNBALANCED GENOMIC ALTERATIONS

Initially, the proband was screened for unbalanced genomic alterations by high resolution whole-genome array analysis. A 651.76kb deletion was found at 14q24.3, between the markers C-6KQCH and S-3RPCO (chr14:76,673,181-77,324,937 [GRCh37/hg19]) (Figure 4.3).

Through the analyses of the genomic array it was possible to observe that this deletion disrupts two genes, G-patch domain containing 2 like (*GPATCH2L*) and chromosome 14 open reading frame 166 (*C14orf166*, OMIM*610858; https://www.omim.org/entry/610858). Besides the disrupted genes, the deletion encompasses three additional genes, estrogen related receptor beta (*ESRRB*), vasohibin 1 (*VASH1*; OMIM*609011; https://www.omim.org/entry/609011) and angel homolog I (*ANGEL1*).

Even though none of the disrupted genes seems to have an associated phenotype, one of the genes encompassed by the deletion, *ESRRB*, is associated with an autosomal recessive disorder. Several mutations on this gene have been reported has being associated with autosomal recessive deafness (DFNB35; OMIM#608565; http://omim.org/entry/608565) (Collin et al. 2008) and even with progressive dental decay (Weber et al. 2014). Furthermore, in mice, *ESRRB* is involved in prenatal inner ear development. In contrast, in the human population this gene appears to be expressed only postnatally in the cochlea (Collin et al. 2008).

About 224kb upstream of the deletion breakpoint it is possible to find transforming growth factor beta 3 (*TGFB3*) gene. *TGFB3* gene is a member of the transforming growth factors beta family, known for their pivotal role in neuronal development, and consequently for their involvement neurodevelopmental disorders (Samanta et al. 2008). This gene is associated with two autosomal dominant disorders, arrhythmogenic right ventricular dysplasia (ARVD1; OMIM#107970; https://www.omim.org/entry/107970) and Loeys-Dietz syndrome 5 (LDS5; OMIM#615582; https://www.omim.org/entry/615582).

ARVD1 is characterized by myocardial dystrophy. This disorder causes severe arrhythmias and is the major genetic cause of juvenile sudden death (Basso et al. 2009). On the other hand, LDS5 is characterized mainly by aortic aneurisms, cleft palate, bifid uvula, mitral valve disease and skeletal abnormalities, including clubfoot deformities, skeletal overgrowth and cervical spine instability (Bertoli-Avella et al. 2015). Additionally, at least one reported case presenting a missense mutation in this gene caused an additional phenotype of distal arthrogryposis, low muscle mass and growth retardation (Rienhoff et al. 2013).

Almost 300kb upstream of the deletion breakpoint another gene is associated with an autosomal recessive disorder. Intraflagellar transport 43 (*IFT43*) gene is associated with cranioectodermal displasya 3 (CED3; OMIM#614099; https://www.omim.org/entry/614099), a multi-system syndrome involving growth retardation and a number of characteristic skeletal, ectodermal and facial features, which vary between individuals depending on the affected locus (Arts et al. 2011; Arts and Knoers 2013). Recently, a microdeletion involving this gene was found on an individual presenting mild intellectual disability, skeletal anomalies, congenital heart defect, myopia and facial dysmorphism (Stokman et al. 2016).

Figure 4.3 Overview of the chromosome 14q24.3 deletion region.
A. Ideogram of chromosome 14.
B. Detailed physical map across the deleted region. Horizontal lines with folded gray arrows indicate the genes in sense (above the map) and antisense (below the map) orientation. Disrupted genes (*GPATCH2L* and *C14orf166B*) are marked with a hashtag. The breakpoints are depicted by a black arrows. Genes associated with autosomal dominant disorders are in bold (*TGFB3* – OMIM*190230) and genes associated with autosomal recessive disorders are underlined (*ESSRB* – OMIM*602167 – and *IFT43* – OMIM*614068).
C. Genomic array analysis of the 651.756 Kb deletion using the CytoScan HD array (Affymetrix). The red area highlights the deleted region. Log 2 ratio and copy number state of this region are also shown. Bellow, green rectangles mark genes *TGFB3*, *ITF43* and *ESRRB*, associated with developmental disorders.

4.4. COMPARATIVE MAPPING OF THE TRANSLOCATION'S BREAKPOINTS

In order to map the translocation's breakpoints to a reference genome, genomic amplicons of flow-sorted derivative chromosomes 5, 6 and 7 were analyzed by array painting (Figure 4.4).

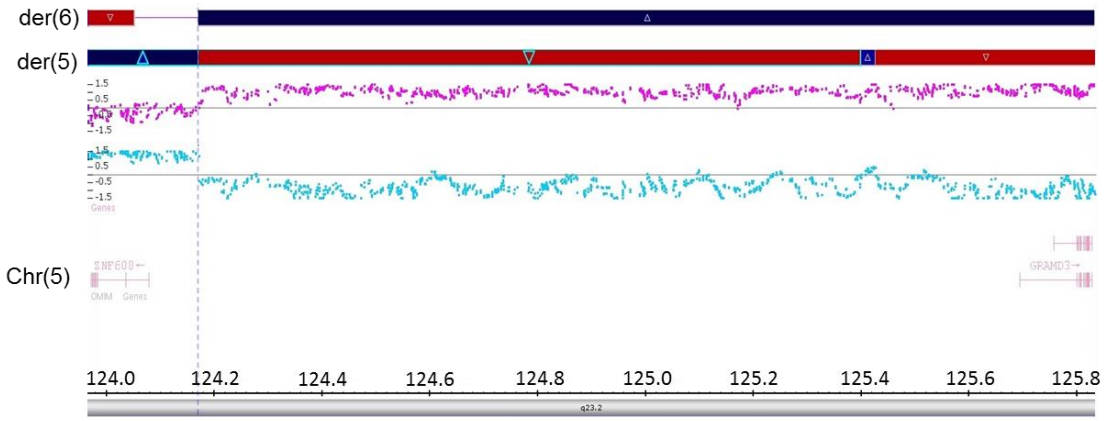

Figure 4.4: Array analysis of the t(5;6) breakpoint using CytoScan HD array from Affymetrix. A dashed line highlights the position of the breakpoints in der(5) and der(6). Allele peaks from both derivative chromosomes are shown. Flanking genes of chromosome 5 are depicted bellow.

In t(2;7)(q23.3;q32.1), the translocation breakpoint on der(2) is mapped within a 1.8kb region between copy number (CN) markers C-4LCOM and C-7BFEG (g.153,545,790 and g.153,547,538, respectively [GRCh37/hg19]). On der(7), the translocation breakpoint is localized within a 1.9kb region between CN markers C-4OFRR (g.127,652,367 [GRCh37/hg19]) and C-4SKW (g.127,654,318 [GRCh37/hg19]).

Regarding the t(5;6)(q23.2;q26), the translocation breakpoint on der(5) is localized between CN markers C-5XADF (g.124,172,389 [GRCh37/hg19]) and C-7KHAB (g.124,175,834 [GRCh37/hg19]), within a 3.4kb region. The translocation breakpoint on der(6) is localized within a 9.6kb genomic region delimited by the CN markers C-6RHHK (g.163,495,249 [GRCh37/hg19]) and C-5JCXT (g.163,504,881 [GRCh37/hg19]).

Recent advances in sequencing technologies have provided new tools for the identification of chromosome rearrangements' breakpoints with nucleotide resolution. In this way, and in order to further delimitate the breakpoint regions, whole-genome sequencing of large-insert jumping libraries (liWGS) was comparatively applied.

Regarding the t(2;7)(q23.3;q32.1), the liWGS approach localized the der(2) breakpoint within a 273bp region, at g.153,546,754-153,547,027 [GRCh37/hg19] and the der(7) breakpoint within a 159bp region between g.127,652,400-127,652,559 [GRCh37/hg19].

As for the t(5;6)(q23.2;q26), the der(5) breakpoint (BP3) is mapped to a 1.5kb region, at g.124,175,246-125,626,104 [GRCh37/hg19]. On der(6), the translocation breakpoint is defined within a 541bp region at g.128,510,992-128,511,533 [GRCh37/hg19], being repositioned to 6q22.33.

When comparing the size of the translocation breakpoint regions found by array painting with the ones found using liWGS, it is possible to understand that liWGS allows for a significant restriction of the disrupted regions. The breakpoint regions found by liWGS for both translocations appear within a much smaller genomic region. The breakpoint regions of t(2;7)(q23.3;q32.1) go from 1.8kb and 1.9kb regions to 273bp and 159 bp regions, respectively, and the breakpoint regions of t(5;6)(q23.2;q26) are restricted from 3.4kb and 9.6kb regions to 1.5kb and 541bp regions, demonstrating the much higher resolution provided by the application of this new technology.

Besides the identification of the translocations breakpoints, the liWGS approach allowed for the identification of three additional breakpoints at der(6).

At 5q23.2, a breakpoint defining a 1.45Mb inversion was found within positions g.124,176,814-125,626,573[GRCh37/hg19].

At 6q22.33, another breakpoint is localized within g.128,462,041-128,511,533 [GRCh37/hg19], defining a 48.3kb excision. The excised 48.3kb region is then inserted at 6q26, causing one more breakpoint between g.163,468,284 [GRCh37/hg19] at 6q26, and g.128,463,801 [GRCh37/hg19] at 6q22.33 (Figure 4.5).

The array painting approach is insensitive to the orientation of the genomic DNA, not being able to detect the inversion breakpoint in der(6) at 5q23.2. Furthermore, since the excised region, which is posteriorly inserted at 6q26, is only 48.3kb in size, the resolution of the array painting does not allow for the observation of this small structural rearrangement. In this way, without the application of liWGS it would not be possible to unravel these cryptic genomic alterations.

Additionally, the liWGS methodology confirmed the deletion previously found by high resolution whole genome array analysis. This approach detected a 648.005kb deletion mapped at 14q24.3: 76,673,103-77,321,108 [GRCh37/hg19].

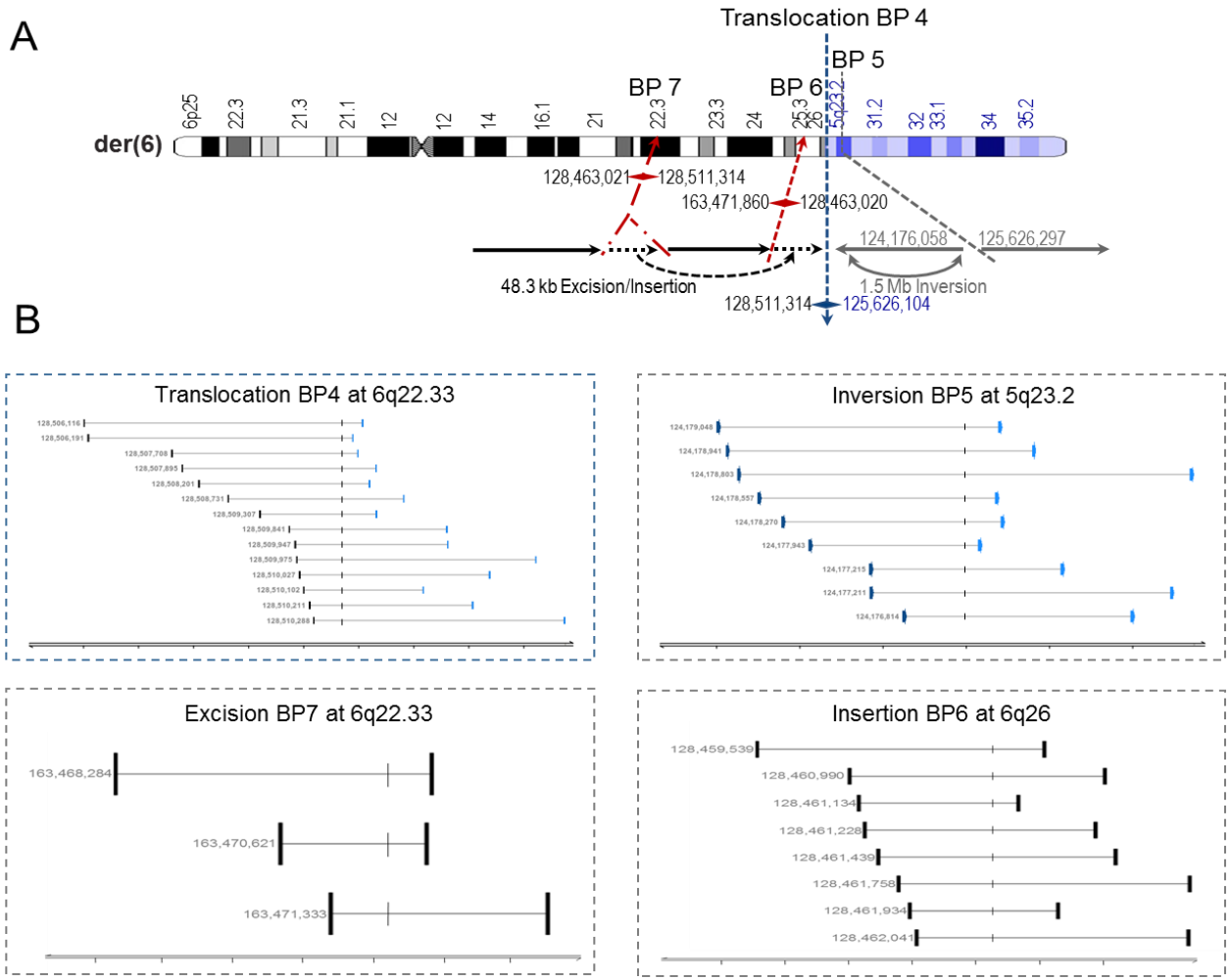

Figure 4.5: Mapping of the breakpoints by li-WGS.
A. Ideogram of derivative chromosome 6 with the indication of the translocation breakpoint (BP4 at 6q22.33), the inversion breskpoint (BP5 at 5q23.2), the insertion breakpoint (BP6 at 6q26) and the excision breakpoint (BP7 at 6q22.33).
B. Breakpoint supporting read pair from the regions of der(6). The breakpoint position at a nucleotide level is depicted by small black dashes.

4.5. AMPLIFICATION AND SEQUENCING OF JUNCTION FRAGMENTS

In order to position the translocation's breakpoints at nucleotide resolution primers flanking each of the seven identified breakpoints were designed. PCR conditions were optimized in order to obtain specific control and junction fragments (Figure 4.6). Amplified fragments were sequenced by Sanger sequencing in both orientations (Figure 4.7, 4.8 and 4.9).

Concerning t(2;7)(q23.3;q32.1), the breakpoint at der(2) is localized at position g.153,546,999-153,547,000 [GRCh37/hg19]. A 31bp deletion of the chromosome 2 sequence was found at the breakpoint junction. At der(7) the breakpoint is localized at position g.127,652,519-127,652,520 [GRCh37/hg19], presenting a 10bp deletion of the chromosome 7 sequence at the breakpoint junction (Figure 4.7).

When it comes to t(5;6)(q23.2;q26), the der(5) breakpoint is localized at position g.124,175,407-124,175,408 [GRCh37/hg19], presenting a 668bp deletion of the chromosome 5 sequence at the breakpoint junction, as well as a 12bp insertion. As for the der(6) translocation breakpoint, it is localized at position g. 128,511,147-128,511,148 [GRCh37/hg19] and a 31bp deletion of the chromosome 6 sequence was found at the breakpoint junction (Figure 4.8).

Regarding the breakpoints only identified by the liWGS approach on der(6), the 5q23.2 inversion breakpoint is localized at position g.124,176,336-124,176,337 [GRCh37/hg19], presenting a 71bp deletion of chromosome 5 sequence at the breakpoint junction. The insertion breakpoint at 6q26 is localized at position g.163,472,151-163,472,152 [GRCh37/hg19], presenting a 17bp deletion at the breakpoint junction. As for the excision breakpoint at 6q22.33, it is localized at g.128,463,242-128,463,243 [GRCh37/hg19], presenting a 68bp deletion of chromosome 6 sequence at the breakpoint junction (Figure 4.9).

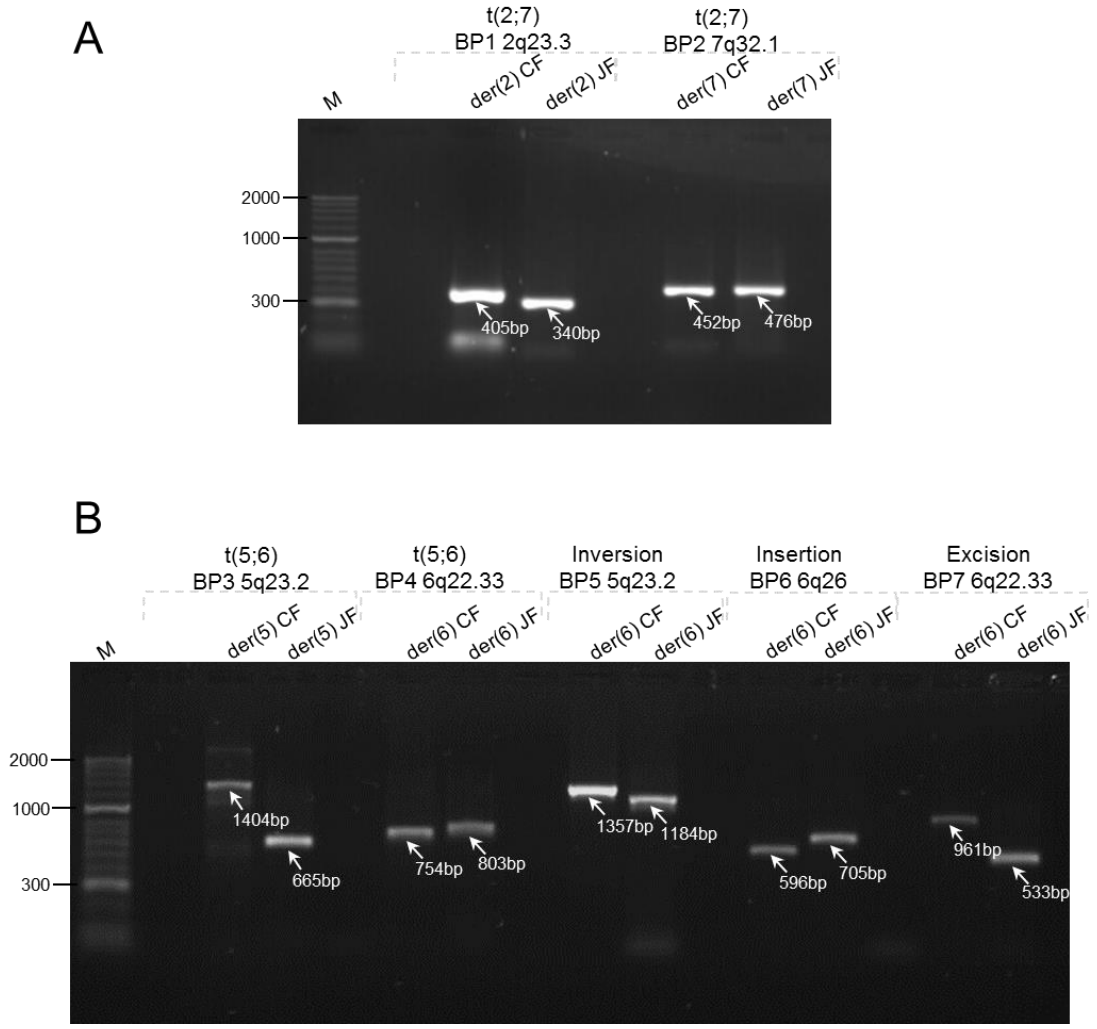

Figure 4.6: Amplification and DNA sequence of the der(2), der(7), der(5) and der(6) junction and control fragments.
A. Amplification of the t(2;7) junction and control fragments using chromosome 2 and 7 specific PCR primers. A 340bp junction fragment (BP1 2q23.3) was obtained for der(2) and a 476bp junction fragment (BP2 7q32.1)) was obtained for der(7).
B. Amplification of the t(5;6) junction and control fragments, including the additional rearrangements found by liWGS, using chromosome 5 and 6 specific PCR primers. A 665bp junction fragment (BP3 5q23.2) was obtained for der(5) and a 803bp junction fragment (BP4 6q22.33) was obtained for the translocation breakpoint on der(6). As for the three additional rearrangements on der(6), a 1184bp junction fragment (BP5 5q23.2) was obtained for the inversion distal breakpoint, a 705bp junction fragment (BP6 6q26) was obtained for the insertion breakpoint and a 533bp junction fragment (BP7 6q22.33 was obtained for the excision breakpoint.
(BP - Breakpoint; M - HyperLadder II marker; NC - negative control; CF - control fragment; JF - Junction fragment).

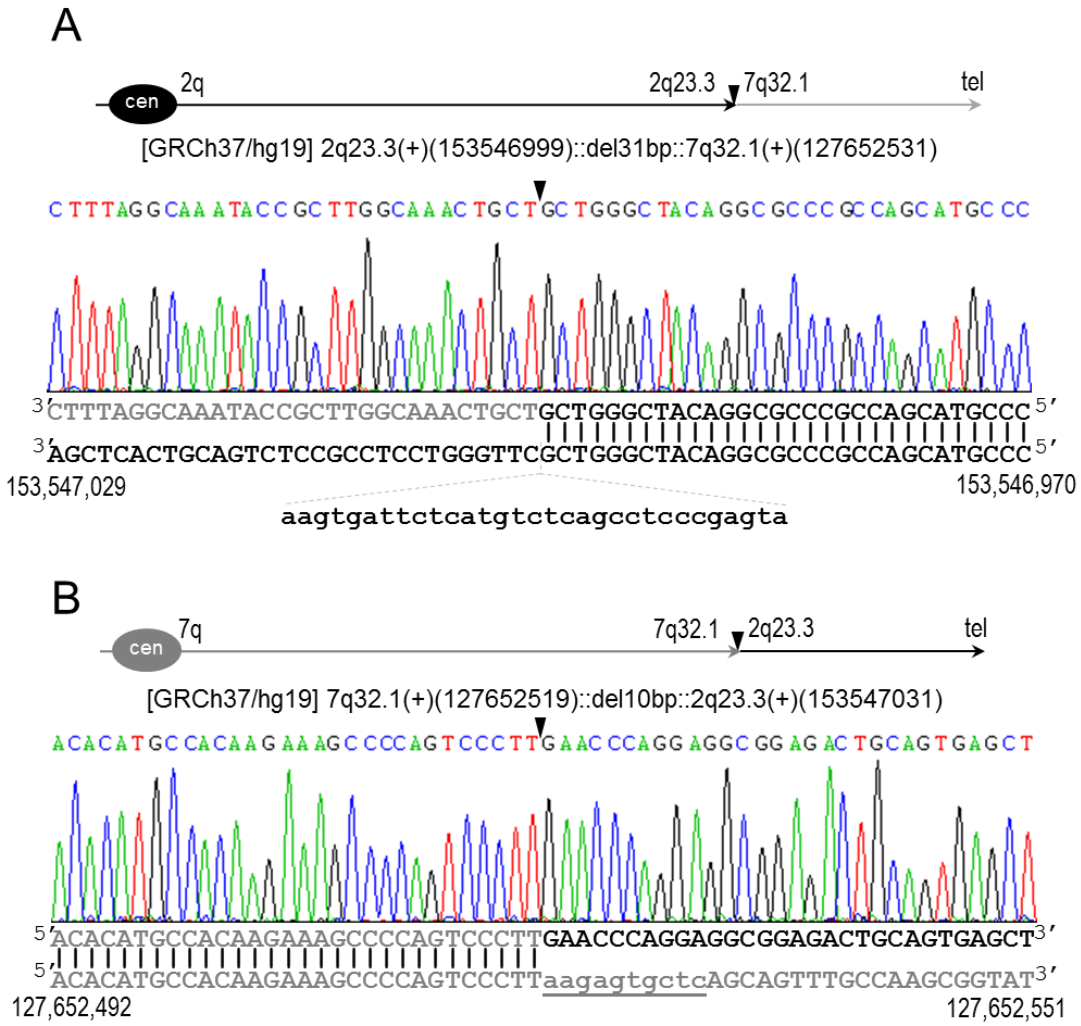

Figure 4.7: Breakpoints of t(2;7)(q23.3;q32.1) at nucleotide resolution. Breakpoints of both der(2) and der(7) are depicted by a black inverted triangle.
A. der(2) breakpoint sequence aligned with the reference genome A 31bp deletion was found at the breakpoint junction.
B. der(7) breakpoint sequence aligned with the reference genome. A 10bp deletion was found at the breakpoint junction.
Refrence Genome GCRh37/hg19.

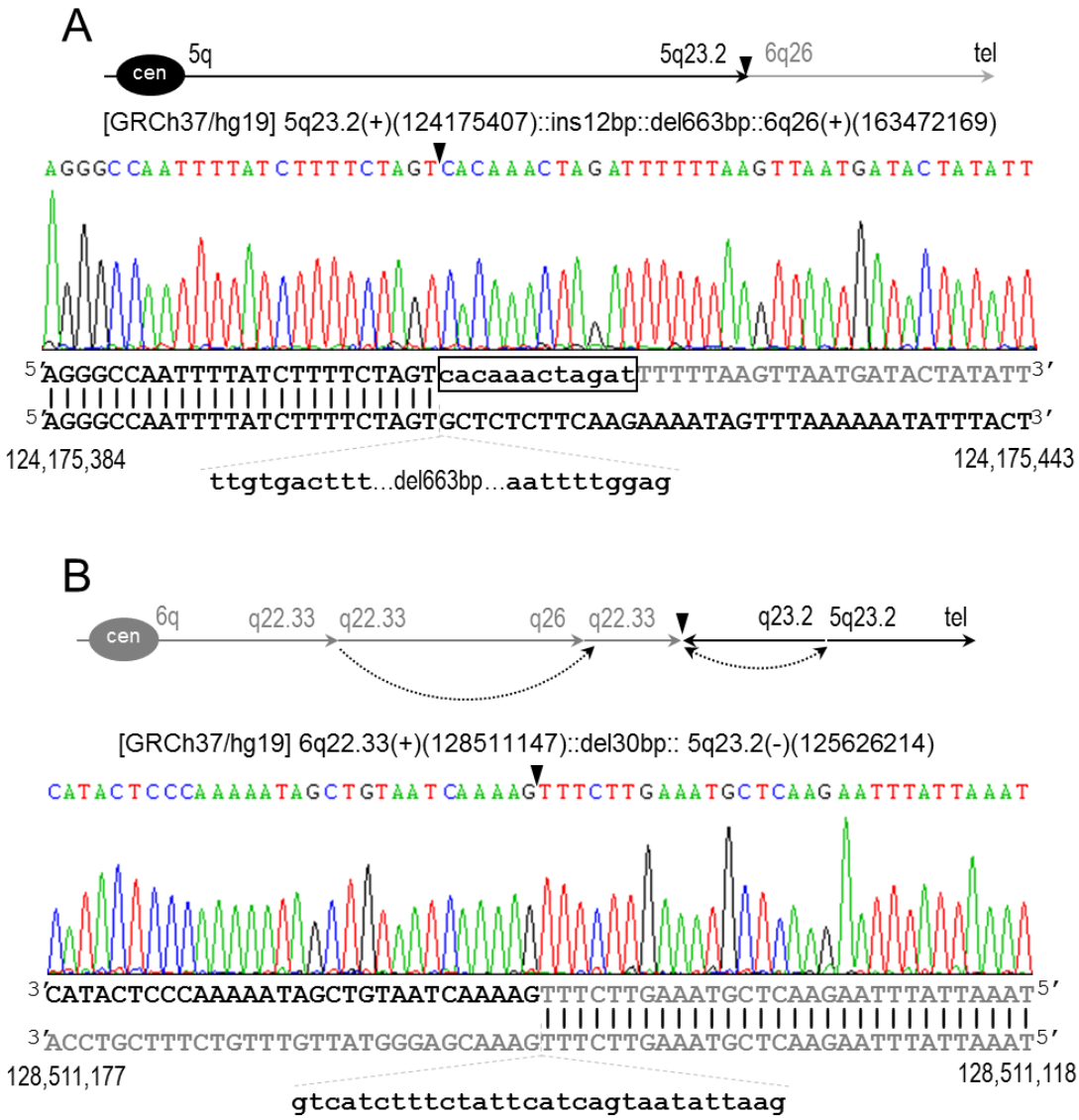

Figure 4.8: Breakpoints of t(5;6)(q23.2;q26) at nucleotide resolution. Breakpoints of both der(5) and der(6) are depicted by a black inverted triangle.
A. der(5) translocation breakpoint sequence aligned with the reference genome. A 12bp insertion at the breakpoint region is marked by a black box. A 663bp deletion was found at the breakpoint junction.
B. der(6) translocation breakpoint sequence aligned with the reference genome. A 30bp deletion was found at the breakpoint region.
Reference Genome GRCh37/hg19.

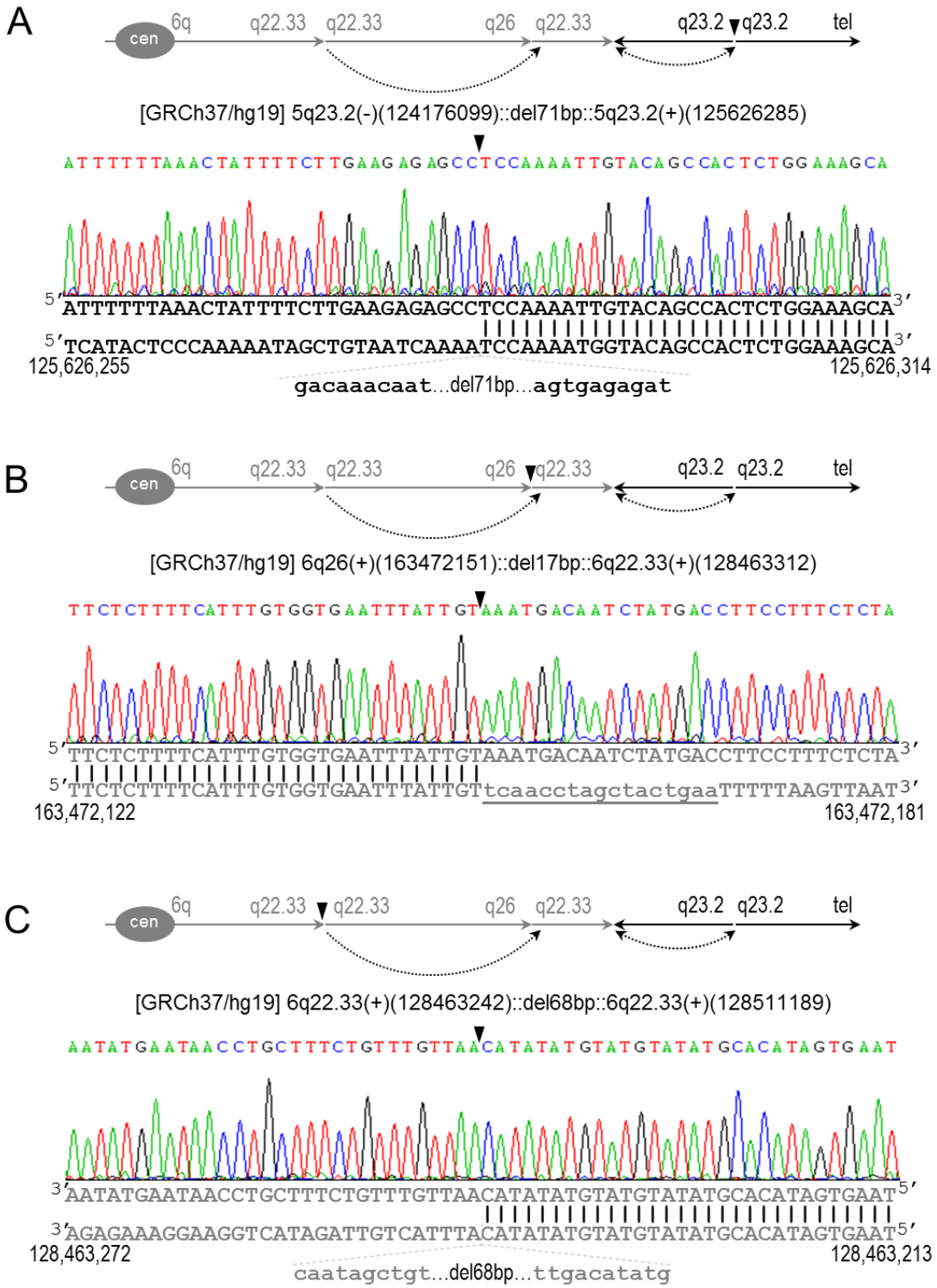

Figure 4.9: Breakpoints of the cryptic rearrangements found in der(6) at nucleotide resolution. All breakpoints are depicted by a black inverted triangle.
A. Inversion breakpoint (BP5 at 5q23.2) sequence aligned with the reference genome. A 71bp deletion was found at the breakpoint junction.
B. Insertion breakpoint (BP6 at 6q26) sequence aligned with the reference genome. A 17bp deletion was found at the breakpoint junction.
C. Excision breakpoint (BP7 6q22.33) sequence aligned with the reference genome. A 68bp deletion was found at the breakpoint region.
Reference Genome GRCh37/hg19.

4.6. IDENTIFICATION OF POSSIBLE CANDIDATE GENES AT THE TRANSLOCATION'S BREAKPOINTS

Genomic regions surrounding the translocation's breakpoints were analyzed in order to screen for genes that might help explain the proband's phenotype.

The breakpoints at t(2;7)(q23.3;q32.1) disrupt two genes, pre-mRNA processing factor 40 homolog A (*PRPF40A*) at der(2) and Staphylococcal nuclease and tudor domain containing 1 (*SND1*, OMIM *602181; https://www.omim.org/entry/602181) gene at der(7).

PRPF40A is a protein coding gene associated with Huntington's disease (HD; OMIM#143100; https://www.omim.org/entry/143100) and Rett Syndrome. This gene codes for WW domains that bind to both huntingtin, the protein responsible for Huntington disease, and methyl-CpG-binding protein, associated with Rett syndrome. Both referred diseases result in progressive neurodegeneration due to the accumulation of mutant neurotoxic proteins (Kato et al. 2006).

Upstream of the *PRPF40A* gene, at a distance of 600kb, is the calcium channel, voltage-dependent, beta-4 subunit (*CACNB4*) gene, which encodes for a member of the beta subunit family of voltage-dependent calcium channel complex proteins and plays an important role on membrane polarization. *CACNB4* has been associated with three epilepsy related autosomal dominant disorders: episodic ataxia type 5 (EA5; OMIM#613855; https://www.omim.org/entry/613855), idiopathic generalized epilepsy (EIG9; OMIM#607682; https://www.omim.org/entry/607682) and juvenile myoclonic epilepsy (EJM6; OMIM#607682; https://www.omim.org/entry/607682) (Escayg et al. 2000). Furthermore, this gene is currently in study to be a part of a gene panel to diagnose refractory epilepsy (Segal et al. 2016).

About 1Mb upstream of the der(2) breakpoint it is possible to find the nebulin (*NEB*) gene. This gene is associated with an autosomal recessive nemaline myopathy (NEM2; OMIM#256030, https://www.omim.org/entry/256030), which is characterized by progressive muscle weakness throughout the body (Figure 4.10A).

Besides disrupting *SND1* gene, as mentioned above, the der(7) breakpoint is surrounded by three disorder associated genes.

RNA binding motif protein 28 (*RBM28*) is located 300kb downstream of the *SND1* gene and is one of the genes involved in regular ribosomal function (Freed et al. 2010). This gene is associated with an autosomal recessive disorder characterized by alopecia, progressive neurological defects, and endocrinopathy syndrome (ANES; OMIM#612079; https://www.omim.org/entry/612079). RBM28 also seems to be involved in the development of the neural system (Nousbeck et al. 2008).

Also associated with an autosomal recessive disorder, the leptin (*LEP*) gene is located 200kb downstream of the der(7) breakpoint. *LEP* gene is associated with morbid obesity due to leptin deficiency (LEPD, OMIM#614972; https://www.omim.org/entry/614962), an autosomal recessive disorder characterized by an abnormally elevated body fat percentage.

Additionally, paired domain gene 4 (*PAX4*) is located 400kb upstream of the der(7) breakpoint. *PAX4* gene has been associated with autosomal dominant non-insulin dependent diabetes mellitus (NIDDM; OMIM#125853; https://www.omim.org/entry/125853) and ketosis prone diabetes mellitus (KPD; OMIM#612227; https://www.omim.org/entry/612227) (Figure 4.10B).

Regarding the t(5;6)(q23.2:q26), the der(5) translocation breakpoint disrupts an intergenic region and does not present any disorder associated gene in close proximity. On the other hand, the der(6) translocation breakpoint disrupts the Parkin coregulated (*PACRG*; OMIM*608427; https://www.omim.org/entry/608427) gene, which shares a bi-directional promoter with Parkin RBR E3 Ubiquitin Protein Ligase (*PARK2*) gene. *PARK2* gene has been associated with autosomal recessive juvenile Parkinson disease (PARK2; OMIM#600116; https://www.omim.org/entry/600116) (Lesage et al. 2007; Kay et al. 2010). *PACRG* gene and *PARK2* gene are abundantly expressed in the human brain, having been found in *substancia nigra* nuclear extracts (Kitada et al. 1998; West et al. 2003).

Even though it is not formally associated with any human disorder, 100kb downstream of the bidirectional promoter shared by *PACRG* gene and *PARK2* gene, the homolog of quaking mouse (*QKI*; OMIM#609590; https://www.omim.org/entry/609590) gene also plays a role in brain development. In fact, this gene seems to be involved in myelination and oligodendrocyte differentiation (Ebersole et al. 1996) (Figure 4.11C).

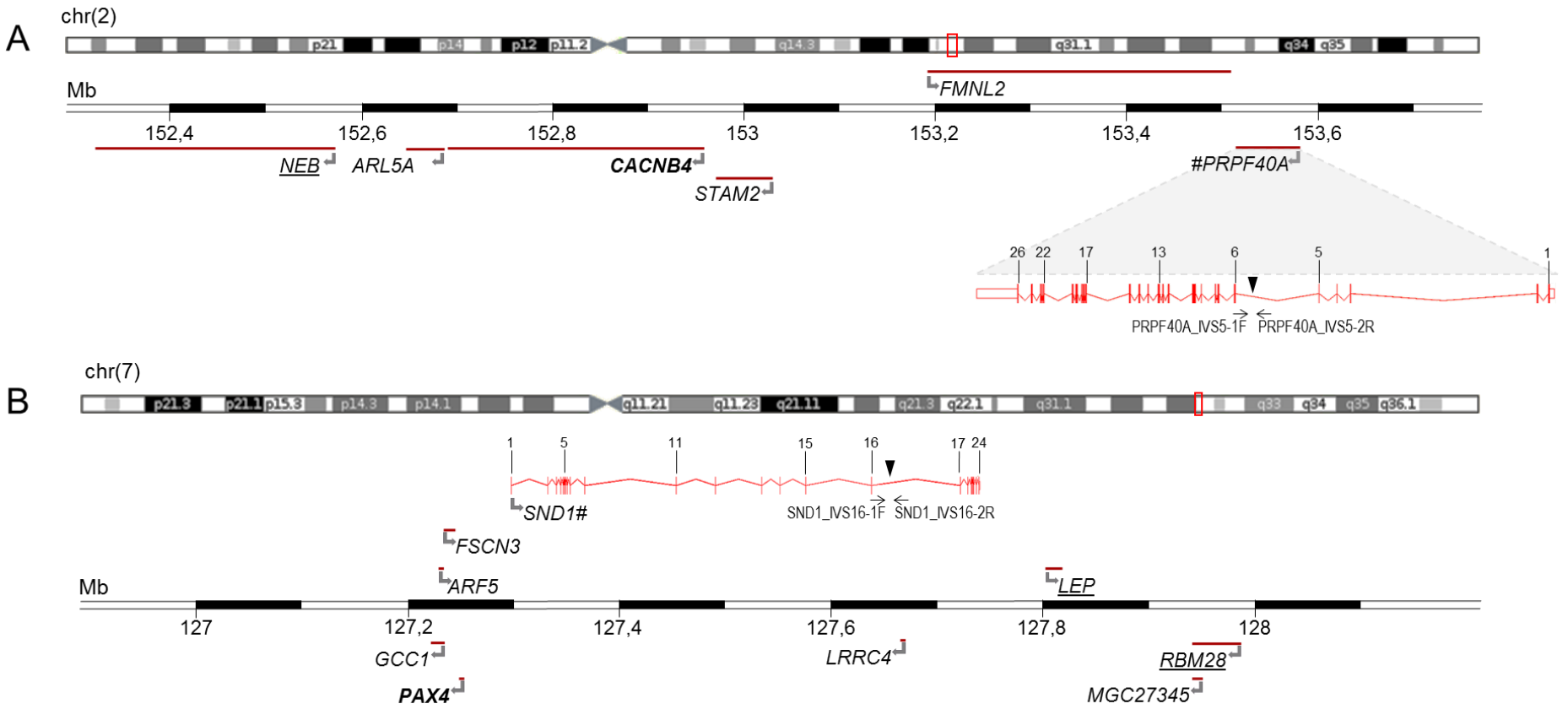

Figure 4.10: Physical maps across the breakpoint regions of t(2;7)(q23.3;q32.1).
A. Breakpoint region of chr(2) (BP1 2q23.3).
B. Breakpoint region of chr(7) (BP2 7q32.1).
Genes are represented by red horizontal lines with folded arrows indicating the orientation (sense or antisense). Disrupted genes (*PRPF40A* – OMIM*612941 - and *SND1* – OMIM*602181) are marked with a hashtag and their genomic organization is depicted. Black vertical lines show exon numbers. The breakpoint is depicted by a black inverted triangle and the used primers are indicated by horizontal arrows. Genes associated with autosomal dominant disorders are in bold (*CACNB4* - OMIM*601949 – and *PAX4* – OMIM*167413) and genes associated with autossomic recessive disorders are underlined (*NEB* – OMIM*161650 – and *RBM28* – OMIM*612074).

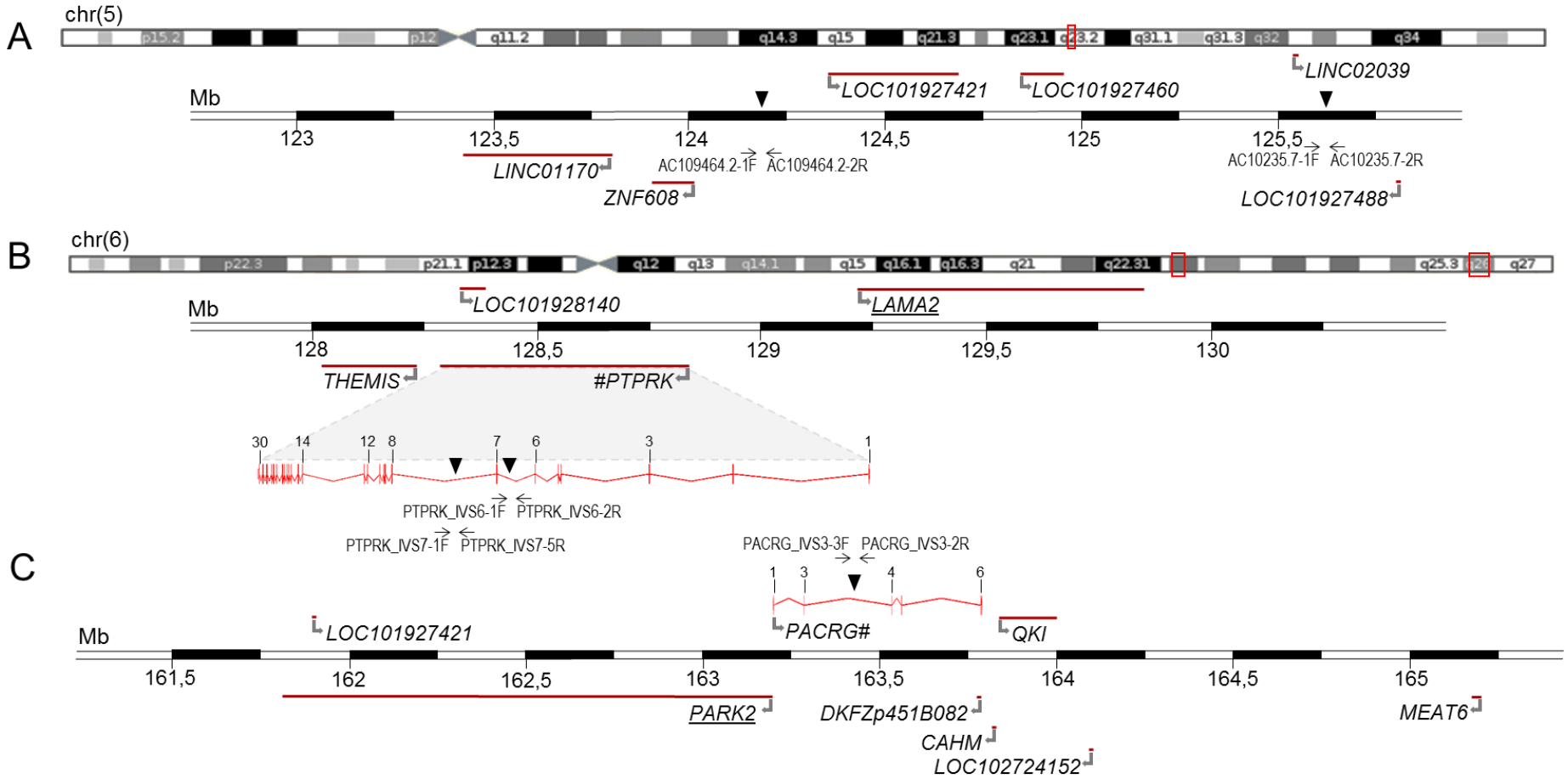

Figure 4.11: Physical maps across the breakpoint regions of t(5;6)(q23.2;q26), including the rearrangements found by liWGS.
A. Breakpoint regions of chr(5). Translocation breakpoint (BP3 5q23.2) is at g.124,175,407 and the inversion breakpoint (BP5 5q23.2) is at g. 124,176,099.
B. Breakpoint regions of chr(6). Translocation breakpoint region of chr(6) (BP4 6q.22.33) is at IVS6 of PTPRK and the excision breakpoint (BP7 6q22.33) is at IVS7 of the same gene.
C. Insertion breakpoint region of chr(6) (BP6 6q26).
Genes are represented by red horizontal lines with folded arrows indicating the orientation (sense or antisense). Disrupted genes (*PACRG* – OMIM*608427 – and *PTPRK* – OMIM*602545) are marked with a hashtag and their genomic organization is depicted. Black vertical lines show exon numbers. The breakpoints are depicted by a black inverted triangle and the used primers are indicated by horizontal arrows. Genes associated with autosomal recessive disorders are underlined (*PACRG* and *LAMA2* – OMIM*156225).

4.7. IDENTIFICATION OF POSSIBLE CANDIDATE GENES AT THE CRYPTIC ALTERATIONS' BREAKPOINTS

Due to the application of liWGS three new breakpoints, corresponding to three cryptic alterations, were found in der(6).

The inversion breakpoint found at 5q23.2 is located 150kb from the der(5) translocation breakpoints and does not disrupt any gene, nor is it near to any disorder related gene (Figure 4.11A).

Regarding the excision/insertion breakpoints at 6q22.33, both disrupt Protein Tyrosine Phosphatase, Receptor Type K (*PTPRK*; OMIM#602545; https://www.omim.org/entry/602545) gene which is mostly associated with tumor suppression. However, localized 70kb downstream of this gene there is the laminin alpha-2 (*LAMA2*) gene, which is associated with autosomal recessive merosin-deficient congenital muscular dystrophy type 1A (MDC1A; OMIM#607855; https://www.omim.org/entry/607855) (Schéele et al. 2007). Furthermore, in 2009 a homozygous mutation in the *LAMA2* gene was reported as being associated with both MDC1A and severe brain malformations, including polymicrogyria. Additionally, the individual presenting this mutation developed seizures consistent with absence epilepsy (Vigliano et al. 2009) (Figure 4.11B).

4.8. PLURIPOTENCY INDUCTION FROM INDIVIDUAL SPECIFIC CELLS

4.8.1 Characterization of electroporation ready plasmid DNA

In order to assess the feasibility of the development of an individual specific cellular model to study chromosomal rearrangements, an attempt at pluripotency induction from individual derived LCL was made using episomal plasmid DNA as the transfection vector.

Episomal plasmid DNA was isolated and subjected to enzymatic restriction. The digestion patterns observed in agarose gel were as expected. pCXLE-hOCT3/4-shp53-F presented bands with 5513bp, 3334bp and 2834bp and pCXLE-hSK presented bands with 5513bp, 4293bp and 2941bp (Figure 4.12). The same was observed in pCXLE-hUL, confirming the identity of the isolated plasmids.

In the same agarose gel undigested plasmid DNA was also applied in order to assess the quality of the extraction. Due to DNA passage through the membrane on the purification column and to constant pipetting throughout the process, plasmid DNA may appear nicked, linearized or in a supercoiled conformation. All plasmids presented a higher concentration of supercoiled conformation, which increases the chances of a successful transfection (Figure 4.12).

The integrity of the isolated plasmids was assessed by PFGE (Figure 4.13). Most analyzed plasmids seemed to be integrate, appearing in a single high concentration band. Both pmax-GFP, provided by Lonza in the eletroporation kit, and pEP4-E02S-ET2K, extracted in house, presented more than one band in the agarose gel, which may be result of degradation because of long time storage and several frosting-defrosting cycles.

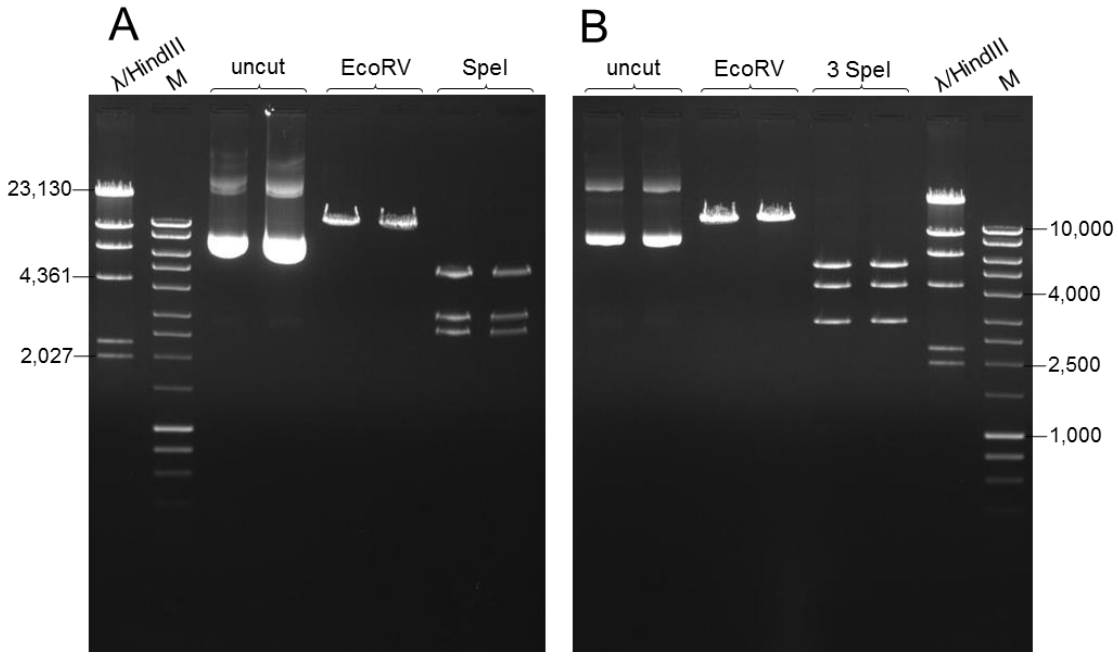

Figure 4.12: Agarose gels depicting whole plasmids and their enzymatic digestion with EcoRV e SpeI.
A Agarose gel depicting two samples of the pCXLE-hOCT3/4-shp53-F plasmid;
B. Agarose gel depicting two sample of the pCXLE-hSK plasmid.
(M - Hyper Ladder I marker).

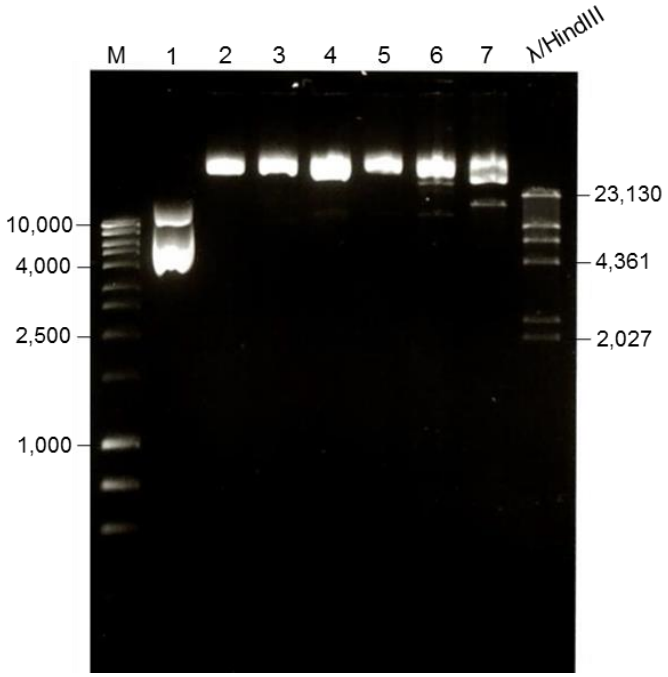

Figure 4.13: PFGE agarose gel depicting all the plasmids used to induce pluripotency.
(M - Hyper Ladder I Marker, 1 - pmax-GFP Lonza; 2 - pCXLE-EGFP (7); 3 - pCXLE-EGFP (8); 4 - pCXLE-hOCT3/4-shp53-F; 5 - pCXLE-hUL; 6 - pCXLE-hSK; 7 - pEP4-EO2S-ET2K)

4.8.2 Feeder Layer

In order to successfully go through the induction process, electroporated cells need a support matrix that is able to provide all the factors needed to cell proliferation. To use as a feeder-layer, human foreskin fibroblasts (HFF) were irradiated to become mitotically inactive.

After irradiation, HFF cells were plated and were expected to become adherent in 24 hours. Following optimization assays, a first batch of irradiated cells was plated and did not become adherent. Instead, amorphous cell clumps unable to sustain electroporated cells formed. It was possible to observe cell fusion, suggesting that irradiation had a great impact on cell membrane integrity. After a series of assays, it was possible to understand that this was due to the tubes used when the cells were irradiated. Even though both chosen tubes were made of polypropylene, it was found that only Falcon tubes were appropriate for irradiation (Figure 4.14).

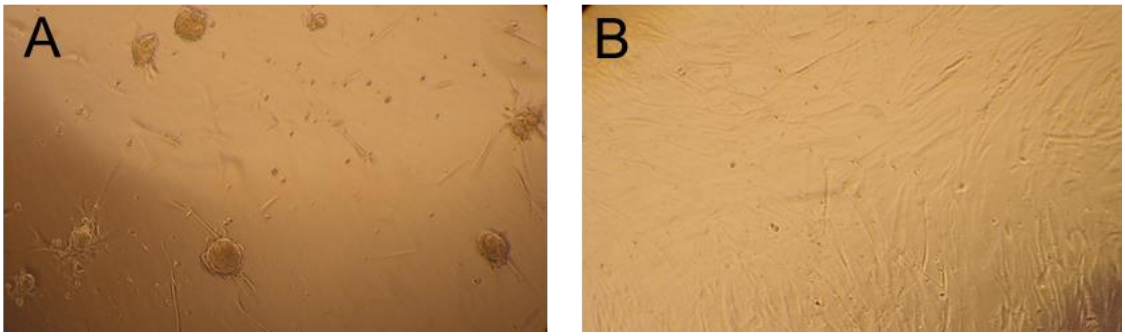

Figure 4.14: Inactivated human foreskin fibroblasts as feeder-layer.
A. Aggregated HFF 24 hours after irradiation.
B. Adhered inactivated HFF 24 hours after inactivation, ready to be used as a feeder layer for iPSCs.

4.8.3 Electroporation Optimization Assay and Plasmid Transfection

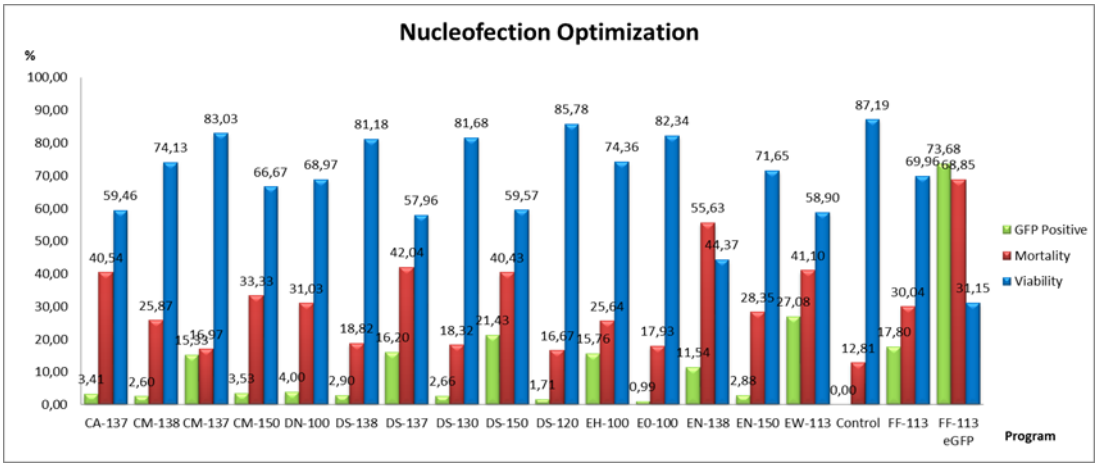

Figure 4.15: Optimization assay of Lonza's SF kit using LCL culture.
All the suggested programs were used and a negative control, without pmax-GFP, was subjected to electroporation to establish a baseline of cell mortality. Twenty-four hours after the assay, the cells correspondent to each of the programs were counted with a hemocytometer in order to establish viability, mortality and GFP positive cells. For the program FF-113, which presented the best efficiency results in Lonza's database, an assay was also made using the pCXLE-eGFP plasmid isolated in house. The bars in green represent the percentage of counted GFP positive cells, the bars in red represent the mortality rate and the bars in blue represent the viability rate.

After plasmid DNA characterization and HFF inactivation, an optimization assay was made using the SF kit from Lonza (Figure 4.15).

Cells electroporated using program FF-113 and the in house isolated GFP (pCXLE-eGFP) presented an electroporation efficiency of 74%. However the mortality rate was of 69%, meaning that the effective number of cells able to proceed with the induction process was very low. The two programs with closer efficiencies were EW-113, which showed an efficiency of 27% and a mortality rate of 41% when used with pmax-GFP, and DS-150, which showed an efficiency of 21% and a mortality rate of 40% with the same plasmid.

Using the three programs referred above, several electroporation assays were performed. In all assays GFP was added to the induction plasmid mix in order to be possible to calculate electroporation efficiency. Even though there were always a high concentration of GFP positive cells in the center of the wells in the 12- or 24-well plates (Figure 4.16), electroporation efficiency was always extremely low. Regardless, electroporated cells were plated onto the feeder-layer and followed for at least 21 days.

Usually, after 7 days of incubation most electroporated cells were adherent. It was considered that any non-adherent cells were not successfully transfected. In some of the assays round structures similar to pluripotent colonies were observed at around 14 days after electroporation. These structures were monitored for about two weeks. However, these colony-like structures never expanded and were not able to proliferate when passed to a new feeder layer.

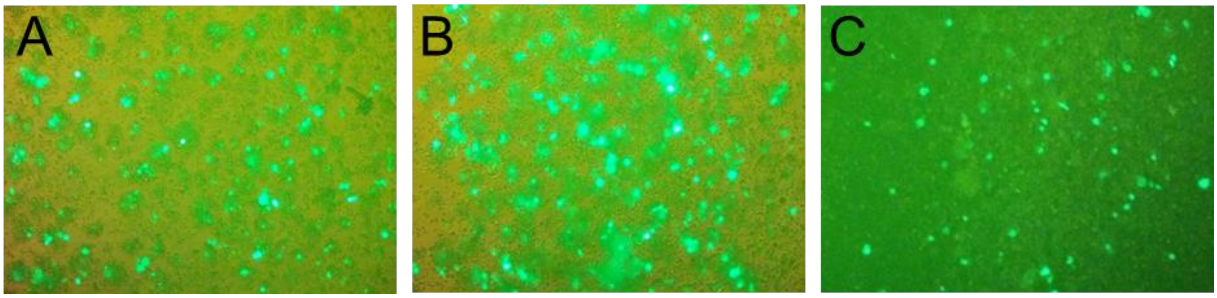

Figure 4.16: Observed fluorescence in the iPSC assays.
A. Fluorescence observed in the center of the well correspondent to the FF-113 assay;
B. Fluorescence observed in the center of the well correspondent to the EW-113 assay;
C. Fluorescence observed in the center of the well correspondent to the DS-150 assay.

5. DISCUSSION

5.1. IDENTIFICATION OF THE BREAKPOINT REGIONS OF THE REPORTED *DE NOVO* DOUBLE TRANSLOCATION

The proband is a male with a *de novo* balanced double translocation between the long arms of chromosomes 2 and 7 and the long arms of chromosomes 5 and 6 [t(2;7)(q23.3;q32.1),t(5;6)(q23.2;q26)dn], associated with a phenotype characterized by intrauterine growth retardation, severe developmental delay, brain malformations and epilepsy.

Derivative (der) chromosome 6 is, in fact, a complex rearrangement, presenting cryptic alterations including an inversion at 5q23.2 and an excision/insertion at 6q22.33.

Altogether these chromosomal rearrangements introduced seven breakpoints, four of which disrupt protein coding genes (*PRPF40A*, *SND1*, *PACRG* and *PTPRK*).

In addition, a 648.005 kb deletion at 14q24.3 (g.76,673,103-77,321,108) was also identified in the proband.

Regarding this deletion, even though neither the disrupted genes nor any of the genes encompassed by the deletion can explain the phenotype presented by the proband, *TGFB3* gene, located 224kb upstream of the deletion breakpoint, may be accountable for modelling brain abnormalities consistent with developmental delay (Samanta et al. 2008).

The use of array painting with genomic amplicons of the der chromosomes allowed for the mapping of the translocations' breakpoints. By posteriorly using the liWGS approach, it was possible to confirm and further delimit the translocations' breakpoints, demonstrating the possibility of breakpoint mapping at a much higher resolution than the previously used techniques. The true strength of the liWGS approach is the unraveling of cryptic alterations. In fact, only the use of this approach allowed for the identification of the further rearrangements on der(6).

Besides *TGFB3* gene, most of the genes present in the double translocation's breakpoint regions, as well as in the regions affected by the cryptic alterations, may explain the observed phenotype.

PRPF40A, the gene disrupted by the 2q23.2 breakpoint, is associated to progressive neurodegenerative disorders. Furthermore, it has previously been indicated as one the genes modeling non-syndromic intellectual disability (Kaufman et al. 2010). Concomitantly, 300 kb downstream of the 7q32.1 breakpoint, *RBM28* gene is also associated with both neurological defects and neural system development, with particular emphasis on pituitary hypoplasia (Nousbeck et al. 2008). Pituitary hypoplasia causes a shortage on growth hormone from a very early stage of intrauterine development (Fang et al. 2016), which may explain the intrauterine growth retardation observed in the proband.

The 6q26 translocation breakpoint disrupts the *PACRG* gene, which shares a bidirectional promoter with *PARK2* gene. The deletion of this promoter and several mutations in these genes have consistently been associated with early onset parkinsonism (Lesage et al. 2007; Kay et al. 2010). However, it is on the Quaking viable mouse model that it is possible to find phenotypical characteristics that may help explain the proband's phenotype.

The Quaking viable mouse, presents alterations in the expression of the *PACRG* gene and *PARK2* gene, significantly reducing the expression of *QKI* and completely depleting *PACRG* expression (Lockhart et al. 2004). The phenotype presented by this model includes male sterility related with low motility of the spermatozoan flagella and dysmyelination of the central nervous system (Wilson et al. 2010).

In humans, deletions encompassing the *QKI* gene, located 100 kb downstream of the *PARK2/PACRG* bidirectional promoter, were reported in 18 individuals. Even though the phenotype varied among individuals, all presented intellectual disabilities, hypotonia, seizures, brain abnormalities and dysmorphic facial features. Furthermore, a *de novo* balanced translocation involving chromosomes 5 and 6 – t(5;6)(q23.1;q26) – that disrupts the *QKI* gene was reported as being associated with a phenotype of borderline mental retardation, dysmorphic features and relative microcephaly (Backx et al. 2010), which is also consistent with the presented proband's phenotype.

Located 70kb distal of the cryptic excision/insertion breakpoint at 6q22.33, *LAMA2* gene has been associated with severe brain malformations. Animal models deficient in the α2 chain of laminin present severe muscle dystrophy and are not able to survive past 5 weeks of age (Guo et al. 2003). Furthermore, in the mouse neuromuscular system, the absence of the α2 chain causes severe muscle weakness, hypotonia, joint contractures, white matter abnormalities, peripheral neuropathy and respiratory compromise (Schéele et al. 2007). Besides this, in 2009 a homozygous mutation on this gene was shown to be causative for a phenotype of polymicrogyria with subsequently developed seizures consistent with absence epilepsy (Vigliano et al. 2009).

Since the described brain malformations, such as polymicrogyria, have been reported has being responsible for epilepsy, it would be possible to explain the complex refractory epilepsy presented by the proband as a consequence of these malformations. However, the *CACNB4* gene, located 600kb upstream of the translocation breakpoint on der(2), has been associated with several epilepsy related phenotypes. The spontaneously arose "lethargic" mouse model, which has several mutations on the Ca^{2+} beta subunit (*Cchb4*) gene, an orthologous of *CACNB4* gene, presents a phenotype of ataxia and seizures (Burgess et al. 1997). Additionally, mutations on the *CACNB4* gene are responsible for abnormal neuronal membrane polarization, often causing ataxia and seizures (Segal et al. 2016).

In conclusion, it is possible to state that the identified apparently balanced double translocation - t(2;7)(q23.3;q32.1),t(5;6)(q23.2;q26)dn – is actually a CCR, identifiable only through the application

of large insert whole genome sequencing. At the present moment, liWGS seems to be the only available technology able to fully unravel the spectrum of otherwise cryptic genomic alterations.

Even though it is not possible to exclude the contribution of the deletion to the observed phenotype, this study suggests that the genes affected by the translocations breakpoints are accountable for the reported phenotype. In this way, *PRPF40A* is suggested as the candidate gene for the developmental delay presented by the proband. Regarding the intrauterine growth retardation, the main candidate gene to explain the phenotype is *RBM28*. As for the presented severe brain malformations, the suggested candidate genes are *PACRG*, *PARK2*, *QKI* and *LAMA2*. The epilepsy phenotype presented by the proband can not only be explained by the severe brain malformations associated with the previous stated genes, but also by the involvement of the *CACNB4* gene.

Currently, expression profiling of the proband's LCLs is being carried out for assessing the effect of these chromosomal rearrangements on the expression of the genes from the breakpoint regions.

Furthermore, it would be of great benefit for the further enlightenment of the pathogenicity of this case to be able to analyze samples from the proband's parents. Even though the reported double translocation is *de novo*, it is possible that the deletion found at chr14 may be inherited, not causing any deleterious phenotype, which would sustain our findings.

5.2. ESTABLISHMENT OF AN INDIVIDUAL-SPECIFIC IPSC MODEL

Since it is not usually possible to access relevant biological samples from the proband, an individual specific cellular model would be of tremendous utility in the study of the molecular pathogenesis in the reported CCR.

Pluripotency induction experiments were initiated using proband derived LCLs as the starter cell culture. Due to their wide availability, patients' derived LCLs have been recurrently used for iPSC generation (Barrett et al. 2014; Thomas et al. 2015).

Due to high mortality rates and low electroporation efficiency observed during these experiments, not comparable to either Barrett et al. (2014) or Thomas et al. (2015), it was not possible to induce LCL derived iPSC colonies. However, it was possible to identify the critical steps for the induction and to establish the protocol for preparation and maintenance of the HFF used as feeder cells.

Since the pluripotency induction efficiency itself usually does not surpass 0,01%, high electroporation efficiency is critical. In that way, it was found that using LCL with a lower passage number together with a higher mitotic rate would tremendously benefit the electroporation efficiency.

Even though electroporation of individual-derived cells, especially LCLs, using episomal plasmids is less effective than the use of viral vectors, this methodology is safer because episomal plasmids do not integrate the DNA of induced cells. Concomitantly, a 2015 study shows that iPSC made from LCL using episomal plasmid vectors recover the donor's gene expression profile (Thomas et al. 2015). In this way, this methodology is the most appealing when thinking about the available main starting material and in the future applications.

Differentiation of individual specific iPSC into tissue specific cells will allow to overcome the difficulties in the obtainment of relevant biopsies. Thus, characterization of the molecular pathogenesis of congenital anomalies caused by chromosomal rearrangements will highly benefit from this methodology.

Hereafter, besides optimizing the electroporation with episomal plasmids, other methodologies will be studied. Transfection using Sendai virus is an option to take into account, since it provides a higher transfection efficiency and, consequently, a higher reprogramming efficiency. Additionally, new methodologies on the culture, propagation and induction of human urine cells are becoming more and more popular (Zhou et al. 2014). The easy and non-invasive accessibility of these samples makes this another methodology to take into account when developing individual specific iPSCs.

6. CONCLUSIONS AND FUTURE PERSPECTIVES

Comparative mapping of the t(2;7)(q23.3;q32.1),t(5;6)(q23.2;q26)dn breakpoints using both array painting and liWGS supports the importance of the use of the latter approach on the study of chromosomal structural rearrangements. In fact, the use of the liWGS methodology unraveled a CCR on der(6) involving a 48 kb excision/insertion that disrupts the *PTPRK* gene, and a 1.5 Mb inversion of the 5q23.2 region.

Thus, it is possible to state that, presently, NGS-based approaches are the only available technologies able to unveil the full spectrum of structural genomic alterations.

Regarding the phenotype observed in the proband, the intrauterine growth retardation is most likely explained by the *RBM28* gene, located at 7q32.1. As for the severe developmental delay, current data points to *PRPF40A*, disrupted at 2q23.3, as the most likely candidate gene. Brain malformations observed in the proband, which contribute to both developmental delay and epilepsy, may be explained by the disrupted *PACRG* gene, by *PARK2* and *QKI* genes, located at 6q26, and also by *LAMA2* gene, located 70 kb distal from the excision breakpoint at 6q22.33. Furthermore, *CACNB4* gene, located 600 kb distal from the breakpoint at 2q23.3, allows to further explain the refractory epilepsy observed in the proband.

Even though causal relationship between the reported phenotype and observed genomic alterations was unveiled, the complexity of the clinical phenotype and the number of affected genes did not allow for the full understanding of the molecular pathogenesis of this double *de novo* chromosomal translocation. Thus, in order to further dissect the genotype-phenotype association, the development of a patient specific model is of extreme importance.

During the course of this study, it was not possible to obtain an individual-specific iPSC-based model. However, further development of disease specific models is currently underway.

Optimization of iPSC protocols applicable to the available patient primary cells would be highly beneficial towards the study of the molecular pathogenesis of complex structural chromosomal rearrangements.

Besides electroporation with episomal plasmids, transfection of key reprogramming factors using Sendai virus is another non-integrative approach to take into account. The use of Sendai virus would allow for higher transfection and reprogramming efficiencies, and would be most likely beneficial towards achieving several iPSC-based models.

Likewise, aside from blood-derived primary cells and skin epithelial cells, which obtainment requires an invasive procedure, it is now possible to culture and propagate urine derived cells. These fibroblast-like cells can be obtained by non-invasive aseptic urine collection followed by plating of the

resulting pellet, and were already successfully reprogrammed and differentiated to cardiomyocytes (Zhou et al. 2014).

Additionally, developing a viable animal model that would mimic the disruption of the affected genes would represent an excellent tool for the study of the molecular pathogenesis.

Due to the complexity of this hypothesis, the study of a previous established mouse model, namely the Quaking viable mouse model, which presents a loss of function mutation in the *QKI* gene, significantly reducing its expression and completely depleting *PACRG* expression (Lockhart et al. 2004), could be of value specifically for the understanding of the phenotypical effect of the affected genes on der(6) translocation breakpoint. In this case, the study of the effects on the central nervous system, namely by performing histological sections and expression profiling of the tissue, would probably cast a light on the understanding of genotype-phenotype association.

In conclusion, application of liWGS for mapping complex genomic rearrangements and development of patient-specific models would be of tremendous value in the study of these rearrangements and, specially, in unravelling therapeutical targets that may help chromosomal translocation carriers with severe phenotypes.

7. BIBLIOGRAPHY

Agarwal S, Tafel A, Kanaar R (2006) DNA double-strand break repair and chromosome translocations. DNA Repair 5:1075–81. doi: 10.1016/j.dnarep.2006.05.029

Alberts B, Johnson A, Lewis J, Roberts K, Walter P (2008) Molecular Biology of the Cell. Garland Science, New York

Arts H, Knoers N (2013) Cranioectodermal Dysplasia. NCBI https://www.ncbi.nlm.nih.gov/books/NBK154653/ accessed 16 September 2016

Arts HH, Bongers EMHF, Mans DA, van Beersum SEC, Oud MM, Bolat E, Spruijt L, Cornelissen EAM, Schuurs-Hoeijmakers JHM, de Leeuw N, Cormier-Daire V, Brunner, HG, Knoers NVAM, Roepman R (2011) C14ORF179 encoding IFT43 is mutated in Sensenbrenner syndrome. J Med Genet 48:390–395. doi: 10.1136/jmg.2011.088864

Backx L, Fryns JP, Marcelis C, Devriendt K, Vermeesch J, Van Esch H. (2010) Haploinsufficiency of the gene Quaking (QKI) is associated with the 6q terminal deletion syndrome. Am J Med Genet Part A 152:319–326. doi: 10.1002/ajmg.a.33202

Barrett R, Ornelas L, Yeager N, Mandefro B, Sahabian A, Lenaeus L, Targan SR, Svendsen C, Sareen D. (2014) Reliable Generation of Induced Pluripotent Stem Cells From Human Lymphoblastoid Cell Lines. Stem Cells Trans Med. doi: 10.5966/sctm.2014-0121

Basso C, Corrado D, Marcus FI, Nava A, Thiene G. (2009) Arrhythmogenic right ventricular cardiomyopathy. Lancet 373:1289–1300. doi: 10.1016/S0140-6736(09)60256-7

Bershteyn M, Hayashi Y, Desachy G, Hsiao EC, Sami S, Tsang KM, Weiss LA, Kriegstein AR, Yamanaka S, Wynshaw-Boris A. (2014) Cell-autonomous correction of ring chromosomes in human induced pluripotent stem cells. Nature. doi: 10.1038/nature12923

Bertoli-Avella AM, Gillis E, Morisaki H, Verhagen JMA, De Graaf BM, Van De Beek G, Gallo E, Kruithof BPT, Venselaar H, Myers LA, Laga S, Doyle AJ, Oswald G, Van Cappellen GWA, Yamanaka I, Van Der Helm RM, Beverloo B, De Klein A, Pardo L, Lammens M, Evers C, Devriendt K, Dumoulein M, Timmermans J, Bruggenwirth HT, Verheijen F, Rodrigus I, Baynam G, Kempers M, Saenen J, Van Craenenbroeck EM, Minatoya K, Matsukawa R, Tsukube T, Kubo N, Hofstra R, Goumans MJ, Bekkers JÁ, Roos-Hesselink JW, Van De Laar IMBH, Dietz HC, Van Laer L, Morisaki T, Wessels MW,Loeys BL. (2015) Mutations in a TGF-β ligand, TGFB3, cause syndromic aortic aneurysms and dissections. J Am Coll Cardiol 65:1324–1336. doi: 10.1016/j.jacc.2015.12.056

Brault V, Pereira P, Duchon A, Hérault Y (2006) Modeling chromosomes in mouse to explore the function of genes, genomic disorders, and chromosomal organization. PLoS Genet 2:0911–0919. doi: 10.1371/journal.pgen.0020086

Burgess DL, Jones JM, Meisler MH, Noebels JL (1997) Mutation of the Ca2+ channel b subunit gene Cchb4 is associated with ataxia and seizures in the lethargic (lh) mouse. Cell 88:385–392. doi: 10.1016/S0092-8674(00)81877-2

Chen W, Kalscheuer V, Tzschach A, Menzel C, Ullmann R, Schulz MH, Erdogan F, Li N, Kijas Z, Arkesteijn G, Pajares IL, Goetz-Sothmann M, Heinrich U, Rost I, Dufke A, Grasshoff U, Glaeser

B, Vingron M, Ropers, HH (2008) Mapping translocation breakpoints by next-generation sequencing. Genome Res 18:1143–1149. doi: 10.1101/gr.076166.108

Chen W, Ullmann R, Langnick C, Menzel C, Wotschofsky Z, Hu H, Döring A, Hu Y, Kang H, Tzschach A, Hoeltzenbein M, Neitzel H, Markus S, Wiedersberg E, Kistner G, van Ravenswaaij-Arts CMA, Kleefstra T, Kalscheuer VM, Ropers HH. (2010) Breakpoint analysis of balanced chromosome rearrangements by next-generation paired-end sequencing. Eur J Hum Genet 18:539–543. doi: 10.1038/ejhg.2009.211

Cheung VG, Nowak N, Jang W, Kirsch IR, Zhao S, Chen XN, Furey TS, Kim UJ, Kuo WL, Olivier M, Conroy J, Kasprzyk A, Massa H, Yonescu R, Sait S, Thoreen C, Snijders A, Lemyre E, Bailey JA, Bruzel A, Burrill WD, Clegg SM, Collins S, Dhami P, Friedman C, Han CS, Herrick S, Lee J, Ligon AH, Lowry S, Morley M, Narasimhan S, Osoegawa K, Peng Z, Plajzer-Frick I, Quade BJ, Scott D, Sirotkin K, Thorpe AA, Gray JW, Hudson J, Pinkel D, Ried T, Rowen L, Shen-Ong GL, Strausberg RL, Birney E, Callen DF, Cheng JF, Cox DR, Doggett NA, Carter NP, Eichler EE, Haussler D, Korenberg JR, Morton CC, Albertson D, Schuler G, de Jong PJ, Trask BJ .(2001) Integration of cytogenetic landmarks into the draft sequence of the human genome. Nature 409:953–958. doi: 10.1038/35057192

Clancy BS, Shaw KM (2008) DNA Deletion and Duplication and the Associated Genetic Disorders Clustering of Breakpoints : Recombination Hotspots. Nature Edu 1(1):23.

Coe BP, Witherspoon K, Rosenfeld JA, van Bon BWM, Vulto-van Silfhout AT, Bosco P, Friend KL, Baker C, Buono S, Vissers LELM, Schuurs-Hoeijmakers JH, Hoischen A, Pfundt R, Krumm N, Carvill GL, Li D, Amaral D, Brown N, Lockhart PJ, Scheffer IE, Alberti A, Shaw M, Pettinato R, Tervo R, de Leeuw N, Reijnders MRF, Torchia BS, Peeters H, O'Roak BJ, Fichera M, Hehir-Kwa JY, Shendure J, Mefford HC, Haan E, Gécz J, de Vries BBA, Romano C, Eichler EE. (2014) Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet 46:1063–1071. doi: 10.1038/ng.3092

Collin RWJ, Kalay E, Tariq M, Peters T, van der Zwaag B, Venselaar H, Oostrik J, Lee K, Ahmed ZM, Çaylan R, Li Y, Spierenburg HA, Eyupoglu E, Heister A, Riazuddin S, Bahat E, Ansar M, Arslan S, Wollnik B, Brunner HG, Cremers CWRJ, Karaguzel A, Ahmad W, Cremers FPM, Vriend G, Friedman TB, Riazuddin S, Leal SM, Kremer H (2008) Mutations of ESRRB Encoding Estrogen-Related Receptor Beta Cause Autosomal-Recessive Nonsyndromic Hearing Impairment DFNB35. Am J Hum Genet 82:125–138. doi: 10.1016/j.ajhg.2007.09.008

Cooper GM, Coe BP, Girirajan S, Rosenfeld, Jill A, Vu TH, Baker C, Williams C, Stalker H, Hamid Riz, Hannig V, Abdel-Hamid H, Bader P, McCracken E, Niyazov D, Leppig K, Thiese H, Hummel M, Alexander N, Gorski J, Kussmann J, Shashi V, Johnson K, Rehder C, Ballif BC, Shaffer LG, Eichler EE. (2011) A copy number variation morbidity map of developmental delay. Nat Genet 43:838–46. doi: 10.1038/ng.909

Corsello G, Giuffrè M. (2012) Congenital malformations. J Matern Neonatal Med 25:25–29. doi: 10.3109/14767058.2012.664943

David D, Marques B, Ferreira C, Araújo C, Vieira L, Soares G, Dias C, Pinto M (2013) Co-segregation of trichorhinophalangeal syndrome with a t(8;13)(q23.3;q21.31) familial translocation that appears to increase TRPS1 gene expression. Hum Genet 132:1287–99. doi: 10.1007/s00439-013-1333-0

De Gregori M, Ciccone R, Magini P, Pramparo T, Gimelli S, Messa J, Novara F, Vetro A, Rossi E, Maraschio P, Bonaglia MC, Anichini C, Ferrero GB, Silengo M, Fazzi E, Zatterale A, Fischetto R, Previderé C, Belli S, Turci A, Calabrese G, Bernardi F, Meneghelli E, Riegel M, Rocchi M, Guerneri S, Lalatta F, Zelante L, Romano C, Fichera M, Mattina T, Arrigo G, Zollino M, Giglio S, Lonardo F, Bonfante A, Ferlini A, Cifuentes F, Van Esch H, Backx L, Schinzel A, Vermeesch JR, Zuffardi, O. (2007) Cryptic deletions are a common finding in "balanced" reciprocal and complex chromosome rearrangements: a study of 59 patients. J Med Genet 44:750–62. doi: 10.1136/jmg.2007.052787

Dolk H, Loane M, Garne E (2010) The Prevalance of Congenital Anomalies in Europe. In Paz MPP, Groft SC. Rare Diseases Epidemiology, Advances in Experimental Medicine and Biology, 1st ed. Springer, pp 349–364

Dowey SN, Huang X, Chou B, Cheng L. (2012) Generation of integration-free human induced pluripotent stem cells from postnatal blood mononuclear cells by plasmid vector expression. Nat Protoc 7:2013–2021. doi: 10.1038/nprot.2012.121

Durmaz AA, Karaca E, Demkow U, Toruner G, Schoumans J, Cogulu O (2015) Evolution of genetic techniques: Past, present, and beyond. Biomed Res Int. doi: 10.1155/2015/461524

Ebersole TA, Chen Q, Justice MJ, Artzt K (1996) The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins. Nat Genet 12:260–5. doi: 10.1038/ng0396-260

Escayg A, De Waard M, Lee DD, Bichet D, Wolf P, Mayer T, Johnston J, Baloh R, Sander T, Meisler MH. (2000) Coding and noncoding variation of the human calcium-channel beta4-subunit gene CACNB4 in patients with idiopathic generalized epilepsy and episodic ataxia. Am J Hum Genet 66:1531–9. doi: 10.1086/302909

Fang Q, George AS, Brinkmeier ML, Mortensen AH, Gergics P, Cheung LY, Daly AZ, Ajmal A, Perez Millan MI, Ozel AB, Kitzman JO, Mills RE, Li JZ, Camper SA. (2016) Genetics of Combined Pituitary Hormone Deficiency: Roadmap into the Genome Era. Endocr Rev 37:636–675. doi: 10.1210/er.2016-1101

Fiegler H, Gribble SM, Burford DC, Carr P, Prigmore E, Porter KM, Clegg S, Crolla JA, Dennis NR, Jacobs P, Carter NP. (2003) Array painting: a method for the rapid analysis of aberrant chromosomes using DNA microarrays. J Med Genet 40:664–670. doi: 10.1136/jmg.40.9.664

Ford CE, Jones KW, Polani PE, Almeida JC, Brigs JH. (1959) A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). The Lancet 7075:711-3.

Franklin R, Gosling R (1953) Molecular Configuration in Sodium Thymonucleate. Nature 171:740–741.

Freed EF, Bleichert F, Dutca LM, Baserga SJ (2010) When ribosomes go bad: diseases of ribosome biogenesis. Mol Biosyst 6:481–93. doi: 10.1039/b919670f

Fujie. Y, Fusaki N, Katayama T, Hamasaki M, Soejima Y, Soga M, Ban H, Hasegawa M, Yamashita S, Kimura S, Suzuki S, Matsuzawa T, Akari H, Era T. (2014) New type of sendai virus vector provides transgene-free iPS cells derived from chimpanzee blood. PLoS One 9:1–19. doi: 10.1371/journal.pone.0113052

Fusaki N, Ban H, Nishiyama A, Saeki K, Hasegawa M (2009) Efficient induction of transgene-free

human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. Proc Jpn Acad Ser B Phys Biol Sci 85:348–62. doi: 10.2183/pjab.85.348

Gardner RM, Sutherland GR, Shaffer L (2012) Chromosome Abnormalities and Genetic Counseling, 4th ed. Oxford University Press, New York

Griffiths A, Wessler S, Lewontin R, Carrol S (2007) An Introduction to Genetic Analysis, 9ª. W.H. Freeman and Company

Guo L, Zhang X, Kuang W, Xu H, Liu LA, Vilquin JT, Miyagoe-Suzuki Y, Takeda S, Ruegg MA, Wewer UM, Engvall, E. (2003) Laminin a2 deficiency and muscular dystrophy; genotype-phenotype correlation in mutant mice. Neuromuscul Disord 13:207–2015. doi: 10.1016/S0

Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res 33:514–517. doi: 10.1093/nar/gki033

Han W, Zhao Y, Fu X (2010) Induced Pluripotent Stem Cells: The Dragon Awakens. Bioscience 60:278–285. doi: 10.1525/bio.2010.60.4.6

Hanscom C, Talkowski M (2014) Design of large-insert jumping libraries for structural variant detection using Illumina sequencing. Curr Protoc Hum Genet 7.22.1-7.22.9. doi: 10.1002/0471142905.hg0722s80

Higgins AW, Alkuraya FS, Bosco AF, Brown KK, Bruns GAP, Donovan DJ, Eisenman R, Fan Y, Farra CG, Ferguson HL, Gusella JF, Harris DJ, Herrick SR, Kelly C, Kim H, Kishikawa S, Korf BR, Kulkarni S, Lally E, Leach NT, Lemyre E, Lewis J, Ligon AH, Lu W, Maas RL, Macdonald ME, Moore SDP, Peters RE, Quade BJ, Quintero-Rivera F, Saadi I, Shen Y, Shendure J, Williamson RE, Morton CC (2008) Characterization of Apparently Balanced Chromosomal Rearrangements from the Developmental Genome Anatomy Project. Am. J. Hum. Genet. 82:712–722. doi: 10.1016/j.ajhg.2008.01.011.

Houge G, Liehr T, Schoumans J, Ness GO, Solland K, Starke H, Claussen U, Stromme P, Akre B, Vermeulen S. (2003) Ten years follow up of a boy with a complex chromosomal rearrangement: going from a > 5 to 15-breakpoint CCR. Am J Med Genet A 118a:235–240. doi: 10.1002/ajmg.a.10106

Hussain T, Mulherkar R (2012) Lymphoblastoid Cell lines: a Continuous in Vitro Source of Cells to Study Carcinogen Sensitivity and DNA Repair. Int J Mol Cell Med 1:75–87.

Jacobs PA, Strong JA (1959) A Case of Human Intersexuality Having a Possible XXY Sex-Determining Mechanism. Nature 183:302–303. doi: 10.1038/183302a0

Jiang J, Zhang L, Zhou X, Chen X, Huang G, Li F, Wang R, Wu N, Yan Y, Tong C, Srivastava S, Wang Y, Liu H, Ying Q. (2016) Induction of site-specific chromosomal translocations in embryonic stem cells by CRISPR/Cas9. Sci Rep 6:21918. doi: 10.1038/srep21918

Kallioniemi A, Kallioniemi O, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. (1992) Comparative Genomic Hynridization for Molecular Cytogenetic Analysis of Dolid Tumors. Science 258:818–820. doi: 10.1016/B978-0-12-383834-6.00029-X

Kato Y, Miyakawa T, Kurita JI, Tanokura M (2006) Structure of FBP11 WW1-PL ligand complex reveals the mechanism of proline-rich ligand recognition by Group II/III WW domains. J Biol

Chem 281:40321–40329. doi: 10.1074/jbc.M609321200

Kaufman L, Ayub M, Vincent JB (2010) The genetic basis of non-syndromic intellectual disability: A review. J Neurodev Disord 2:182–209. doi: 10.1007/s11689-010-9055-2

Kay DM, Stevens CF, Hamza TH, Montimurro JS, Zabetian CP, Factor SA, Samii A, Griffith A, Roberts JW, Molho ES, Higgins DS, Gancher S, Moses L, Zareparsi S, Poorkaj P, Bird T, Nutt J, Schellenberg GD, Payami H. (2010) A comprehensive analysis of deletions, multiplications, and copy number variations in PARK2. Neurology 75:1189–1194. doi: 10.1212/WNL.0b013e3181f4d832

Kent WJ (2002) BLAT — The BLAST -Like Alignment Tool. Genome Res 12:656–664. doi: 10.1101/gr.229202.

Kent WJ, Sugnet CW, Furey TS, Roskin KM (2002) The Human Genome Browser at UCSC. J Med Chem 19:1228–31. doi: 10.1101/gr.229102.

Kirkpatrick M (2010) How and why chromosome inversions evolve. PLoS Biol. doi: 10.1371/journal.pbio.1000501

Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, Yokochi M, Mizuno Y, Shimizu N. (1998) Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature 392:605–608. doi: 10.1038/33416

Kleinjan D-J, Coutinho P (2009) Cis-ruption mechanisms: disruption of cis-regulatory control as a cause of human genetic disease. Brief Funct Genomic Proteomic 8:317–32. doi: 10.1093/bfgp/elp022

Kloosterman WP, Guryev V, Van Roosmalen M, Duran KJ, De Bruijn E, Bakker SCM, Letteboer T, Nesselrooij BV, Hochstenbach R, Poot M, Cuppen E. (2011) Chromothripsis as a mechanism driving complex de novo structural rearrangements in the germline. Hum. Mol. Gen 20:1916–1924. doi: 10.1093/hmg/ddr073

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

Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders EAC, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M. (2007) Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome. Science 318:420–426. doi: 10.1126/science.1149504.Paired-End

Langer-Safer PR, Levine M, Ward DC (1982) Immunological method for mapping genes on Drosophila polytene chromosomes. Proc Natl Acad Sci USA 79:4381–5. doi: 10.1073/pnas.79.14.4381

Lassar AB, Paterson BM, Weintraub H (1988) Transfection of a DNA Locus That Mediates the Conversion of lOTV2 Fibroblasts to Myoblasts. 47:649–656.

Lejeune J, Gautier M, Turpin R. (1959) Ètude des chromosomes somatiques de neuf enfant mongoliens. Comptes Rendus 248:1721–1722.

Lesage S, Magali P, Lohmann E, Lacombez L, Teive H, Janin S, Cousin P-Y, Dür A, Brice A. (2007) Deletion of the parkin and PACRG Gene Promoter in Early-Onset Prakinsonism. Hum Mutat 28:27–32. doi: 10.1002/humu.20436

Liao Y, Wang L, Zhang D, Liu C (2014) Identification of a balanced complex chromosomal rearrangement involving chromosomes 3, 18 and 21 with recurrent abortion: case report. Mol Cytogenet 7:39. doi: 10.1186/1755-8166-7-39

Lizier N, Kerkis I, Wenceslau C (2013) Generation of Induced Pluripotent Stem Cells from Dental Pulp Somatic Cells. Intech 7: 131-149 doi.org/10.5772/55856

Lockhart PJ, O'Farrell CA, Farrer MJ (2004) It's a double knock-out! The quaking mouse is a spontaneous deletion of parkin and parkin co-regulated gene (PACRG). Mov Disord 19:101–104. doi: 10.1002/mds.20000

Luthardt FW, Keitges E (2001) Chromosomal Syndromes and Genetic Disease. In Enciclopedia of Life Sciences. Nature Pub Group pp 1–12

Mack AA, Kroboth S, Rajesh D, Wang WB (2011) Generation of Induced Pluripotent Stem Cells from CD34 + Cells across Blood Drawn from Multiple Donors with Non-Integrating Episomal Vectors. PLoS 11:1-14. doi: 10.1371/journal.pone.0027956

Manzini S, Viiri LE, Marttila S, Aalto-Setälä K (2015) A Comparative View on Easy to Deploy non-Integrating Methods for Patient-Specific iPSC Production. Stem Cell Rev Reports 11:900–908. doi: 10.1007/s12015-015-9619-3

Meng X, Su R, Baylink DJ, Neises A, Kiroyan JB, Lee WY, Payne KJ, Gridley DS, Wang J, Lau KH, Li G, Zhang XB. (2013) Rapid and efficient reprogramming of human fetal and adult blood CD34 + cells into mesenchymal stem cells with a single factor. Nat Publ Gr 23:658–672. doi: 10.1038/cr.2013.40

Merling RK, Sweeney CL, Choi U, Ravin SSD, Myers TG, Otaizo-Carrasquero F, Pan J, Linton G, Chen L, Koontz S, Theobald NL, Malech HL. (2016) Transgene-free iPSCs generated from small volume peripheral blood. Hemat Stem Cell 121:98–108. doi: 10.1182/blood-2012-03-420273.R.K.M.

Metzker ML (2010) Sequencing technologies — the next generation. Nat Rev Genet 11:31–46. doi: 10.1038/nrg2626

Miller G (1982) Immortalization of Human Lymphocytes by Epstein-Barr Virus. Yale J Biol Med 55:305–310.

Moorhead PS, Nowell PC, Mellman WJ, Battips DM, Hungerford DA. (1960) Chromosome preparations of leukocytes cultured from human peripheral blood. Exp Cell Res 20:613–616. doi: 10.1016/0014-4827(60)90138-5

Neitzel H (1986) A routine method for the establishment of permanent growing lymphoblastoid cell lines. Hum Genet 73:320–6.

Nelson DL, Cox MM (2008) Principles of Biochemistry, 5th ed. W H Freeman & Co, US

Nousbeck J, Spiegel R, Ishida-Yamamoto A, Indelman M, Shani-Adir A, Adir N, Lipkin E, Bercovici S, Geiger D, van Steensel MA, Steijlen PM, Bergman R, Bindereif A, Choder M, Shalev S, Sprecher E. (2008) Alopecia, Neurological Defects, and Endocrinopathy Syndrome Caused by

Decreased Expression of RBM28, a Nucleolar Protein Associated with Ribosome Biogenesis. Am J Hum Genet 82:1114–1121. doi: 10.1016/j.ajhg.2008.03.014

Okita K, Yamakawa T, Matsumura Y, Sato Y, Amano N, Watanabe A, Goshima N, Yamanaka S. (2013) An efficient nonviral method to generate integration-free human-induced pluripotent stem cells from cord blood and peripheral blood cells. Stem Cells 31:458–66. doi: 10.1002/stem.1293

Ordulu Z, Kammin T, Brand H, Pillalamarri V, Redin CE, Collins RL, Blumenthal I, Hanscom C, Pereira S, Crandall BF, Gerrol P, Hayden MA, Hussain N, Kanengisser-Pines B, Kantarci S, Levy B, Macera MJ, Quintero-Rivera F, Spiegel E, Stevens B, Ulm JE, Warburton D, Wilkins-Haug LE, Yachelevich N, Gusella JF, Talkowski M(2016) Structural Chromosomal Rearrangements Require Nucleotide-Level Resolution : Lessons from Next-Generation Sequencing in Prenatal Diagnosis. Am J Hum Genet 99:1–19. doi: 10.1016/j.ajhg.2016.08.022

Pai GS, Thomas GH, Mahoney W, Migeon BR (1980) Complex chromosome rearrangements. Report of a new case and literature review. Clin Genet 18:436-444

Pellestor F, Anahory T, Lefort G, Puechberty J, Liehr T, Sarda P, He B. (2011) Complex chromosomal rearrangements : origin and meiotic behavior. Hum Reprod Update 17:476–494. doi: 10.1093/humupd/dmr010

Pinkel D, Segraves R, Sudar D, Clark, S, Poole I, Kowbel D, Collins C, Kuo W L, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG. (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 20:207–211. doi: 10.1038/2524

Potter M (2008) Brief historical sketch of chromosomal translocations and tumors. J Natl Cancer Inst - Monogr 20892:2–7. doi: 10.1093/jncimonographs/lgn013

Quelin C, Saillour Y, Poirier K, Roubertie A, Boddaert N, Desguerre I, Letourneur F, Beldjord C, Chelly J, Bahi-Buisson N. (2012) Focal polymicrogyria are associated with submicroscopic chromosomal rearrangements detected by CGH microarray analysis. Eur J Med Genet 55:527–530. doi: 10.1016/j.ejmg.2012.06.004

Ramos-Mejía V, Montes R, Bueno C, Ayllón V, Real PJ, Rodríguez R, Menendez P (2012) Residual expression of the reprogramming factors prevents differentiation of iPSC generated from human fibroblasts and cord blood CD34+ progenitors. PLoS One. doi: 10.1371/journal.pone.0035824

Riegel M (2014) Human molecular cytogenetics: From cells to nucleotides. Genet Mol Biol 37:194–209.

Rienhoff HY, Yeo CY, Morissette R, Khrebtukova I, Melnick J, Luo S, Leng N, Kim YJ, Schroth G,Westwick J, Vogel H, Mcdonnell N, Hall JG, Whitman M. (2013) A mutation in TGFB3 associated with a syndrome of low muscle mass, growth retardation, distal arthrogryposis and clinical features overlapping with marfan and loeys-dietz syndrome. Am J Med Genet Part A 161:2040–2046. doi: 10.1002/ajmg.a.36056

Robinton D, Daley GQ (2012) The promise of induced pluripotent stem cells in research and therapy. Nature 481:295–305. doi: 10.1038/nature10761

Roukos V, Misteli T (2014) The biogenesis of chromosome translocations. Nat Cell Biol 16:293–300. doi: 10.1038/ncb2941

Roux KH (2009) Optimization and Troubleshooting in PCR. Cold Spring Harb Protoc 4:1–7. doi:

10.1101/pdb.ip66

Rowley JD (1973) Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature 243:290–3. doi: 10.1038/243290a0

Samanta J, Bonaguidi MA, Kessler JA (2008) The TGF-b Family in Neural and Neuronal Differentiation and Development. In: Derynck R, Miyazono K, TGF-b Fam., 1st ed. Cold Spring Harbor Press, New York, pp 819–860

Schluth-Bolard C, Labalme A, Cordier M-PP, Till M, Nadeau G, Tevissen H, Lesca G, Boutry-Kryza N, Rossignol S, Rocas D, Dubruc E, Edery P, Sanlaville D. (2013) Breakpoint mapping by next generation sequencing reveals causative gene disruption in patients carrying apparently balanced chromosome rearrangements with intellectual deficiency and/or congenital malformations. J Med Genet 50:144–150. doi: 10.1136/jmedgenet-2012-101351

Scouarnec S Le, Gribble SM (2011) Characterising chromosome rearrangements : recent technical advances in molecular cytogenetics. J Hered 108:75–85. doi: 10.1038/hdy.2011.100

Seabright M (1971) a Rapid Banding Technique for Human Chromosomes. Lancet 298:971–972. doi: 10.1016/S0140-6736(71)90287-X

Segal E, Pedro H, Valdez-Gonzalez K, Parisotto S, Gliksman F, Thompson S, Sabri J, Fertig E. (2016) Diagnostic Yield of Epilepsy Panels in Children With Medication-Refractory Epilepsy. Pediatr Neurol 64:66–71. doi: 10.1016/j.pediatrneurol.2016.06.019

Schéele S, Nyström A, Durbeej M, Talts JF, Ekblom M, Ekblom P. (2007) Laminin isoforms in development and disease. J Mol Med 85:825–836. doi: 10.1007/s00109-007-0182-5

Solinas-Toldo S, Lampel S, Stilgenbauer S, Nickolenko J, Benner A, Döhner H, Cremer T, Lichter P. (1997) Matrix-based comparative genomic hybridization: Biochips to screen for genomic imbalances. Genes Chromosom Cancer 20:399–407. doi: 10.1002/(SICI)1098-2264(199712)20:4<399::AID-GCC12>3.0.CO;2-I

Stadtfeld M, Hochedlinger K (2010) Induced pluripotency : history , mechanisms , and applications. Cold Spring Harb Lab Pres 24:2239–2263. doi: 10.1101/gad.1963910.

Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan-Golan Y, Kohn A, Rappaport N, Safran M, Lancet D.. (2016) The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr Protoc Bioinforma 54:1.30.1-1.30.33. doi: 10.1002/cpbi.5

Stokman MF, Oud MM, van Binsbergen E, Slaats GG, Nicolaou N, Renkema KY, Nijman, IJ, Roepman R, Giles RH, Arts HH, Knoers NVAM, van Haelst MM. (2016) De novo 14q24.2q24.3 microdeletion including IFT43 is associated with intellectual disability, skeletal anomalies, cardiac anomalies, and myopia. Am J Med Genet Part A 170:1566–1569. doi: 10.1002/ajmg.a.37598

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K (2007) Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell 131:861–872. doi: 10.1016/j.cell.2007.11.019

Takahashi K, Yamanaka S (2006) Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell 126:663–676. doi: 10.1016/j.cell.2006.07.024

Talkowski ME, Ernst C, Heilbut A, Chiang C, Hanscom C, Lindgren A, Kirby A, Liu S, Muddukrishna B, Ohsumi TK, Shen Y, Borowsky M, Daly MJ, Morton CC, Gusella JF. (2011) Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. Am J Hum Genet 88:469–81. doi: 10.1016/j.ajhg.2011.03.013

Talkowski ME, Rosenfeld JA, Blumenthal I, Heilbut A, Chiang C, Hanscom C, Lindgren A. (2012) Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. 149:525–537. doi: 10.1016/j.cell.2012.03.028.Sequencing

Thomas SM, Kagan C, Pavlovic BJ, Burnett J, Patterson K, Pritchard JK, Gilad Y. (2015) Reprogramming LCLs to iPSCs Results in Recovery of Donor-Specific Gene Expression Signature. PLoS Genet 11:e1005216. doi: 10.1371/journal.pgen.1005216

Thomson JA, Itskovitz-eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. Embryonic Stem Cell Lines Derived from Human Blastocysts. Science 282:1145-1147.

Tjio JH, Levan A (1956) The Chromosome Number of Man. Hereditas 42:1–6. doi: 10.1111/j.1601-5223.1956.tb03010.x

Utami KH, Hillmer AM, Aksoy I, Chew EGY, Teo ASM, Zhang Z, Lee CWH, Chen PJ, Seng CC, Ariyaratne PN, Rouam SL, Soo LS, Yousoof S, Prokudin I, Peters G, Collins F, Wilson M, Kakakios A, Haddad G, Menuet A, Perche O, Tay SKH, Sung KWK, Ruan X, Ruan Y, Liu ET, Briault S, Jamieson RV, Davila S, Cacheux V. (2014) Detection of chromosomal breakpoints in patients with developmental delay and speech disorders. PLoS One 9:e90852. doi: 10.1371/journal.pone.0090852

Vigliano P, Dassi P, Blasi CD, Mora M, Jarre L. (2009) LAMA2 stop-codon mutation: Merosin-deficient congenital muscular dystrophy with occipital polymicrogyria, epilepsy and psychomotor regression. Eur J Paediatr Neurol 13:72–76. doi: 10.1016/j.ejpn.2008.01.010

Vorsanova SG, Yurov YB, Iourov IY (2010) Human interphase chromosomes: a review of available molecular cytogenetic technologies. Mol Cytogenet 3:1. doi: 10.1186/1755-8166-3-1

Wang L, Wang L, Huang W, Su H, Xue Y, Su Z, Liao B, Wang H, Bao X, Qin D, He J, Wu W, So KF. (2013) Generation of integration-free neural progenitor cells from cells in human urine. Nature Meth 10:84–89. doi:10.1038/nmeth.2283

Watson JD, Crick FHC (1953) Molecular structure of nucleic acids. Nature 171:737–738. doi: 10.1097/BLO.0b013e3181468780

Weber ML, Hsin H-Y, Kalay E, BroŽková, Dana S, Shimizu T, Bayram M, Deeley K, Küchler EC, Forella J, Ruff TD, Trombetta VM, Sencak RC, Hummel M, Briseño-Ruiz J, Revu SK, Granjeiro JM, Antunes LS, Antunes LA, Abreu FV, Costa MC, Tannure PN, Koruyucu M, Patir A, Poletta FA, Mereb JC, Castilla EE, Orioli IM, Marazita ML, Ouyang H, Jayaraman T, Seymen F, Vieira AR. (2014) Role of estrogen related receptor beta (ESRRB) in DFN35B hearing impairment and dental decay. BMC Med Genet 15:81. doi: 10.1186/1471-2350-15-81

Wellesley D, Dolk H, Boyd PA, Greenlees R, Haeusler M, Nelen V, Garne E, Khoshnood B, Doray B, Rissmann A, Mullaney C, Calzolari E, Bakker M, Salvador J, Addor MC, Draper E, Rankin J. (2012) Rare chromosome abnormalities , prevalence and prenatal diagnosis rates from population-based congenital anomaly registers in Europe. Eur. J. Hum. Genet 20:521–526. doi:

10.1038/ejhg.2011.246

West AB, Lockhart PJ, O'Farell C, Farrer MJ (2003) Identification of a novel gene linked to parkin via a bi-directional promoter. J Mol Biol 326:11–19. doi: 10.1016/S0022-2836(02)01376-1

Weyden L Van Der, Bradley A (2008) Mouse Chromosome Engineering for Modeling Human Disease. Annu Rev Genomics Hum Genet 7:247–276. doi: 10.1146/annurev.genom.7.080505.115741.Mouse

WHO (2016) Congenital Anomalies. http://www.who.int/mediacentre/factsheets/fs370/en/. Accessed 26 Dec 2016

Wilson GR, Wang HX, Egan GF, Robinson PJ, Delatycki MB, O'Bryan MK, Lockhart PJ. (2010) Deletion of the Parkin co-regulated gene causes defects in ependymal ciliary motility and hydrocephalus in the quakingviable mutant mouse. Hum Mol Genet 19:1593–1602. doi: 10.1093/hmg/ddq031

Yamanaka S (2012) Induced pluripotent stem cells: past, present, and future. Cell Stem Cell 10:678–84. doi: 10.1016/j.stem.2012.05.005

Yunis JJ (1976) High resolution of human chromosomes. Science 191:1268–1270.

Zhang X-B (2013) Cellular reprogramming of human peripheral blood cells. Genomics Proteomics Bioinformatics 11:264–74. doi: 10.1016/j.gpb.2013.09.001

Zhou T, Benda C, Duzinger S, Huang Y, Li X, Li Y, Guo X, Cao G, Chen S, Hao L, Chan YC, Ng KM, Ho JC, Wieser M, Wu J, Redl H, Tse HF, Grillari J, Grillari-Voglauer R, Pei D, Esteban MA. (2011) Generation of induced pluripotent stem cells from urine. J Am Soc Nephrol 22:1221–1228. doi: 10.1681/ASN.2011010106

Zhou YF, Yao JL, Yang XJ, Li HX, Han LH, Jiang, Wen-Ping. (2014) Induced pluripotent stem cells reprogrammed from human urine cells as a new strategy for cardiomyocytes. Exp. Clin. Cardiol. 20:721–737.

Zhu J, Cifuentes H, Reynolds J, Lamba DA (2016) Generation of retinal photoreceptors from cGMP-Manufactured Human IPSC line. ARVO Paper Session

8. ANNEX

8.1. LCL ELECTROPORATION AND INDUCTION

Adapted from Thomas et al., *Reprogramming LCLs to iPSCs Results in Recovery of Donor-Specific Gene Expression Signature,* PLOS Genetics, 2015

Before starting

1. Maintain LCLs in an incubator at 37ºC with 5% CO_2 in complete RPMI 1640**.** Make sure you have enough cells, you will need 0,2-1x10^6 cells for each assay using the 16-well nucleofector strips and 1-5x10^6 cells for each assay using the nucleofector cuvettes.
2. Prepare 12 well plates with **1mL of Reprogramming Medium** and equilibrate in a CO_2 incubator at 37ºC**:**

Reagents	Final Concentration
Knockout DMEM/F12 (Gibco, #12660-012)	-
Knockout Serum Replacement	20%
MEM Non Essential Aminoacids (Gibco, #11140-050)	1%
L-Glutamin (Gibco, #25030-081)	1%
Penicilin-Streptomycin (Gibco #15140-148)	0,5%
L-ascorbic acid 2-phosphate sesquimagnesium salt hydrate (Sigma Aldrich, #A8960)	50 µg/mL
Add only before use!	
bFGF (stock 25 µg/mL) (Gibco, #13256-029)	12,5µg/mL
Sodium Butyrate (stock 500mM) (Sigma Aldrich, #B5887)	0,5 mM
B-mercaptoethanol (stock 14,3M) (Sigma Aldrich, #M3148)	0,1 mM

3. Pre-warm an aliquot of reprogramming medium in order to add 400µL to each cuvette at the end of the transfection.
4. Program the 4D-Nucleofector™ System (Lonza, AAF-1002B and AAF-1002X) with the appropriate conditions.
5. Prepare the DNA. The total DNA volume must never surpass more than 10% of total reaction volume. When using Lonza's pmaxGFP vector always use 0,4µg.

LCL Reprogramming

Day -2:

Split the LCL cell line you are going to reprogram into two T25 flasks containing 5-10mL per flask at a density of 0,35x10^6 cell/mL.

Day 0:

1. Count the cells with a Neubauer chamber, you should have a density between 0,8 and 1,2x10^6 cells/mL. Make sure the cells have at least 65% viability;
2. Centrifuge between 2 to 4x10^6 cells in a 15mL tube at 1000rpm for 5 minutes;
3. Aspirate all the medium and resuspend the pellet in 10mL of PBS;
4. Centrifuge at 1000rpm for 5 minutes;
5. Aspirate the medium completely and resuspend the pellet in 100µL of Nucleofector solution made with:
    a. 82µL of Nucleofector solution SF
    b. 18µL of Supplement
6. Add to the ressuspended cells **2µg** of each of the following four plasmids:
    a. **pEP4 E02S ET2K (Addgene, #20927)**
    17522bp, expresses Oct4, Sox2, SV40LT and KLF4
    b. **pCXLE-hOCT3/4-shp53-F (Addgene #27077)**
    11681bp, expresses Oct3/4 and shRNA against p53
    c. **pCXLE-hUL (Addgene, #27080)**
    12051bp, expresses L-Myc and Lyn28
    d. **pCXLE-hSK (Addgene, #27078)**
    12693bp, expresses Sox2 and KLF4
7. Transfer the mixture to the Nucleofector cuvette and place it in the appropriate support in the Nucleofector 4D X unit;
8. Nucleofect the sample with the programof choice.
9. When the program finishes remove the Nucleofector cuvette and incubate 10 minutes at room temperature;
10. Resuspend the cells in 400µL of **Reprogramming Medium** and transfer to the pre-incubated 12 well plate.

Day 2:

1. Using a 1mL pipette, transfer the cells to a 15mL conical tube and centrifuge at 1000rpm for 5 minutes;
2. Add 0,5mL of fresh Reprogramming Medium to the used wells on the 12 well plate to maintain the cells that may have adhered and add another 0,5mL of Reprogramming medium to a new well of a 12-well plate;
3. Carefully aspirate the medium of the pelleted cells, leaving around 100µL to avoid losing any cells;
4. Flick the tube to resuspend the cells and add 1mL of Reprogramming Medium;

5. Transfer 0,5mL of the cell suspension to the pre-used well and the other 0,5mL to the new well you prepared.

Days 4 and 6:

1. Using a 1mL pipette, transfer the cells to a 15mL conical tube and centrifuge at 1000rpm for 5 minutes;
2. Add 0,5mL of fresh Reprogramming Medium to the used wells on the 12 well plate to maintain the cells that may have adhered;
3. Carefully aspirate the medium of the pelleted cells, leaving around 100µL to avoid losing any cells;
4. Flick the tube to resuspend the cells and add 0,5mL of Reprogramming Medium;
5. Transfer the cell suspension to the pre-used wells.

Day 7:

1. Flash-thaw an aliquot of inactivated HFF by submerging it in a 37ºC water bath;
2. Transfer the cells to a 15mL tube and fill it with **IMDM + 10% FBS**;
3. Centrifuge at 1400rpm for 10 minutes;
4. Discard the supernatant by inverting and resuspend the pellet in order to count the cells with a Neubauer chamber;
5. Plate the cells at a density of 20000 irradiated HFF per cm^2 (190000 cells per well of a 6-well plate).

Allow the plate to incubate at least 24 hours at 37ºC in order for the cells to adhere.

Day 8:

1. Aspirate HFF medium from the 6-well plate and wash once with 1mL of PBS;
2. Add 2mL of Reprogramming medium and incubate at 37ºC;
3. Transfer the cells from the two corresponding wells of the 12-well plate to one 15mL tube and centrifuge at 1000rpm for 5 minutes;
4. Count the cells with a Neubauer chamber;
5. Calculate the volume of your cell suspension needed to have $0,2x10^6$ cells and transfer it into a new 15mL tube;
6. Bring the volume up to 2mL and divide into two 1,5mL microtubes in order to extract DNA and RNA.
7. Plate the remaining cells at 9000, 18000 and 36000 cells per well on the 6-well plate with the inactivated HFF.

Day 10:

1. Observe under the microscope to see if your cells have attached. If a significant number of cells are attached to the HFFs aspirate the old medium and replace it with fresh Reprogramming medium.

   If the cells are not attached, transfer the old medium to a 15mL tube and add 2mL of fresh medium to the plate. Then centrifuge the tube at 1000rpm for 5 minutes and resuspend the cells in 2mL of fresh Reprogramming medium. Aspirate the medium from the plate and return the cells to the wells with the HFFs.

Day 12 to day 30+:

1. Change medium in the HFF plates every two days. From this moment on stop adding sodium butyrate to the medium!
2. Observe the cell everyday to check for colonies, they should appear between days 16 to 18.
3. When a colony appears prepare a HFF plate in order to make the first colony picking (usually around day 24), if the colony overgrows the cells will start to differentiate;
4. Make three passages before choosing the best clones to freeze.