UNIVERSIDADE D
COIMBRA

Rute Luísa Cabrita Pinto

INTERNSHIP REPORT AT THE *GIANNINA GASLINI* INSTITUTE

MASTER'S DEGREE IN LABORATORY CLINICAL GENETICS

Curricular Internship Report in the context of the Master's Degree in Laboratory Clinical Genetics, oriented by Dr. Domenico Coviello M.D. Ph.D. and Prof. Isabel Maria Marques Carreira, M.D. Ph.D., presented to the Faculty of Medicine of the University of Coimbra

August of 2022

1 2 9 0

UNIVERSIDADE D
COIMBRA

ISTITUTO GIANNINA GASLINI

ISTITUTO PEDIATRICO
DI RICOVERO E CURA
A CARATTERE SCIENTIFICO

Rute Luísa Cabrita Pinto

Curricular Internship Report in the context of the Master's Degree in Laboratory Clinical Genetics, oriented by Dr. Domenico Coviello M.D. Ph.D. and Prof. Isabel Maria Marques Carreira, M.D. Ph.D., presented to the Faculty of Medicine of the University of Coimbra

September of 2022

Acknowledgments

First of all, I want to express my gratitude to my family for always encouraging and supporting me to pursue my academic and personal ambitions, regardless of the situation. In particular, to my brothers, my mum, and my dad for always bringing me peace when things don't happen as I planned.

To my mentor, Dr. Domenico Coviello (Director of the Human Genetics Laboratory), for his kind welcome and support in helping me advance in the genetics profession throughout my internship, whether in diagnosis or in research.

To my professor Doctor Isabel Carreira, mentor, and coordinator of the Laboratory Clinical Genetics Master's degree, for her kind support and availability in answering all my questions.

To my colleague and friend, Joana Silva, for the cooperation and strength given since the moment we both arrived to Italy.

And, finally, to the ERASMUS+ program that made this internship possible.

Resume

The elaboration of the following report aims to describe the developed activities during the curricular internship for the conclusion of the second year of the master's degree in Laboratory Clinical Genetics, at the University of Coimbra's Faculty of Medicine. Thanks to the funding from the ERASMUS+ program, the internship took place in the Human Genetics Laboratory at the *Giannina Gaslini* Institute in Genova, Italy.

This report includes a general approach to all activities developed in the facility during the eleven months (October to August) of the internship. The first chapter is focused on the introduction to the Human Genetics Laboratory. The second chapter is focused on the work developed in the research field, where I had the opportunity to develop my laboratory skills in the project that I was integrated into, as well as to participate in the writing of an article. The third chapter aims to describe the genetic diagnostic activities in the laboratory. These are divided into three sections: Molecular Genetics, Molecular Cytogenetics, and Conventional Cytogenetics (pre- and postnatal). For molecular genetics, an introduction to some genetic diseases was made, together with the application of diverse laboratory techniques for its diagnosis in suspected patients, as well as the discussion of clinical cases. In molecular cytogenetics, I witnessed the execution of the NIPS protocol and the importance of the NGS technology in it. In conventional cytogenetics, I performed a karyotype on peripheral blood and watched the set-up of cellular cultures for prenatal diagnosis. The fourth chapter, which is the last, discusses the purpose and activities of the Human Genetics Laboratory's Genetic Biobank.

Keywords: Human Genetics; Diagnosis of Genetic Diseases; Gene Expression; Sotos Syndrome.

Resumo

A elaboração do seguinte relatório visa descrever as actividades desenvolvidas durante o estágio curricular para a conclusão do segundo ano do mestrado em Genética Clínica Laboratorial, na Faculdade de Medicina da Universidade de Coimbra. Graças ao financiamento do programa ERASMUS+, o estágio teve lugar no Laboratório de Genética Humana do Instituto *Giannina Gaslini*, em Génova, Itália.

Este relatório inclui uma abordagem geral de todas as actividades desenvolvidas nas instalações durante os onze meses (de Outubro a Agosto) de estágio. O primeiro capítulo centra-se na introdução ao Laboratório de Genética Humana. O segundo capítulo centra-se no trabalho desenvolvido no campo da investigação, onde tive a oportunidade de desenvolver as minhas competências laboratoriais no projecto em que fui integrada, bem como participar na redacção de um artigo. O terceiro capítulo visa descrever as actividades de diagnóstico genético no laboratório. Estas estão divididas em três partes: Genética Molecular, Citogenética Molecular, e Citogenética Convencional (em pré e pós-natal). No diagnóstico em genética molecular, foi feita uma introdução a algumas doenças genéticas, juntamente com a aplicação de diversas técnicas laboratoriais para o seu diagnóstico em doentes, e a discussão de casos clínicos. Na citogenética molecular, testemunhei a execução do protocolo NIPS e a importância da tecnologia NGS no mesmo. Na citogenética convencional, realizei um cariótipo sobre o sangue periférico e assisti ao estabelecimento de culturas celulares para o diagnóstico em pós-natal. O quarto capítulo, que é o último, discute o objectivo e actividades do Biobanco Genético do Laboratório de Genética Humana.

Palavras-chave: Genética humana; Diagnóstico de Doenças Genéticas; Expressão Génica; Síndrome de Sotos.

Contents

Figures

Tables

List of Abbreviations and Acronyms

ACMG – America College of Medical Genetics

CAVD - Congenital Absence of the Vas Deferens

CBC - Complete Blood Count

CF - Cystic Fibrosis

cfDNA – cell-free DNA

CFTR - Cystic Fibrosis Transmembrane Conductance Regulator

CGH - Comparative Genomic Hybridization

CNVs - Copy Number Variations

cRNA - complementary RNA

Ct - Cycle threshold

dbSNP – database of single nucleotide polymorphisms

ddNTPs - dideoxynucleotides

DMSO - Dimethyl sulfoxide

dNTPs - deoxyribonucleotides

dsDNA – double-strand DNA

FISH - Fluorescent in situ hybridization

FMRP - Fragile X Mental Retardation Protein

FXPOI - Fragile X-associated Primary Ovarian Insufficiency

FXS - Fragile X Syndrome

FXTAS - fragile X-associated Tremor/Ataxia Syndrome

GSEA - Gene Set Enrichment Analysis

H3K36 – Histone methyltransferase that transfers methyl groups to lysine residues in position 36 of Histone 3

H4K20 – Histone methyltransferase that transfers methyl groups to lysine residues in position 20 of Histone 4

Hb - Hemoglobin

HbA - Adult hemoglobin

HbA2 - Adult hemoglobin 2

HbF - Fetal hemoglobin

HbVar - Human Hemoglobin Variants database

HCM - Hypertrophic Cardiomyopathy

HGMD - Human Gene Mutation Database

HPLC – High-Performance Liquid Chromatography

iPSC - induced Pluripotent Stem Cells

KCl - Potassium chloride

Mb - megabase

MCH - Mean Corpuscular hemoglobin

MCV - Mean Corpuscular Volume

MLPA - Multiplex Ligation Dependent Probe Amplification

mPCR - methylation PCR

mRNA - messenger RNA

MsigDB - Molecular Signature Database

NGS – Next Generation Sequencing

NID - Nuclear Interaction Domain

NIPS - Non-Invasive Prenatal Screening

NSD1 - Nuclear receptor Set Domain containing protein 1 gene

PCR - Polymerase Chain Reaction

PCV - Packed cell volume

RBC - Red Blood Cells count

RP-PCR - Repeat-Primed PCR

RT-PCR- Reverse Transcriptase PCR

saRNAs - small activating RNAs

SIGU - Italian Society of Human Genetics

STRING - Search Tool for Retrieval of Interacting Genes/Proteins database

TNGB - Telethon Network of Genetics Biobanks

UTR - untranslated region

VUS - Variant of Uncertain Significance

WES - Whole Exome Sequencing

WGS - Whole Genome sequencing

α - alpha

β - beta

γ - gamma

δ - delta

ISTITUTO GIANNINA GASLINI

ISTITUTO PEDIATRICO
DI RICOVERO E CURA
A CARATTERE SCIENTIFICO

CHAPTER I

Introduction to the Human Genetics Laboratory

Internship Report

Master's Degree in Laboratory Clinical Genetics

Rute Pinto
2021/2022

Chapter I

Introduction to the Human Genetics Laboratory

1.1 *Giannina Gaslini* Institute

Figure 1 - (2019). Giannina Gaslini Institute in Genova, Italy. Retrieved from https://eatris.eu/institutes/irccs-istituto-giannina-gaslini-igg/.

The *Giannina Gaslini* Institute is placed in Genova, and began its activity back in 1938, consisting of 20 buildings well connected by an underground system. The founder, Gerolamo Gaslini, wanted the presence of the University of Genova within the Institute through lessons on pediatrics to ensure the qualified support of research and training of new generations of pediatricians, nurses, and technicians. In 1959, the Institute was formally recognized as a National Scientific Institute within the National Health System.

With the popular principle "*Curiamo i bambini*" ("we cure children"), the Institute has a comprehensive organization of diverse multidisciplinary fields, including pediatric and surgical specialties, scientific laboratories, and university professorships. In addition, also religious activities can be provided to patients and family, by religious assistants, to offer them spiritual support. The *Giannina Gaslini* Institute has always been a national and international reference point in many disciplines. Every year is visited by a thousand children of ninety nationalities from all over the world, and over twenty thousand from all Italian regions. In the last decade, the Institute has obtained and maintained the accreditation of excellence based on the standards of the International Joint Commission,

which have as their objective the continuous improvement of care with maximum safety for patients and operators (Istituto Giannina Gaslini, 2019).

1.2 Human Genetics Laboratory

I spent my eleven-month internship (from October 2021 to August 2022) in the Human Genetics Laboratory, a unit of the *Giannina Gaslini* Institute that specializes in pre-and postnatal genetic and genomic testing, within the National Health System. The main objectives of these laboratory activities are to identify rare genetic disorders and evaluate the genetic component of complex and oncological diseases. It is a modern facility with high-quality systems, including a pneumatic system that connects the laboratory with other parts of the hospital, and enables the quick and practical distribution of samples. Additionally, this laboratory is also devoted to some research activities. The laboratory's director is the Dr. Domenico Coviello, M.D. Ph.D.

The facility is divided into three principal sections,

- Building 3, responsible for prenatal cytogenetics and molecular genetics;
- Building 15, responsible for postnatal cytogenetics and Array-CGH;
- Building 16, responsible for oncohematological cytogenetics.

Diagnostic approaches used in the human genetics laboratory can range from the most traditional ones, such as conventional karyotypes and Sanger sequencing, to the most sophisticated ones, such as SNP-Array and Next Generation Sequencing (NGS), including the most recent bioinformatics databases and tools.

The laboratory's diagnostic activities involve mainly:

- Conventional cytogenetic testing for the diagnosis of germline and somatic chromosomal abnormalities, and molecular cytogenetics for the evaluation of genomic rearrangements, such as FISH and Array-CGH;

- Molecular testing is available for the assessment of patients with diverse clinical indications, including for example Cardiomyopathies, Cystic Fibrosis,

Mediterranean Fever, Short stature associated syndromes, Male sterility, and Hemoglobinopathies;

- Prenatal molecular screening for the analysis of trisomy 13, 18, and 21 by NGS analysis of fetal cell-free DNA in maternal serum;

- Investigations for the diagnosis of Thalassemia and screening of other Hemoglobinopathies.

1.3 Reference

Istituto Giannina Gaslini. (2019). Istituto Giannina Gaslini – Annual Report 2019. Retrieved from https://www.gaslini.org.

CHAPTER II

Research in Molecular Medicine

Internship Report

Master's Degree in Laboratory Clinical Genetics

Rute Pinto

2021/2022

Chapter II

Research in Molecular Medicine

Research is the light of scientific knowledge. Particularly in health, it is an essential tool for understanding how human beings function, whether is in terms of their cells, specific biologic processes, unidentified disorders, or to test novel therapeutic drugs in view of innovative therapies.

In the Human Genetics Laboratory, I was allowed to be integrated into a research project, directed by the laboratory director, Doctor Domenico Coviello, M.D. Ph.D, and the researcher Dr. Giuseppina Conteduca, Ph.D. The research project was focused on the study of Sotos Syndrome, a genetic overgrowth disorder caused by pathogenic germline variants in the *NSD1* gene (Tatton-Brown, RP Cole, & Rahman, 2019). The objective of this project was to elucidate more about the role of *NSD1* pathogenic variants at the molecular and genetic level in view to develop new innovative therapies for this disease. This chapter is constituted by a brief introduction to the Sotos Syndrome and its causative gene, a description of the work developed during the research project, and moreover, my contribution to the research work.

2.1 Overview of Sotos Syndrome

Sotos1 Syndrome (Sotos) (MIM #117550) is an overgrowth developmental syndrome, with an autosomal dominant hereditary pattern, caused by germline mutations in the *NSD1* gene (Tatton-Brown, RP Cole, & Rahman, 2019). These include intergenic mutations (such as missense, truncating, and splice-site mutations), and deletions in the 5q35 chromosomic region involving the *NSD1* gene. Additionally, the observation that a single functional allele is not sufficient to maintain the normal phenotype, indicates that Sotos is driven by an haploinsufficiency mechanism (Kurotaki et al., 2003). Since the majority of the Sotos cases that have been documented are caused by germline *NSD1* variants, its diagnosis is suspected in the presence of a consistent clinical picture and

requires the molecular diagnosis of a heterozygous pathogenic variant or deletions spanning the *NSD1* gene. (Tatton-Brown, RP Cole, & Rahman, 2019).

Relatively to its clinical features, and concordant with Tatton-Brown & Rahman (2006), more than 90% of Sotos patients appear to have three cardinal characteristics: excessive growth, a non-progressive neurological disorder with a learning disability, and a characteristic facial appearance as the broad and prominent forehead, sparse frontotemporal hair, downslanting palpebral fissures, long and narrow face, and long chin. Other Sotos characteristics were also present as major features in more than 15% of cases including advanced bone age, cardiac defects, and renal alterations.

Figure 2 - Classical Sotos traces in childhood (a) and in adulthood (b) (Tatton-Brown & Rahman, 2006).

NSD1 Gene and Protein

The *N*uclear receptor *S*et *D*omain containing protein *1* gene (*NSD1)*, located at chr5q35, is a DNA coding-protein segment with 23 exons of length ("Human gene NSD1," 2013). NSD1 protein participates in the methylation process, which is a post-translational modification on the N terminal tail of histones. Specifically, *NSD1* encodes a lysine methyltransferase that transfers methyl groups to lysine residues in position 36 of Histone 3 (H3K36). However, it can only methylate in that position till the second degree. Other lysine methyltransferases, like SETD2, can methylate the same residues until the third degree. Both the degree and position of methylation influence chromatin regulation and gene expression, supporting the evidence that histone methyltransferases

have crucial functions in the regulation of biological processes. An increasing amount of evidence indicate that the critical role of the *NSD1* gene in Sotos syndrome is associated with the NSD1 protein activity, which controls transcriptional master regulators (Tauchmann & Schwaller, 2021).

The NSD1 lysine methyltransferase protein is a complex epigenetic regulator with multiple domains, and distinct functions. These domains include a catalytic domain composed of a SET domain (known as enhancer-of-zeste, trithorax) and its adjacent SET-associated cysteine-rich domains, with the function to regulate chromatin. Additionally, the protein domains include five plant homeodomains and two proline-tryptophan-proline domains, present in most methyltransferases and implicated in interactions between proteins. Finally, it includes two nuclear receptor interaction domains (NID $^{-L}$ and NID $^{+L}$) typically present in coactivators and corepressors, suggesting that NSD1 methyltransferase may work as a transcription intermediary factor able to regulate, both positively and negatively, transcription. However, one of the most highlighted domains is the SET domain, due to its histone specificity (Tauchmann & Schwaller, 2021). The research work by Rayasam (2003) found that the SET domain has methylation specificity for H3K36 but also for H4K20, although the biochemical assays didn't confirm the potential H4K20 substrate.

Over the years, the clinical importance of the *NSD1* gene has been more evident. The evidence that mice embryos with homozygous disruption on the *NSD1* gene did not develop till gastrulation, failing the embryogenesis, highlights the role of *NSD1* in post-implantation. The same experience, but with an heterozygous *NSD1* gene disruption, claimed better results. Embryos were able to finish the embryogenesis and had milder Sotos phenotype when compared to humans. However, it is important to consider that the impact of the disrupted genotype on the rats phenotypic is less severe when compared to the human phenotype (Rayasam, 2003), which suggests that homozygous disruption of the NSD1 gene in humans don't allow embryos to develop. Therefore, gene expression experiments along with SOTOS phenotype indicate that *NSD1* may have a crucial role during embryogenesis by modeling human development (Tauchmann & Schwaller, 2021). Furthermore, an interesting study showed that NSD1 could bind close to the promotor and thereby regulate genes involved in crucial processes such as bone morphology, keratin biology, and cell growth (Lucio-Eterovic et al., 2010). Since several

of these physiologic and important processes appear to be altered in the clinical picture of Sotos, it would be intriguing to investigate further in this direction.

In literature, the identified *NSD1* variants associated to Sotos were observed to be deletions and loss-of-function variants. On the other hand, curiously, microduplications were identified in the *NSD1* gene, in the chromosomic region 5q35.2–q35.3, to develop a "Reverse SOTOS phenotype" in patients (Novara et al., 2014). Apparently, while Sotos syndrome is driven by loss-of-function, Reverse Sotos is driven by NSD1 gain-of-function, inducing features like short statute (the opposite of overgrowth syndromes like Sotos), microcephaly and learning disabilities, but with no distinctive facial appearance or typical osseous abnormalities. It was already known that NSD1 abnormalities were a hallmark of developmental syndromes, but this finding suggests that *NSD1* gene dosage may determine its phenotype.

Furthermore, research suggests that somatic mutations on the *NSD1* gene may have a role in cancer. The first reported evidence was the association of *NSD1* genetic abnormalities to pediatric myelodysplastic syndromes and aggressive acute myeloid leukemia through a fusion protein, a gain-of-function product that results from the chromosomal translocation that fuses NSD1 and NUP98 proteins (Jaju, 2001). Besides this, also *NSD1* loss-of-function contributes to tumor formation through epigenetic alterations, as observed by the *NSD1* epigenetic silencing in human brain tumor cells and human clear cell renal carcinoma by the abnormal methylation its promoter (Tauchmann & Schwaller, 2021). Interestingly, the specific genome-wide methylation signature observed in Sotos syndrome was shown to display a consistent pattern with the one observed in *NSD1* epigenetic silencing, seen in human clear cell renal carcinoma (Su et al., 2017). Although the epigenetic mechanism of *NSD1* silencing is currently considerably known, it remains unclear how a single NSD1 point mutation will contribute to malignant transformation.

2.2 Scope of the Research Project

The major goal of the project was, and it still is, to set up an *in vitro* cell model culture to investigate the molecular component responsible for Sotos syndrome with the perspective of identifying innovative therapies.

This work began in 2020, year in which Sotos patients with an established molecular diagnosis were selected, as well as parents' fibroblasts, to set up fibroblasts cell lines. These were collected at the *Giannina Gaslini* genetic biobank. In the second year (2021), when I was integrated, global gene expression (23 000 genes, by the Agilent Expression chip) was tested to identify over-expressed or down-expressed genes in the presence of the haploinsufficiency mechanism of the *NSD1* gene, by comparing the gene expression profiling of Sotos fibroblasts and controls. For validation of the microarray gene expression patterns, Quantitative Real-Time Polymerase Chain Reaction was performed. In addition, a bioinformatic analysis was performed using the Gene Set Enrichment Analysis and a software for network analysis.

The next step of the project (in 2022) is aimed at the generation of induced Pluripotent Stem Cells (iPSCs), from differentiated fibroblasts, to further generate *in vitro* iPSCs-derived neurons (both from patients and controls). In these reprogrammed cells, several small activating RNAs (saRNAs) will be tested to identify the ones able to overexpress the normal allele. The identified saRNAs are gonna be used to stimulate the *NSD1* promotor, compensating for the unbalanced gene dosage induced by the haploinsufficiency of *NSD1*. If this technique is effective, it will be employed as a new gene treatment for Sotos patients.

My involvement on the research work was mainly focused on the work about the global gene expression, that identified differentially expressed long noncoding RNAs, miR646, and genes controlling the G2/M Checkpoint. The resulted work culminated in the following published article (figure 3).

2.3 Description of the Global Gene Expression work

life

MDPI

Article

NSD1 Mutations in Sotos Syndrome Induce Differential Expression of Long Noncoding RNAs, miR646 and Genes Controlling the G2/M Checkpoint

Giuseppina Conteduca [1,†], Davide Cangelosi [2,†], Simona Coco [3], Michela Malacarne [1], Chiara Baldo [1], Alessia Arado [1], Rute Pinto [1], Barbara Testa [1] and Domenico A. Coviello [1,*]

[1] Laboratory of Human Genetics, IRCCS Istituto Giannina Gaslini, 16147 Genoa, Italy; giusy_conteduca@alice.it (G.C.); michelamalacarne@gaslini.org (M.M.); chiarabaldo@gaslini.org (C.B.); alessiaarado@gaslini.org (A.A.); ruteluisapinto@gmail.com (R.P.); testabarbara@libero.it (B.T.)
[2] Clinical Bioinformatics Unit, IRCCS Istituto Giannina Gaslini, 16147 Genoa, Italy; davidecangelosi@gaslini.org
[3] Lung Cancer Unit, IRCCS Ospedale Policlinico San Martino, 16132 Genoa, Italy; simona.coco@hsanmartino.it
* Correspondence: domenicocoviello@gaslini.org; Tel.: +39-010-5636-3977
† These authors contributed equally to this work.

Abstract: An increasing amount of evidence indicates the critical role of the *NSD1* gene in Sotos syndrome (SoS), a rare genetic disease, and in tumors. Molecular mechanisms affected by *NSD1* mutations are largely uncharacterized. In order to assess the impact of *NSD1* haploinsufficiency in the pathogenesis of SoS, we analyzed the gene expression profile of fibroblasts isolated from the skin samples of 15 SoS patients and of 5 healthy parents. We identified seven differentially expressed genes and five differentially expressed noncoding RNAs. The most upregulated mRNA was stratifin (*SFN*) (fold change, 3.9, Benjamini–Hochberg corrected $p < 0.05$), and the most downregulated mRNA was goosecoid homeobox (*GSC*) (fold change, 3.9, Benjamini–Hochberg corrected $p < 0.05$). The most upregulated lncRNA was lnc-C2orf84-1 (fold change, 4.28, Benjamini–Hochberg corrected $p < 0.001$), and the most downregulated lncRNA was lnc-C15orf57 (fold change, −0.7, Benjamini–Hochberg corrected $p < 0.05$). A gene set enrichment analysis reported the enrichment of genes involved in the KRAS and E2F signaling pathways, splicing regulation and cell cycle G2/M checkpoints. Our results suggest that *NSD1* is involved in cell cycle regulation and that its mutation can induce the down-expression of genes involved in tumoral and neoplastic differentiation. The results contribute to defining the role of *NSD1* in fibroblasts for the prevention, diagnosis and control of SoS.

Keywords: G2/M checkpoint; Sotos syndrome; gene expression; *NSD1*; noncoding RNAs

check for updates

Citation: Conteduca, G.; Cangelosi, D.; Coco, S.; Malacarne, M.; Baldo, C.; Arado, A.; Pinto, R.; Testa, B.; Coviello, D.A. *NSD1* Mutations in Sotos Syndrome Induce Differential Expression of Long Noncoding RNAs, miR646 and Genes Controlling the G2/M Checkpoint. *Life* **2022**, *12*, 988. https://doi.org/10.3390/life12070988

Academic Editors: Sanjeev Gupta and Bijan K. Dey

Figure 3 - Article published in the scope of the Sotos Syndrome project "*NSD1* Mutations in Sotos Syndrome Induce Differential Expression of Long Noncoding RNAs, miR646 and Genes Controlling the G2/M Checkpoint", published on July 2, 2022 (https://doi.org/10.3390/life12070988)

In the light of basic biology knowledge, DNA molecules serve as a template for the formation of RNA molecules, a process known as transcription. When a DNA fragment (gene) is expressed, RNA molecules are formed. That means the possible synthesis of messenger RNA (mRNA) or the synthesis of noncoding RNA. While mRNA are RNA molecules that contain the genetic information for protein synthesis, via a process known as translation, noncoding RNAs do not codify a protein. Yet, although noncoding RNAs are not directly involved in protein synthesis, they play crucial roles in cells, such as the regulation of gene expression during the transcriptional or post-transcriptional phases (Mattick & Makunin, 2006). Noncoding RNAs can be divided into small RNAs, with less than 200 nucleotides, or long RNAs, with above 200 nucleotides (Bhat et al., 2020).

Genes are selectively activated or inactivated as cells become more differentiated, only genes specific for that cell-type functions are expressed. Mutations in genes that regulate gene expression, such as *NSD1*, may alter the profile of expression of cells, with possible pathological implications. Sotos Syndrome is caused by mutations in the *NSD1* gene, which is a transcriptional regulator that controls methylation and regulates the expression of many other genes. To evaluate possible alterations in the gene expression pattern in Sotos patients, RNA from Sotos and patients' fibroblasts were used to guarantee that we were studying the products of the altered expression of each gene. For the analysis of the global gene expression, an Agilent Expression chip was used to identify over-expressed or down-expressed genes due to the haploinsufficiency of the *NSD1* gene.

My contribution to this work was related to the validation of the microarray gene expression patterns, through the Quantitative Real-Time PCR, and to the literature revision about the dysregulated genes identified by microarray. A flowchart of the steps of the project is provided below (figure 4).

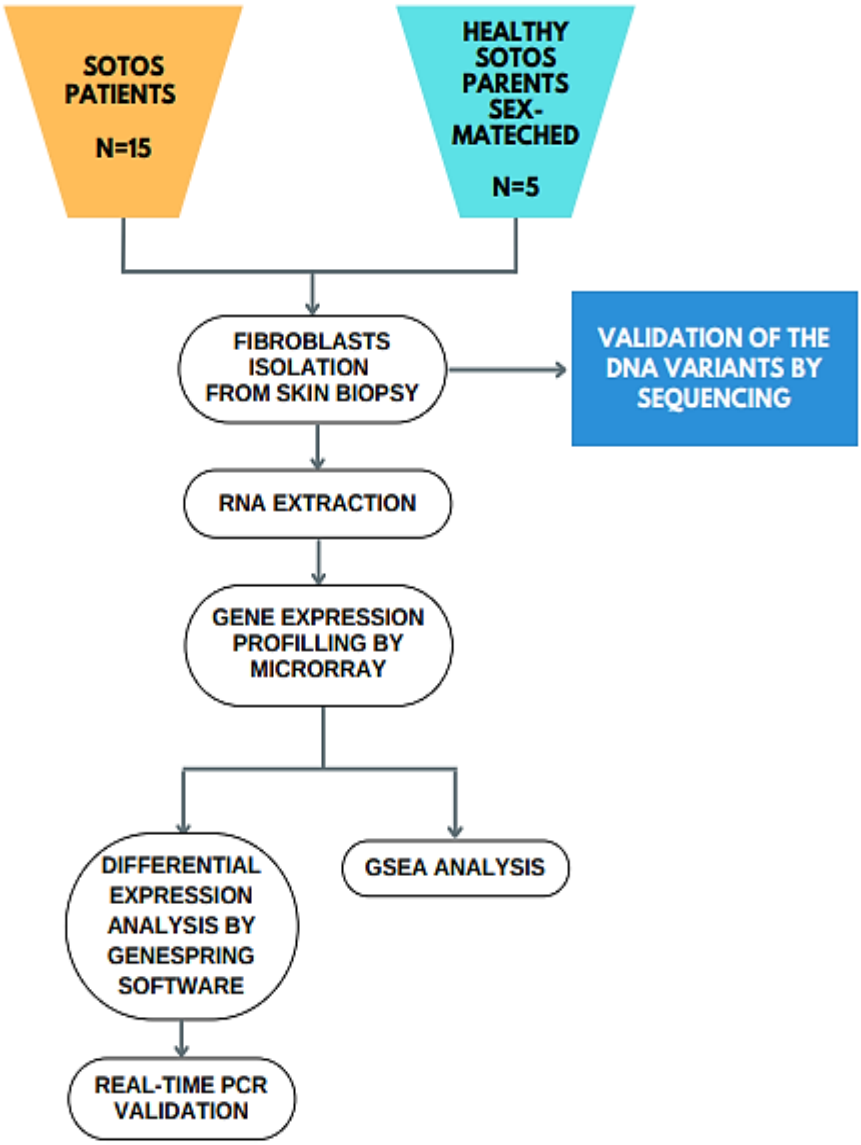


Figure 4 - Flowchart of the strategy used for the global gene expression assessment of fibroblasts from Sotos patients and healthy controls.

Stage 1 - Patients Selection and Establishment of Fibroblasts Cell Lines

To analyze the gene expression profile of fibroblasts, 15 Sotos patients with a well-established molecular diagnosis of Sotos syndrome, and showing classical Sotos features, were selected for a biopsy of the forearm skin. Molecular diagnosis of the patients was made by the identification of the *NSD1* pathogenic variants (point mutations and deletions) on peripheral blood. For the control group, 5 sex-matched healthy Sotos parents were selected for a skin biopsy. Fibroblasts were isolated from samples and cultured to grow in the genetic biobank to perform DNA and RNA extraction. To verify the presence of the *NSD1* pathogenic variants in Sotos patients' fibroblasts, DNA sequencing of the

fibroblast's cell lines was performed. For that, fibroblasts DNA was amplified by PCR using specific primers for the 23 exons that constitute the *NSD1* gene. For confirmation, PCR products were sequenced by automated Sanger sequencing (kit ABI BigDye Terminator Ready Reaction Mix and analyzed by the ABI 3130XL Genetic Analyzer).

Stage 2 – Global Gene Expression Profile

For the global gene expression profiling, microarray technology was used. DNA microarray is a very useful and effective technologic tool to compare the expression of thousands of genes in different samples (Arcellana-Panlilio & Robbins, 2002). DNA microarray technology has been defined as high-density arrays of DNA fragments or oligonucleotides, known as probes, which hybridize with the DNA sample by nucleotide base-pairing. The choice of the microarray chip depends on the goal of the study because probes are specific for a specific fragment. In the case of gene expression, probes are specific for random genes, allowing us to assess the level of activity of genes in each sample (Gabig & Wegrzyn, 2001).

Microarray, or "chip", is a solid surface covered by thousands of spots on their surface. Each spot contains a considerable number of nucleic acid fragments, generated from single genes, allowing to investigate the hybridization of different genes in parallel (Arcellana-Panlilio & Robbins, 2002). Fragments placed in each spot can be of cDNA or oligonucleotides, depending on the hybridized sample. For the global gene expression profiling, microarray can be done with short oligonucleotides, of about 25 to 30 bases, long oligonucleotides, of about 50 to 80 bases, or cDNA, of different lengths (Petersen et al., 2005). In this work, oligonucleotide microarray was used. Oligonucleotide microarrays can be used in a generic way, with random oligonucleotides, or targeted, to identify molecular variants of a target (McPherson & Pincus, 2021). Oligonucleotides bind to the target DNA through nucleotide base-paring, with the advantage of enabling a high targeting of specified genes, which means that it can better discriminate between similar targets (such as splice variants and hereditary genes) (Kreil, Russell, & Russell, 2006).

For the determination of the gene expression profiling, fibroblasts RNA from patients and controls were used, allowing us to analyze only the expressed genes, through its products. RNA was extracted from fibroblasts, quantified, and assessed for its integrity

by the TapeStation from Agilent. Because RNA is a very sensible molecule, its integrity assessment is necessary for the quality of the microarray results. The Agilent One-Color microarray gene expression kit was used for the gene expression profiling, which prepares samples to be hybridized in a gene expression chip and includes step-by-step instructions for microarray execution. Following the manufacturer instructions, first RNA samples from Sotos and controls were labeled.

Fibroblasts RNA were converted into double strand complementary DNA (cDNA), with the addition of the recognition binding site for the T7 RNA Polymerase. Then, the T7 RNA Polymerase enzyme amplifies the cDNA fragments and incorporates fluorescent nucleotides (Cyanine 3-CTP), generating fluorescent cDNA, also described as complementary RNA (cRNA) (provided by the One-Color Microarray-Based Gene Expression Analysis protocol from Agilent Technologies).

Labeled cRNA was applied to the microarray chip and allowed to hybridize under specific conditions. The microarray chip used was the *in situ*-synthesized oligonucleotide microarray, SurePrint G3 microarray chip (Design ID: G4851c), which contained 50,599 probes for 32,776 human mRNAs and 17,438 human noncoding RNAs.

After hybridization, washing was necessary, and slides were scanned by an Agilent microarray scanner (Agilent Technologies). Moreover, data were extracted and then processed by the Genespring software, which normalizes and filters data.

If the hybridized cRNA is present with many copies of a specific gene, many bindings will occur, increasing the fluorescent signal. Therefore, the measurement of the emitted signals was based on the quantity of fluorescent light emitted by the binding between oligonucleotides and the cRNA. After data be processed, Sotos and control samples were compared. If gene expression values of a specific gene in Sotos were significantly above the average media of the controls, that would mean that the specific gene is over-expressed. The same was applied to the reverse situation. If gene expression values of a specific gene in Sotos were significantly below the average media of the controls, that would mean that that specific gene was under-expressed.

Stage 3 - Bioinformatic Analysis of the Microarray Results

Statistical significance tests are commonly used to determine genes that are express differentially under different contexts. To measure the amount of differential expression, the test statistic and its p-value are generated for each gene and compared (Chen, Wang, Tsai, & Lin, 2006). In this work, gene expression profiling was determined through bioinformatics analysis by evaluating and comparing the fold changes and p-values of the microarray values from Sotos fibroblasts and controls, based on the normalized fluorescence signal values of the mRNA and noncoding RNA probes. Fold changes and p-values were calculated with the help of the Student's t-test method (provided by the Genespring software). Results are provided in table 1 and 2.

Nr	Probe Set ID	Gene Symbol	Gene Name	Seq. Name	Log2 Fold Change	*p*-Value	*p*-Value Adjusted
1	A_33_P3389286	*SFN*	stratifin	NM_006142	3.99	1.62×10^{-6}	0.015
2	A_23_P113572	*CD19*	CD19 molecule	NM_001770	3.36	1.28×10^{-6}	0.014
3	A_33_P3305790	*NOS3*	nitric oxide synthase 3 (endothelial cell)	NM_000603	2.4	1.41×10^{-7}	0.0027
4	A_22_P0001791	*ZNF883*	zinc finger protein 883	NM_001101338	1.33	4.95×10^{-6}	0.026
5	A_33_P3334515	*NDRG2*	NDRG family member 2	NM_001282213	−1.24	1.06×10^{-5}	0.047
6	A_24_P317907	*SORBS1*	sorbin and SH3 domain containing 1	NM_001034954	−3.1	1.14×10^{-5}	0.047
7	A_23_P76774	*GSC*	goosecoid homeobox	NM_173849	−3.98	1.38×10^{-5}	0.047

Table 1 - Differentially expressed probe sets of mRNAs in Sotos Samples by microarray analysis

Nr	Probe Set ID	Gene Symbol	Gene Name	Seq. Name	Log_2 Fold Change	*p*-Value	*p*-Value Adjusted
1	A_22_P00002837	Lnc-C20orf197-3	long intergenic nonprotein-coding RNA	lnc-C20orf197-3:13	3.2	4.95×10^{-7}	0.007
2	A_22_P0000283	Lnc-C2orf84-1	long intergenic nonprotein	lnc-C2orf84-1:1	3.2	4.95×10^{-7}	0.007
3	A_22_P00002715	MIR646HG	MIR646 host gene	ENSG00000228340	2.8	8.64×10^{-6}	0.042
4	A_24_P16214	LINC00665	long intergenic nonprotein-coding RNA 665	ENST00000427868	1.2	2.57×10^{-6}	0.016
5	A_32_P98975	C15orf57	chromosome 15 open reading frame 57	NM_052849	−0.7	1.40×10^{-5}	0.047

Table 2- Differentially expressed probe sets of noncoding RNAs in Sotos Samples by microarray analysis.

The fold change (Log_2) parameter represents the average difference in the expression levels between the respective genes of the Sotos fibroblasts and the controls (Chen, Wang, Tsai, & Lin, 2006). A fold change of ≥ 2.0 or ≤ 0.5 was considered to be significantly over-regulated or down-regulated, respectively. The p-value parameter is used to validate an hypothesis against the observed data. A small p-value indicates the evidence of differential gene expression, either by over- or under-expression (Chen, Wang, Tsai, & Lin, 2006). Only genes with adjusted *p*-values of <0.05 were considered to be significantly different expressed.

In table 1, gene expression profiling showed **seven differentially expressed mRNAs**, four upregulated (*SFN, CD19, NOS3,* and *ZNF883*) and three down-regulated (*NDRG2, SORBS1,* and *GSC*). With the most upregulated associated with the *SFN* probe (fold change = 3,9; *p*<0,05), and the most downregulated with the *GSC* probe (fold change = -3,9; *p*<0,05). Additionally, in table 2, **five noncoding RNAs were found to be differentially expressed**, four upregulated (*lnc-C20orf197-3, lnc-C2orf84-1, MIR646HG,* and *lnc-C00665*) and one downregulated (*lnc-C15orf57*) (fold change = -0,7; *p*<0,05). Probes for the *lnc-C2orf84-1* gene (fold change = 3,2; *p*<0,05) were the most upregulated. From the five noncoding RNAs, four were lncRNAs (*lnc-C2orf84-1, lnc-C00665, lnc-C20orf197-3,* and *lnc-C15orf57*) and only one was a small noncoding RNA (*MIR646HG).*

Stage 4 - Quantitative Real-Time PCR Evaluation

Quantitative Real-Time PCR technology was used for the validation of the significantly altered expression of mRNAs and noncoding RNAs, identified by microarray analysis, between fibroblasts from Sotos and controls. For that, cDNA was used. Before Real-Time PCR, specific primers needed to be designed for genes that do not have any primers reported in literature. After primers being designed and ordered, primers needed to be validated. For primers validation, a regular PCR followed by electrophoresis was performed, and the PCR product was sequenced for confirmation. Quantitative Real-Time PCR was performed using LightCycler 480 SYBR Green I Master (Roche Diagnostics, Mannheim, Germany), and repeated 3 times for results reliability. Quantitative Real-Time PCR technique is described in the "Methodology used for research laboratory techniques" (page 33). Real-Time PCR gene expression values were assessed and normalized through the help of the beta-actin gene and the $2^{-\Delta\Delta Ct}$ method. Quantitative expression measurements of the Real-Time PCR were based on the

assessment of the fold change of each expressed gene and compared to the microarray results (figure 5 and 6).

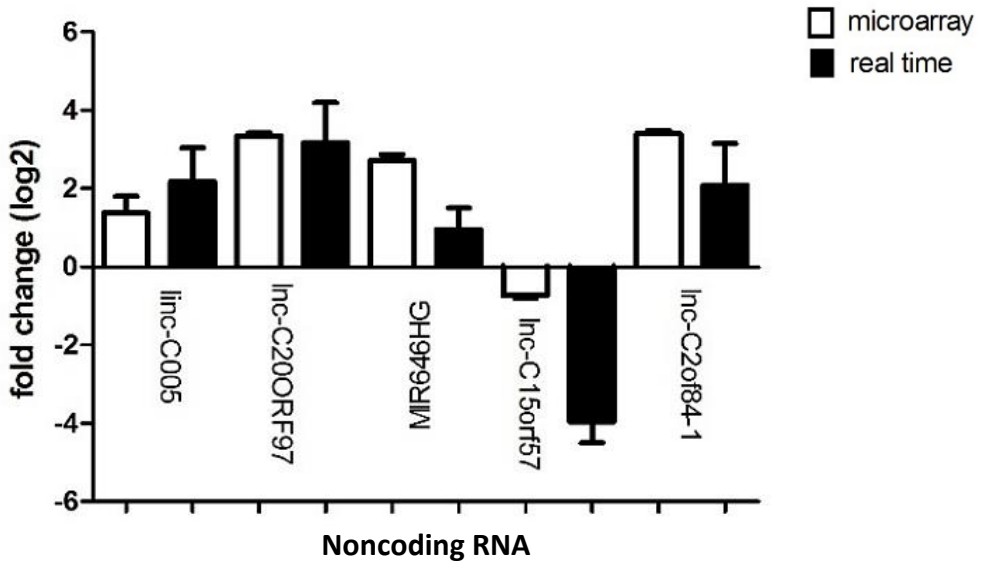


Figure 5 - Quantitative gene expression measurements by Quantitative Real-Time PCR and microarray of the five noncoding RNAs (*lnc-C15orf57*, *MIR646HG*, *lnc-C2orf84-1*, *lnc-C00665*, and *lnc-C20orf197-3*).

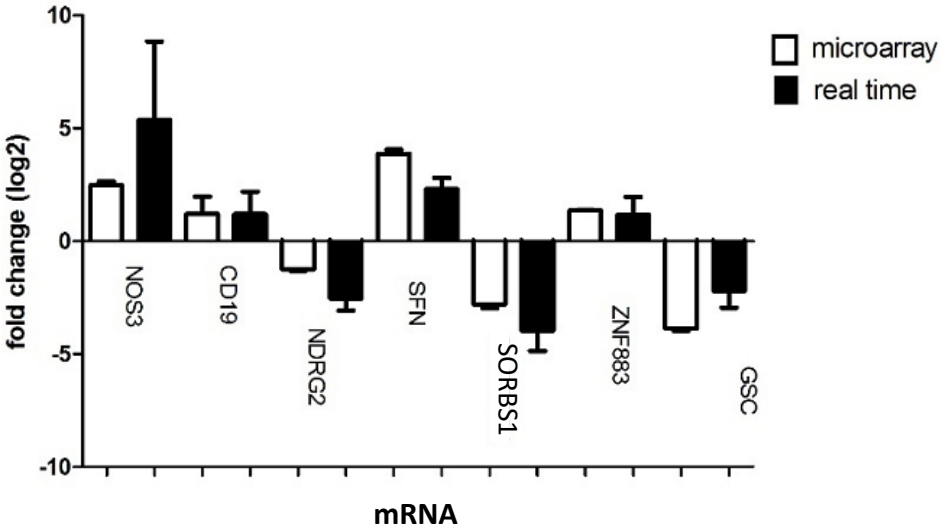


Figure 6 - Quantitative gene expression measurements by Quantitative Real-Time PCR and microarray of the seven mRNAs (*NOS3*, *CD19*, *NDRG2*, *SFN*, *SORBS1*, *ZNF883*, and *GSC*).

In figure 5 and 6 the fold change parameter represents the average difference in the expression level between the respective genes of the Sotos fibroblasts and the healthy controls. The black bars represent the fold changes detected with the microarray, whereas the white bars indicate the average fold change with respect to the triplicate real-time PCR assessments.

The modulation of the gene expression analysis performed by microarray analysis was concordant with the Real-Time PCR results. The same mRNAs and noncoding RNAs were found to be differentially expressed in Sotos fibroblasts.

Microarray chips can screen a lot of genes in one hybridization (>30 000 human mRNAs and about >17 000 human noncoding RNAs), while Real-Time PCR calculates the relative expression of target-genes based on the use of endogenous controls. Real-Time PCR is considered to be the "gold standard" method to confirm microarray gene expression results of each gene. But the diversity of the microarray and Real-Time PCR technical procedures, as well as the biologic diversity of the samples, have a significant impact on its results. It should not be expected the exact same results, and even in some experiences, it may contradict each other (Morey, Ryan, & van Dolah, 2006). However, in this work, both techniques had concordant gene expression results for each gene.

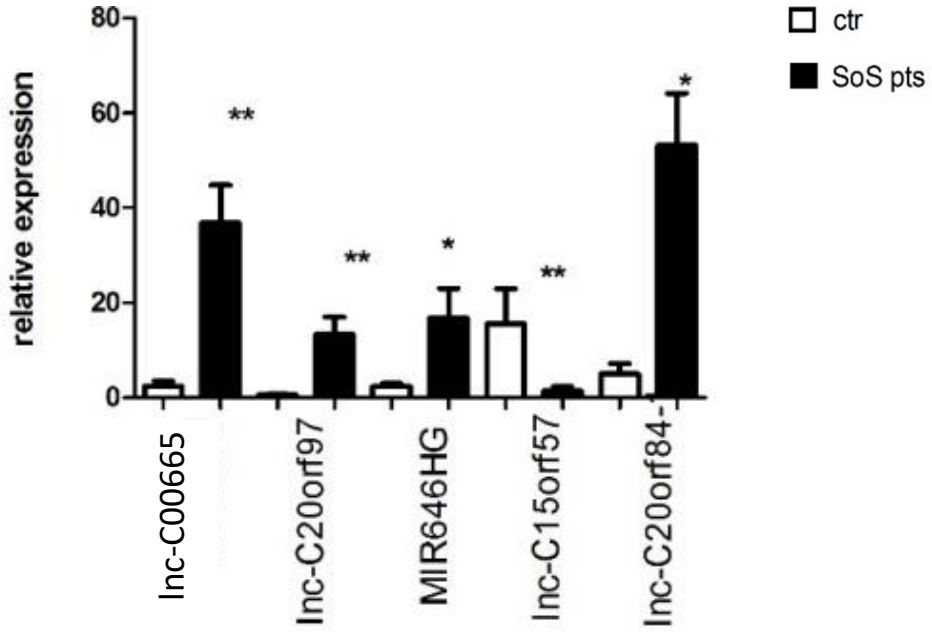


Figure 7 - Gene expression modulation of noncoding RNAs (*lnc-C00665*, *lnc-C20orf97*, *MIR646HG*, *lnc-C15orf57*, and *lnc-C20orf84-1*) of Sotos fibroblasts compared to the control's fibroblasts by Quantitative Real-Time PCR.

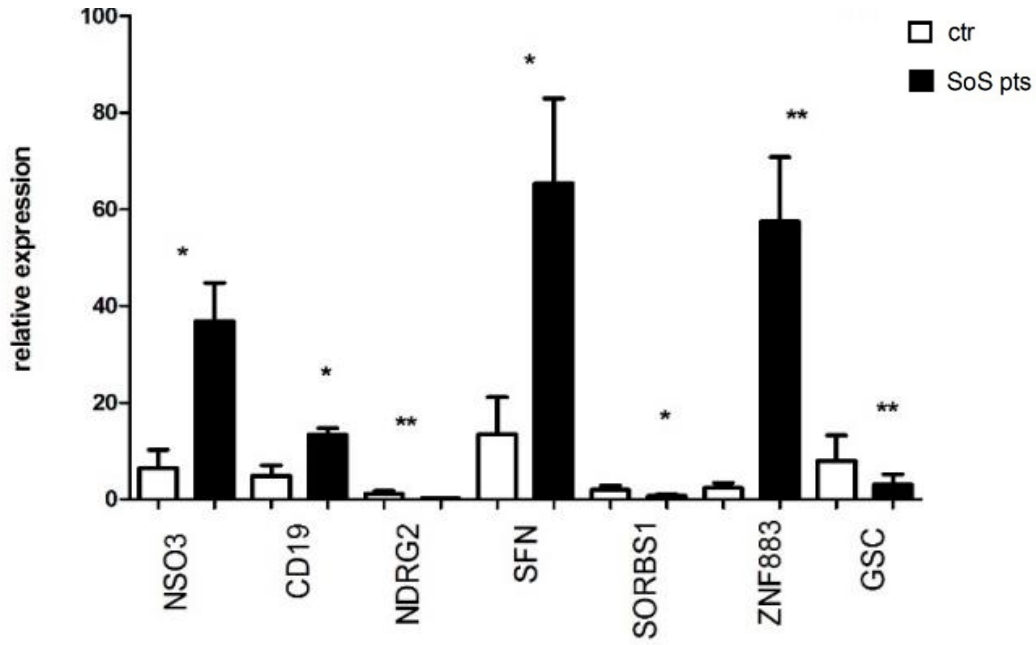


Figure 8 - Gene expression modulation of mRNAs (*NSO3, CD19, NDRG2, SFN, SORBS1, ZNF883,* and *GSC*) of Sotos fibroblast compared to the control's fibroblasts by Quantitative Real-Time PCR.

Real-Time PCR results (figure 7 and 8) were assessed by the Student's t-test. Results showed that *MIR646HG* ($p < 0.05$), *lnc-C2orf84-1* ($p < 0.05$), *lnc-C00665* ($p < 0.001$), *lnc-C20orf197-3* ($p < 0.001$), *NOS3* ($p < 0.05$), *ZNF883* ($p < 0.001$), *CD19* ($p < 0.05$) and *SFN* ($p < 0.05$) were **upregulated** and that those of *GSC* ($p < 0.05$), *lnc-C15orf57* ($p < 0.001$), *NDRG2* ($p < 0.001$) and *SORBS1* ($p < 0.05$) were **downregulated** in Sotos patients, when compared to controls.

Stage 5 - Bioinformatic Investigation of Biological Networks

With the necessity to address the generated big data, the importance of bioinformatics in biology is increasing. In the scope of my internship, I was able to attend to some bioinformatic meetings in order to better understand all the bioinformatic processes that go into producing an article like the one published. I witnessed a short portion of a bioinformatic effort in the STRING database for the paper, a bioinformatic tool that allows researchers to seek biological networks between functional proteins. This analysis employs a statistical algorithm that select the best findings. However, these studies are based on a prediction analysis, therefore, the identified networks need to be supported by experimental studies.

a) Gene Set Enrichment Analysis

Bioinformatic assessment included the Gene Set Enrichment Analysis (GSEA), a bioinformatic tool that is used as an analytical method for gene expression data. GSEA assess group of genes that share the same biological function, regulation, or chromosomal location (Subramanian et al., 2005). In this work, GSEA compared the gene with altered expression with functionally related genes, enabling the identification of the most significantly altered biological processes and pathways associated to the aberrant expression of the *NSD1* gene. Among the different gene sets used for the GSEA, the most relevant ones were "Chemical and genetic perturbations", "hallmark", "gene ontology biological processes", and other gene set collections retrieved from the Molecular Signature Database (MSigDB).

The GSEA bioinformatic analysis found 89 significantly enriched biological processes and pathways in the Sotos fibroblasts when compared to its health controls. The most statistically significant upregulated and downregulated gene sets are indicted in table 3.

Table 3 - Statistically significant enriched biological processes and pathways associated to the mRNAs and noncoding RNAs of Sotos fibroblasts by GSEA.

Over expressed gene sets	Cell cycle and proliferation
	Cell differentiation
	P53-mediated cell cycle arrest
	Cellular senescence
	Cancer
Under expressed gene sets	E2F pathway
	Pediatric cancer markers
	Neoplastic transformation via the STAT3 pathway
	MYC and TFRC targets
	Cell cycle G2/M checkpoint
	TNFA-signaling via NFKB
	Epithelial mesenchymal transition
	Apoptosis process caused by CDKN1A via TP53
	Neoplastic transformation KRAS signaling pathways

These results indicate that these biological processes, including G2/M checkpoints, control transitions between cell cycle phases, cell senescence, and meiotic and mitotic division, and are among the most significantly enriched mRNAs in Sotos patients. In addition, many of these processes are involved with somatic neoplastic disease and tumor development. This suggests that NSD1 participates in cell cycle regulation and that its mutation can induce the down-expression of genes involved in tumoral and neoplastic differentiation.

b) STRING database

Bioinformatic assessment also included the work of identifying possible biological connections between the differentially expressed genes. The network analysis was performed with the help of the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database. This database seeks to incorporate all known and predicted protein-protein interactions, including both physical and functional interactions (Szklarczyk et al., 2020).

Bioinformatic analysis with STRING did not find any direct interactions between the NSD1 protein and other ones. For that reason, analysis was extended to most likely indirect connections. This analysis did find a subset of gene products like the KRAS and pTEN (figure 9). The *KRAS* gene codes for a protein called K-Ras, which is part of a signaling pathway, known as the RAS/MAPK pathway. This protein instructs cells to grow and multiply, or to specialize into a more differentiated cell by transmitting signals from the cell's extracellular matrix to its nucleus (Tidyman & Rauen, 2010). On the other hand, the *PTEN* gene acts as a tumor suppressor gene. Its role on the cell cycle is to regulate the proliferation of cells by interrupting the cell cycle, preventing them to proliferate and divide in an uncontrolled way. The PTEN enzyme makes part of the apoptosis process, a process that stop cells from dividing and triggers it for self-destruction (Yehia, Ngeow, & Eng, 2019). Also, the PTEN enzyme inhibits the PI3K/AKT/mTOR growth-promoting signaling cascade. Pathological variants in the *PTEN* gene are associated with overgrowth (Yehia, Keel, & Eng, 2020).

Therefore, the association between NSD1 and the KRAS/pTEN proteins supports the evidence of NSD1 as a crucial master regulator of gene expression in cell proliferation and differentiation.

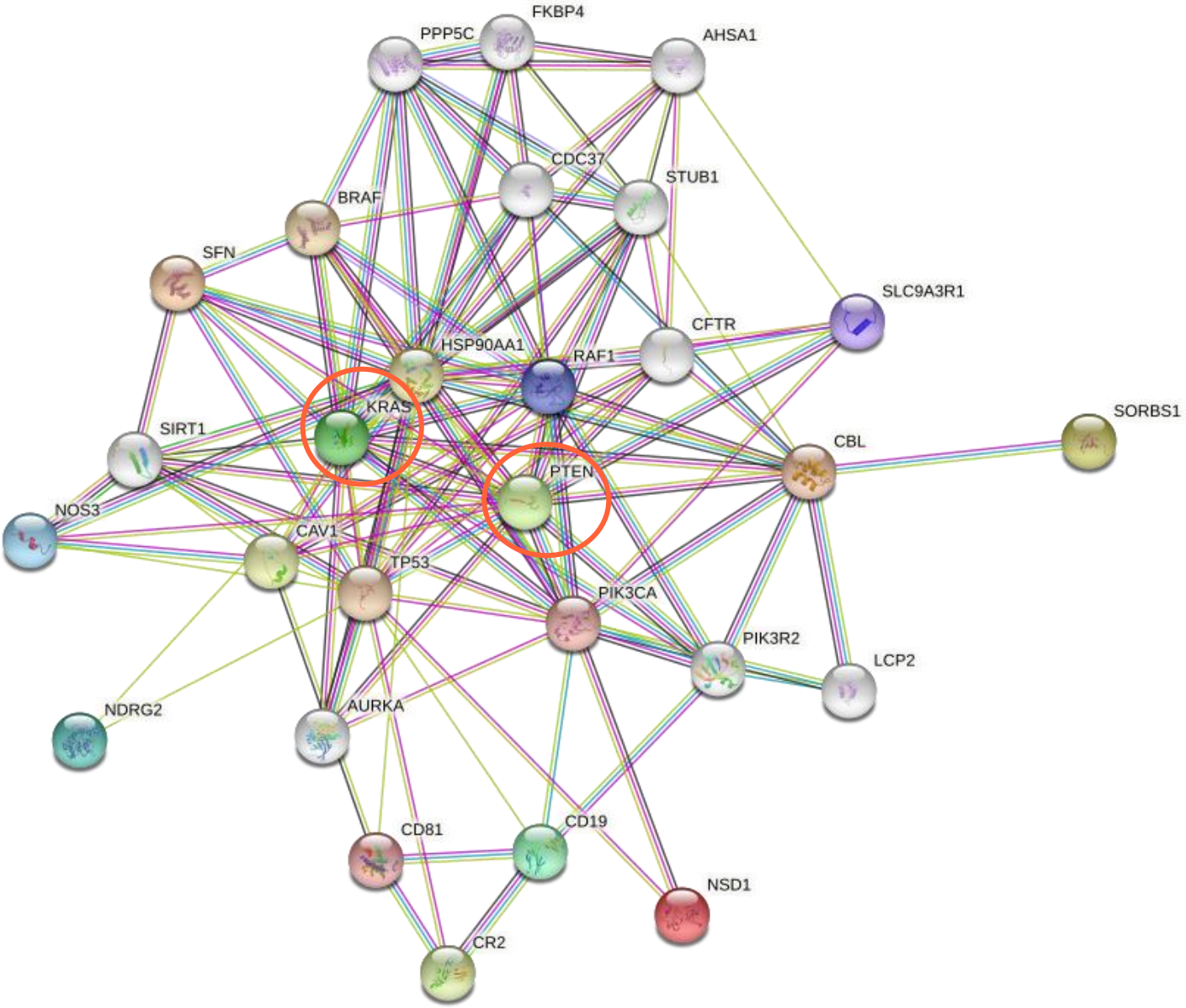


Figure 9 – Output of STRING database; Functional Interaction Network among the differentially expressed genes between Sotos patients and controls. Balls represent gene products and edges represent protein-protein associations. Color indicates different kinds of interactions. Only evidence scores higher than 0.3 are shown.

Stage 6 – Discussion of the Developed Research Work

This investigation identified, on total, five noncoding RNAs and seven mRNAs as being differentially expressed in Sotos patients' fibroblasts relatively to the healthy control fibroblasts. The GSEA explored possible biological functions and mechanisms of these noncoding RNAs and mRNAs.

Although the molecular mechanisms of NSD1 remain unclear, several studies have suggested its role in the regulation of many biological processes. In literature, *NSD1* mutations were reported to induce the alternate methylation of H3K36, which could block cellular differentiation and promote oncogenesis ("Comprehensive genomic characterization of head and neck squamous cell carcinomas," 2015).

Results demonstrated that the identified dysregulated mRNAs, in Sotos samples, were associated with significantly enriched biological processes that control transitions between cell cycle phases, cell senescence, and meiotic and mitotic cell division (including the G2/M checkpoints). Interestingly, most of these functions are involved in neoplastic disease and tumor development, which is consistent with previous studies (Ettel, Zhao, Schechter, & Shi, 2019) (Su et al., 2017).

While KRAS signaling pathway and CDKN1A via TP53-signaling were found to be significantly down-expressed in Sotos samples, genes involved in the regulation of nuclear division, meiotic cell cycle and kinetochore organization were found to be remarkably over-expressed.

Among the most differentially expressed mRNAs, *SFN* gene was the most upregulated (fold change = 3.9) in Sotos samples. Since the SFN protein has crucial roles in the control of the cell cycle and apoptosis, by preventing errors during mitosis, this study suggests that NSD1 may have a crucial role for the cell cycle and mitotic progression, by regulating SFN G2/M checkpoint expression.

GSC, *SORBS1* and *NDRG2* mRNAs were found to be down expressed in Sotos samples, similar with the RNA expression profile of mesenchymal cells (Moradi, et al., 2019). Downregulation of these three genes was shown to trigger epithelial-mesenchymal transition. Furthermore, *GSC* is implicated in mesenchyme-derived tissues during craniofacial development. In mice, *GSC* homozygotic deletions cause craniofacial deformities as well as bone and cartilage defects (Rivera-Perez, Wakamiya, & Behringer,

1999). Therefore, this evidence suggests that *GSC* down-expression may contribute to prolonged bone age and Sotos syndrome-specific facial characteristics such as macrocephaly and long face.

From the noncoding RNAs identified, four were upregulated (*MIR646HG*, *lnc-C2orf84-1*, *lnc-C00665*, and *lnc-C20orf197-3*) and one downregulated (*lnc-C15orf57*). Long noncoding RNAs may provide basic information that could be used to understand pathways related to the disease course of Sotos and to find more effective targeted therapies. Different studies have suggested that long noncoding RNAs alterations may contribute to neuronal and neurodegenerative disease development (Ang, Trevino, & Chang, 2020), and that long noncoding RNAs were also found to be upregulated in neuroblastoma cell lines (Hu, Sun, Hu, & Zhang, 2020), highlighting its probable role in oncogenesis.

Overall, this work highlights the importance of *NSD1* as a gene expression regulator, involved in cell cycle regulation. NSD1 may participate in Sotos syndrome's mechanism of action by regulating the function of noncoding RNAs, inducing the down-expression of *GSC*, *NDRG2*, and *SORBS1* and the over-expression of *SFN* and *ZNF883*. These expression signatures may also be useful tools for screening and monitoring the clinical evolution of Sotos Syndrome and for predicting its prognosis.

2.4 My Research Activities

The practical part developed during the internship was in the research context. The work developed under this scope is described in table 4.

Table 4 - Activities developed in research on the Human Genetics Laboratory

- ☑ **Learn and practice of basic laboratory methods**, useful in the context of research and genetic diagnostic (described in the 2.4.1 point);
- ☑ **Elaboration and presentation of a PowerPoint** work to the laboratory about the following review article:

  Tauchmann, S., & Schwaller, J. (2021). NSD1: A Lysine Methyltransferase between Developmental Disorders and Cancer. *Life*, 11(9), 877. https://doi.org/10.3390/life11090877;
- ☑ **Elaboration of a work of literature review** about thirteen genes, useful for the discussion of results on the published work;
- ☑ **Learn of essential skills by reviewing and correcting drafts of the published article.**

2.4.1 Methodology Used for the Research Activities

In the framework of the research study, I performed several laboratory approaches for my learning experiment. Techniques that were practiced included:

- Manual DNA Extraction;
- DNA Quantification;
- Polymerase Chain Reaction;
- Gel electrophoresis in Agarose;
- Reverse Transcriptase PCR;
- Quantitative Real-Time PCR;
- Automated Sanger Sequencing;
- Manipulation of Fibroblasts Cell Cultures.

Additionally, I observed other methods like RNA extraction, the Real-Time PCR *in silico* analysis, and the observation of the manipulation of induced Pluripotent Stem Cells.

2.4.1.1 Manual DNA Extraction

DNA extraction was performed following the manufacturer instructions reported in the protocol "Purification of Total DNA from Animal blood or Cells (Spin-Column Protocol)" July 2006 edition from QIAGEN company, with this protocol DNA can be extracted and purified from cultured cells or blood of animals or humans. It consists of four essential steps: cell lysis, binding, wash, and elution. To facilitate the execution of the DNA extraction, I wrote my own protocol based on the company's protocol, which is provided in the annex 1 (page 122 to 123).

For the lysis step, proteinase K, and a specific buffer were added and incubated at 56ºC to lyse cells, remove total protein, and ensure the inactivation of DNases and RNases. After that, the lysate is loaded into a column, the "DNeasy Mini spin column". After centrifugation, the DNA was to the filter of the mini spin column, while contaminants are eliminated. The remaining contaminants were eliminated by two washing steps. Finally, the elution step was accomplished by adding water or elution buffer to the mini spin column, followed by centrifugation.

2.4.1.2 DNA Quantification

The equipment used for quantification in the human genetic laboratory for research was the Thermo NanoDrop 1000 Spectrophotometer, by Thermo Fisher Scientific. It measures the nucleic acid concentration and purity of samples up to 3700 ng/µl. The "background" is a spectrum taken from a reference sample "blank". When measuring a sample, 1 µL of sample was utilized, and the intensity of light passing through the sample is recorded. The absorbance was determined automatically by the spectrophotometric equipment through the sample and blank intensities. Using Beer's equation (Absorbance = Wavelength x Path Length x Concentration) and a known amount of absorbance to a certain wavelength, the device estimates the DNA concentration.

2.4.1.3 Polymerase Chain Reaction

Polymerase Chain Reaction (PCR) is a simple amplification technique, in which a single molecule of DNA can generate a large number of similar copies of DNA strands, in a short period of time, and only a few reagents and materials are necessary. The main advantage of this technique is that it's not a very demanding method with the amount of DNA needed, it's considered to be a sensitive assay. The starting DNA template may come from different sources, including genomic DNA, complementary DNA, or even mitochondrial DNA (Kadri, 2020).

For a successful PCR reaction, besides the target DNA template, other components are needed including primers, nucleotides, and DNA polymerase. Specifically, Taq DNA polymerase is an enzyme that adds single nucleotides to the complementary strand being synthesized, even at high temperatures. The added nucleotides are the four known nucleotides on DNA, adenine, thymine, cytosine, and guanine. On the other hand, primers are short DNA fragments with a sequence complementary to a region of the target DNA, their role is to bind the specific sequence where the DNA polymerase starts to amplify on the template, to obtain the PCR product amplification (Garibyan & Avashia, 2013).

Then, a mixture was prepared with all the above-mentioned components and solutions like sterile water DNAase and RNAase free, a buffer solution, that provides an optimal chemical environment for the DNA polymerase activity, and magnesium chloride solution, a cofactor that enhances the activity of Taq DNA polymerase. This combination of elements was prepared on a sample tube and then placed in a thermocycler (Garibyan & Avashia, 2013). This machine is programmed to operate under specific conditions, defining the period of time and the temperature of each cycle (Weier & Gray, 1988). One cycle of PCR reaction consists of 3 essential steps - denaturation, annealing, and elongation. The strands resulting from each synthesis step act as templates for subsequent steps, resulting in exponential amplification (Kadri, 2020).

For the first step "denaturation", the PCR mix is heated at 95ºC. At high temperatures, hydrogen bonds can no longer be maintained, causing the double-stranded DNA to separate into single-stranded DNA. Following this, temperature decreases allowing specific primers to bind by complementarity to specific DNA targets, known as annealing or hybridization. This is a crucial step because, among other things, the annealing

temperature may be too high, and the primers may not have enough specificity due to ambiguity in certain rich CG sequences. In the next step, the temperature is increased to create an optimal environment for the DNA polymerase to recognize the primers and extend them by adding nucleotides, this is the elongation phase. As a result, a new complementary DNA strand is synthesized for each single DNA strand amplified, meaning that each cycle will double the number of DNA molecules (Garibyan & Avashia, 2013).

During the internship, the PCR technique was performed multiple times for the amplification of cDNA fragments. To facilitate the execution of the PCR, I wrote my own protocol based on what I learned in the laboratory, which is provided in the annex 2 (page 124 to 125).

2.4.1.4 Reverse Transcriptase PCR

Reverse Transcriptase PCR (RT-PCR) is a sensitive PCR-based technique that essentially consists in the synthesis of cDNA from an RNA template, and then amplify it. First, RNA must be isolated and purified from the original tissue, and diverse kits are commercially available for this, but preferably, the chosen kit should allow the extracted RNA to be free of contaminants, enzymatic reaction inhibitors, proteins, and polysaccharides. During RT-PCR reaction, RNA is reversed transcripted, by an RNA-dependent DNA polymerase, into cDNA. The choice of the polymerase should be considered along with the length of the amplicons. Because RNA transcripts have the potential to generate secondary structures, RNA integrity should be assessed. Furthermore, the use of random primers on the RT-PCR reaction increases the number of mRNA molecules that may be transcripted from an RNA sample (Malewski, Malewska, & Rutkowski, 2003).

Frequently, the aim of the technique is to quantify a target gene expression. For this effect, usually, the next step is the amplification of the cDNA by conventional PCR using specific primers for the target gene (Real-Time PCR). Nowadays, the conjugation of Real-Time PCR and RT-PCR is a reality, allowing to jump the RT-PCR step, and directly loading the RNA into the Real-Time PCR equipment. It is called Quantitative Real-Time Reverse Transcription PCR, which not only detects the mRNA transcript but also measures the relative amounts of mRNA in a sample. If the aim of the experience

was to detect the presence or absence of a determined RNA molecule, other types of assays can also be used (Malewski, Malewska, & Rutkowski, 2003).

To facilitate the execution of the RT-PCR technique, I wrote my own protocol based on what I learned from the laboratory, which is provided in annex 3 (page 126).

2.4.1.5 Gel Electrophoresis in Agarose

Gel electrophoresis is used to separate and analyze macromolecules and fragments based on their size and charge, including samples of DNA and RNA. During my internship, DNA and cDNA samples were analyzed by electrophoresis to observe the resulted products of the PCR reactions, as well as to assess the sample's quality and concentration. The fragment's migration can depend on the used buffer, concentration of the agarose gel, and molecular weight of the fragments. DNA fragments in this type of gel will migrate when subjected to an electric current so, before starting the experiment, the agarose gel should be prepared (protocol provided in annex 4, page 127). The percentage of the agarose gel impacts the distance between DNA bands of different lengths, meaning that before every run it should be considered the most adequate percentage of agarose having in mind the fragments to analyze. A bigger percentage is used to better separate small fragments of DNA, while lower agarose percentages are ideal to separate large DNA fragments. The bigger the percentages, the most time-consuming will be the run (Lee, Costumbrado, Hsu, & Kim, 2012). The used percentages were 2% and 1,5% (for fragment size between 50 – 300bp).

The gel was submerged in the electrophoresis equipment (figure 10) and a running buffer was added till the agarose gel was completely submerged. After that, samples were loaded into different wells. In the first well, it was added the ladder, a marker characterized by different fragments with a defined molecular weight, which allows for determining PCR fragments' weight by comparison with its fragments. In the other wells were added a mix of PCR samples with loading buffer.

The power supply was turned on after loading, and the run began. Because DNA is a naturally negatively charged molecule, it is attracted to the positive charge and will migrate from the negative electrode to the positive. When the run was about 2/3 of the way through the gel, the power supply was turned off. Finally, the gel was carefully

transported to UVITEC equipment (figure 11), a molecular biology imaging system with a software that allowed to record and capture the image of the gel.

For gel analysis, there are certain aspects that must be considered. Each fragment or mix of fragments in each sample might have a different molecular weight, which will be presented as a column of single and of moderate intensity. Usually, if the intensity of a band is higher than others it means that the DNA quantity with that molecular weight is also higher.

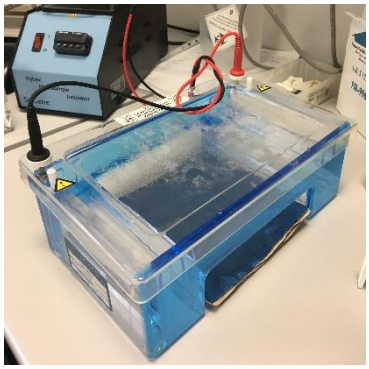


Figure 10 - Gel box immersed with a buffer, with a positive and negative power source.

Figure 11 - UVITEC Cambridge equipment for the visualization record of the agarose gel

One of the goals of electrophoresis in the developed research activities was to validate primers for use in Real-Time PCR. Because some of the target genes did not have specific primers previously reported in the literature, it was necessary to design them. To begin, all the designed primers must be validated to ensure that the Real-Time PCR reaction quantifies the target genes and that the results are credible. For that effect, a PCR reaction was performed, followed by electrophoresis, to validate the primers. The molecular weight of the PCR products was determined on the gel. PCR products that were observed to have a molecular weight consistent with the expectations were sequenced to confirm if primers were amplifying the target cDNA.

Figure 12 shows an electrophoresis run performed in a 1,5% agarose gel. Electrophoresis evaluated the primers for the *GSC* and *Lnc-C20orf84-1* genes. The *ACTB* gene (beta-actin) is a housekeeping gene, essential for cell survival and therefore always transcribed. Frequently, *ACTB* is used in Real-Time PCR runs as endogenous control. As observed, the run was made in duplicate for each gene. The presence of two bands for the *ACTB* gene, with the expected 110 bp, indicates that the *ACTB* primers were amplified,

therefore the run conditions were optimal. But the absence of bands for the *GSC* gene indicates that the primers didn´t amplify the fragment in the PCR reaction despite the run success. Some reasons can explain this event, for example, the annealing temperature of the reaction that might not be the best for this specific set of primers. Therefore, a temperature gradient PCR can be performed to test the best temperature of annealing. Other factors might include the design of the primers, which might not be complementary to the region target, the primers might be forming dimers between themselves, forming secondary connections, or the CG content might be too high. In addition, the *Lnc-C20orf84-1* primers amplified the target fragment, the resulting fragment has a molecular weight of only 50bp, which makes it a very light and soft band, very difficult to see.

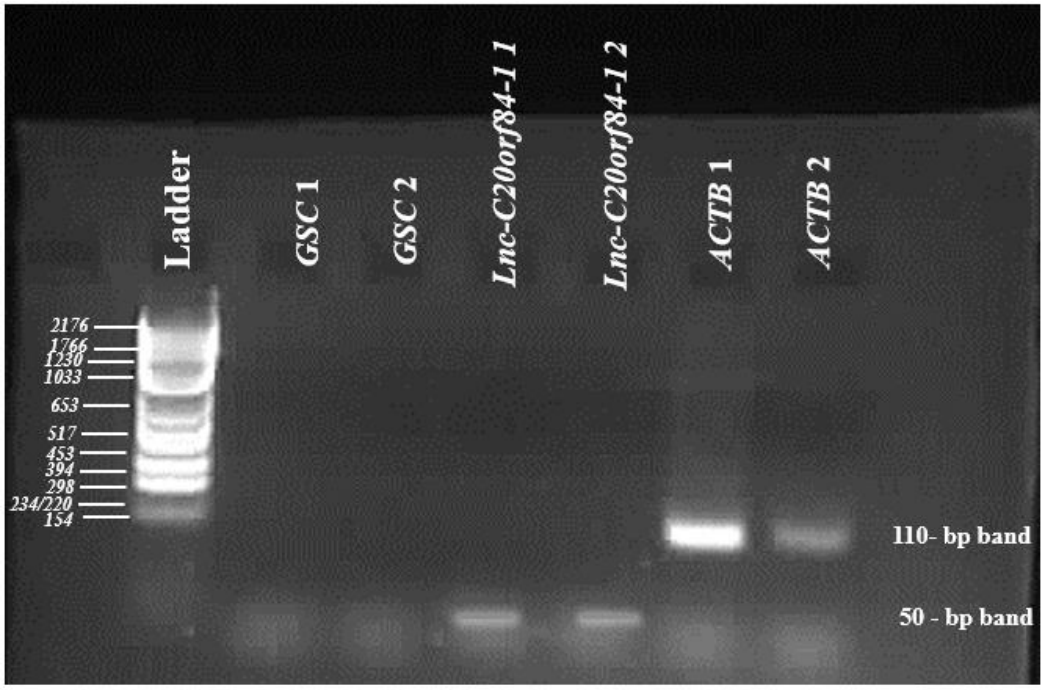


Figure 12 - Electrophoresis analysis on a 1,5% agarose gel; The ladder was obtained through the molecular weight marker VI of Roche, with known concentrations of DNA fragments; *GSC* 1 and *GSC* 2 are the same sample but done in duplicate. The same is applied to the *ACTB* and *Lnc-C20orf84-1* genes. Gel analysis shows that only the primers for the *ACTB* and *Lnc-C20orf84-1* were amplified.

2.4.1.6 Quantitative Real-Time PCR

Quantitative Real-Time PCR is a PCR-based method that allows the detection and quantification of specific sequences of nucleotides in live time. The advantage of Real-Time PCR remains in its ability to quantify PCR products as they are produced, whereas traditional PCR cannot quantify and can only detect the presence or absence of specific fragments at the end of the run using an end-point PCR method. In addition, Real-Time can be combined with RT-PCR, therefore the RNA target is directly loaded into the run. It can also be used to analyze single cells and quantify different targets such as DNA, mRNAs, and proteins (Garibyan & Avashia, 2013).

There are two types of Real-Time methods to detect and quantify the target product, sequence-specific DNA probes fluorescently labeled, or fluorescent dyes that bind to double-strand DNA (Garibyan & Avashia, 2013). In the experiment, the used method was the fluorescent dye, specifically the SYBR Green dye (protocol provided in annex 5, page 128). Because this dye attaches to double-stranded DNA preferentially, it will only bind to the chains being generated, and any Real-Time cycler equipment (the one used in the laboratory of my internship is in figure 13) can be used for this detection. Since the amount of PCR product synthesized is proportional to the amount of fluorescence emitted, Real-Time estimates the amount of DNA synthesized in each cycle. As a result, as the amount of PCR product in each cycle increases, so does the fluorescent signal intensity of each extension phase. Although it's not necessary any specific labeled probes in this method, all double-stranded DNA is going to be dyed, including nonspecific PCR products, such as primer-dimers (which also contribute to the fluorescent signal). So, it is crucial to assure previously that each set of primers is amplifying the target, and not forming dimers between themselves (QIAGEN Company, 2010).

Figure 13 - Station for performing Quantitative Real-Time PCR; LightCycler 480 Instrument II from Roche performs the Real-Time reaction, and data is visualized on the computer.

For the published article, Real-Time PCR was a very useful tool to evaluate the results of microarray assay (to confirm the genes that were found to be differentially expressed between fibroblasts of Sotos patients and fibroblasts of healthy controls). Gene expression in Real-Time PCR is measured by the number of cycles observed to obtain a threshold quantity of cDNA levels in each sample. The Ct (cycle threshold) is defined as the number of cycles required for the fluorescent signal to cross the threshold (to exceed the background level). Therefore, a higher number of Ct identifies the presence of a lower amount of target cDNA in the sample. cDNA from patients and controls were tested with specific primers for the differentially expressed genes and for the beta actin (housekeeping gene), which was used to normalize the measurement values. Gene expression values were assessed and normalized with the beta actin using the $2^{-\Delta\Delta Ct}$ method.

The $2^{-\Delta\Delta Ct}$ method is a simple formula used in Quantitative Real-Time PCR to calculate the relative fold gene expression of samples (Livak & Schmittgen, 2001). For its calculation, first the ΔΔCt is calculated. The calculation of the ΔCt consists in the difference of the threshold cycles and the target and reference genes:

$$\Delta Ct = Ct \text{ (a target gene)} - Ct \text{ (a reference gene)}$$

The ΔΔCT is the difference between the previous two ΔCT between the target and the reference samples, which is:

$$\Delta\Delta Ct = \Delta Ct \text{ (a target sample)} - \Delta Ct \text{ (a reference sample)}$$

Then, the $2^{-\Delta\Delta Ct}$ is calculated. This method's result is reported as the fold change in target gene expression for a specific sample compared to a reference sample, normalized to a reference gene. A disadvantage of this method is that assumes a uniform PCR amplification efficiency of 100% of all samples (Rao, Huang, & Zhou, 2013).

2.4.1.7 Automated Sanger Sequencing

Even though massive parallel sequencing revolutionized conventional sequencing, first-generation methods, like Sanger Sequencing, continue to be reliable methods. Particularly in genetic diagnosis, Sanger Sequencing continues to be a more suitable method for the sequencing of small fragments of DNA, and as an auxiliary method to confirm sequence variants found by massively parallel sequencing (Hagemann, 2015).

Sanger Sequencing was created in 1977, with the development of the chain-termination method. It makes use of two essential components, deoxyribonucleotides (dNTPs) and dideoxynucleotides (ddNTPs). Although both dNTPs and ddNTPs are monomers analogs of the DNA nucleotides, ddNTPs have a 3'hydroxyl group lacking, essential to continue the synthesis of the new strand (Metzker, 2005).

First, manual Sanger sequencing was introduced. A set of four PCR reactions are prepared in separated, where in each one is added a single type of radioactive labelled ddNTP (adenine, thymine, cytosine, or guanine). Each mixed is composed by one of the four ddNTPs, and regular dNTPs. During each Sanger reaction, ddNTPs and dNTPs are added randomly to the new DNA strand via complementary pairing to the DNA template. However, when a ddNTP is introduced, the new strand's extension is blocked due to the lack of a link between the ddNTP 3'hydroxyl group and the 5'phosphate of the next dNTP. Therefore, a high number of different length fragments are generated due to a block on the extension by the introduction of the random ddNTP. Moreover, for the determination of the DNA sequence, each product of the four PCR reactions is introduced in a different lane in the gel to allow the biologist to know which lane corresponds to each ddNTP (figure 14). The sequence of the DNA template is inferred by comparing the four lanes and the bands produced in each lane (Metzker, 2005), (Heather & Chain, 2016).

Figure 14 - Separation of size fragments by lanes as result of the four separated sequencing runs - manual Sanger sequencing. Each lane is identified only by one ddNTP (ddATP – adenine, ddCTP – citosine, ddGTP – guanine, ddTTP – timine).

Figure 15 - Separation of size fragments in the same lane as result of a single sequencing run by manual Sanger sequencing. Each lane is identified only by one ddNTP (ddATP – adenine, ddCTP – citosine, ddGTP – guanine, ddTTP – timine).

Later, fluorescently tagged ddNTPs were developed, which made it possible to do a single sequencing run using just one mix that contained both the four differently marked ddNTPs, and the normal dNTPs. Then, fragments were able to be runned in the same column on the agarose gel (figure 15).

The development of an automatic method allowed for a faster protocol, in which fragments of different sizes (terminated randomly with the four types of ddNTPS) are runned into a capillary sequencing equipment that separates fragments by size (from the shortest to the longest). Fragments of the same size form a band that is excited by a laser, and a detector do the fluorescent assembly. Due to the presence of the same labeled ddNTPs in the end of the fragments of a single band, it is identified the nucleotide base in that band position. A computer records the fluorescent bases of each band of fragments and compiles the data into an DNA sequence. The computer output results in a chromatogram, a graphic with the fluorescent peaks of each nucleotide's emitted fluorescent signal, as well as its DNA fragment length (Metzker, 2005), (Heather & Chain, 2016), (Clark, Pazdernik, & McGehee, 2019).

Sanger sequencing was performed in the research work for primers validation. Between the genes that were found differentially expressed, *NSO3* was one of the genes

that did not have any primers reported. Therefore, there was the need to design primers first. To validate primers, primers were teste in a PCR run and Sanger sequencing was performed on its PCR product (after verification by electrophoresis). Before the sequencing reaction, PCR products were purified to remove PCR reagents. The commercial kit used for the sequencing reactions was the BigDye Terminator v3.1 - Cycle Sequencing Kit, by Thermo Fisher Scientific. After sequencing, fragments were separated and analyzed on an ABI 3130XL Genetic Analyzer by Applied Biosystems (figure 16).

Figure 16 – Genetic Analyzer from Applied Biosystems; Equipment for genetic sequencing and analysis of DNA fragments

Results were analyzed in the Sequencing Analysis 5.3.1 software. The protocol used to perform the Sanger sequencing and prepare the PCR reactions for sequencing are available in annex 6 (page 129).

The sequencing data was uploaded into the Sequencing analysis software. To proceed to data analysis, the sequence data should be assessed for quality parameters and compared with the reference sequence. The Sequencing Analysis software performed an internal analysis of the quality of the run, but it must also be manually validated by the observation of the chromatogram. Each sequencing run was performed in separate for each primer, the forward and the reverse. Therefore, a chromatogram was generated for each primer (Crossley et al., 2020). For this work, runs were always done in duplicate, which means that we had four chromatograms (two for the forward primer, and two for the reverse primer) for each sequencing run. A very useful tool of the software used is its

ability to adjust the reading direction of the sequences derived from the reverse primers run, being easier to compare to the reference sequence.

For the chromatogram analysis, unreliable regions must be identified, usually located adjacent to the primer-binding sites. The program automatically labels the chromatogram peaks with the corresponding base. Manual review of the sequence chromatograms must check for overlapped peaks, which might indicate non-specific binding sites of primers or closely related genotypes (figure 17) (Crossley et al., 2020). Following an individual quality evaluation of each strand, the forward and reverse strands are combined and assessed as a single sequence.

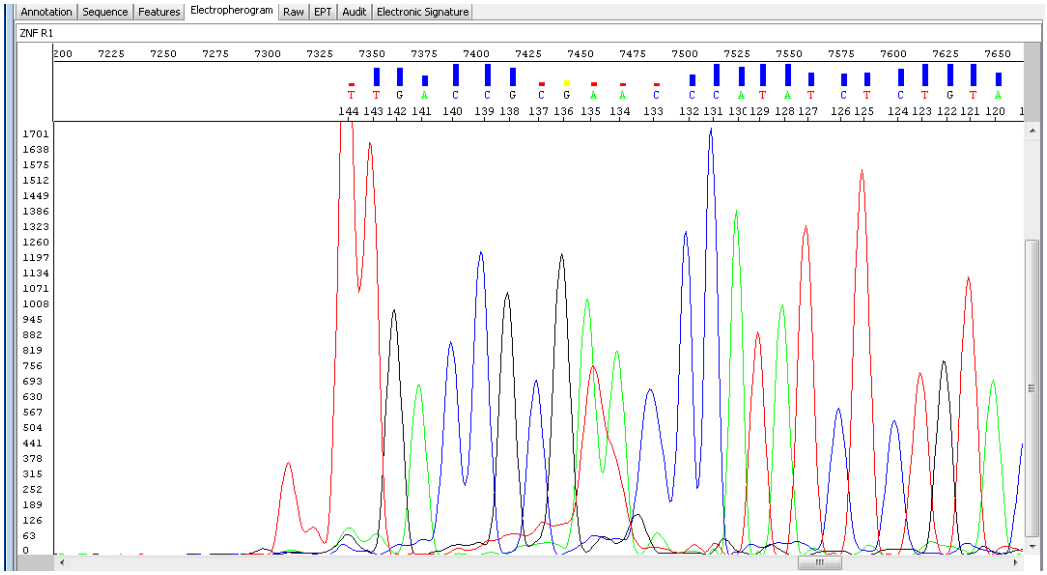


Figure 17 - Example of a chromatogram of the ZNF gene by the reverse primer, with the reading direction adjusted to the reference sequence. The first bases should not be considered for sequence analysis since is adjacent to a primer-binding site, and peaks are not consistent. The bars above each letter with a color (nucleotide bases) represent the quality of each peak.

Figure 18 displays part of the uploaded *NSO3* sequencing data in the software, with the forward primer. The quality assessment for the sequencing run was performed, with visual observation of the quality of the sequence of each strand. As observed, the sequencing run was well succeeded.

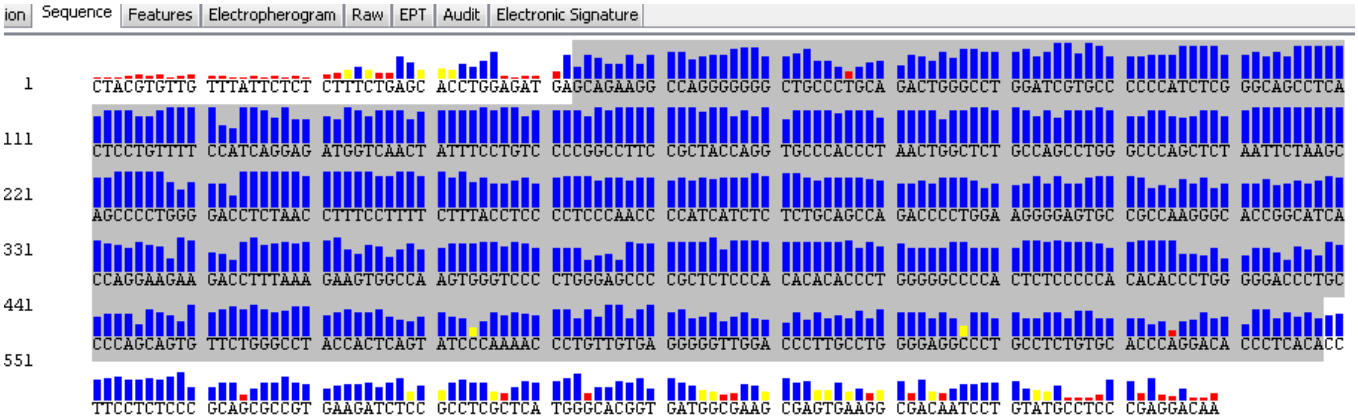


Figure 18 - Results of the sequencing data of the *NSO3* gene, with the forward primer, with the Sequencing Analysis program. In grey, the nucleotide bases that were compared to the reference sequence are highlighted.

By comparison between the reference sequence (figure 19) and the sequencing data (figure 18), with the forward primer, the nucleotides observed were successfully sequenced from the sequence run, with a good quality, and show the same nucleotide bases as the reference sequence. Therefore, after primers had been successfully amplified and verified by electrophoresis, sequencing results confirmed that the forward primer was specific for the target sequence. The reverse primer was also tested in the same run to validate the same outcome with the forward primer.

CCAGCTAGCCAAAGTCACCATcgtggaccaccacgccgccacggcctctt
tcatgaagcacctggagaatgagcagaaggccaggggggctgccctgca
gactgggcctggatcgtgccccccatctcgggcagcctcactcctgtttt
ccatcaggagatggtcaactatttcctgtccccggccttccgctaccagg
tgcccaccctaactggctctgccagcctgggcccagctctaattctaagc
agcccctggggacctctaacctttcctttttctttacctcccctcccaacc
ccatcatctctctgcagccagacccctggaaggggagtgccgccaagggc
accggcatcaccaggaagaagacctttaaagaagtggccaagtgggtccc
ctgggagccccgctctcccacacacaccctgggggccccactctccccca
cacaccctgggggaccctgccccagcagtgttctgggcctaccactcagt
atcccaaaaccctgttgtgaggggttggacccttgcctggggaggccct
gcctctgtgcacccaggacaccctcacaccttcctctcccgcagcgccgt
gaagatctccgcctcgctcatgggcacggtgatggcgaagcgagtgaagg
cgacAATCCTGTATGGCTCCGAGAC

Figure 19 - Partial *NSO3* gene sequence - reference sequence. In yellow are observed the nucleotide bases from the reference sequence that corresponds to the sequence data highlighted with grey in the sequencing data. In green is the sequence of the forward primer (5'CCAGCTAGCCAAAGTCACCAT 3') and in pink is the sequence that corresponds to the reverse primer (3'GTCTCGGAGCCATACAGGATT 5').

2.4.1.8 Manipulation of Fibroblasts Cell Lines

During my traineeship, I had the opportunity to observe the manipulation of human fibroblasts cell lines and iPSCs, performed by the principal researcher Ph.D. Giuseppina Conteduca. Cultured fibroblasts cells were used in the project to do the global gene expression profiling of fibroblasts from Sotos patients and healthy controls. Later, between the month of March and May (for over 8 weeks), I had the opportunity to develop my skills on fibroblasts cell culture lines (figures 20 – 24)

During this time, I learned important culture procedures, which increased my independence in the laboratory. These include the most basic care of how to work in a sterile environment, the preparation of complete and freezing medium, the establishment of cell culture lines from frozen cells, how and when to change mediums, assessing in the microscope the degree of confluence of the cells in the flask, how and when to split the cells through wash and trypsinization, and freezing/conservate cells.

The protocols that I performed are provided in annex 7 (page 130).

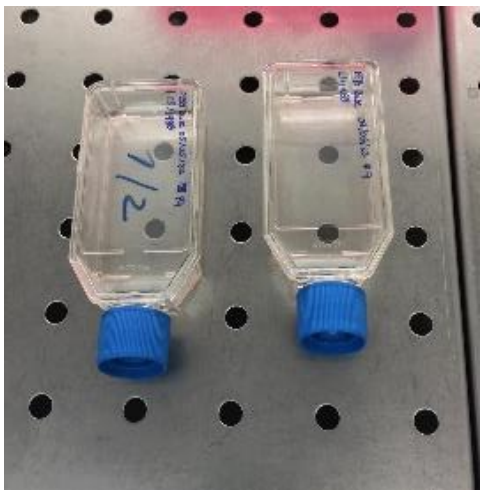


Figure 20 - Observation at the laminar flow chamber of two T25 flasks of cultured fibroblast in confluence.

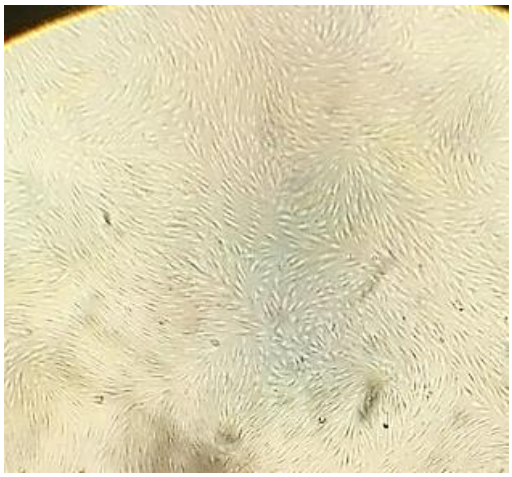


Figure 10 - Observation at the microscope of the confluence of the fibroblast cell line of one of the T25 flasks.

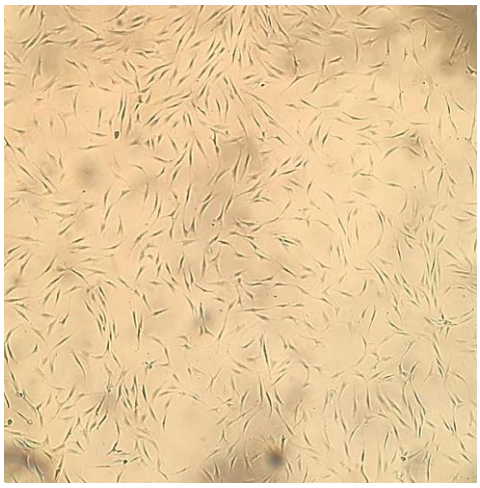


Figure 22- Observation at the microscope two days after the division of the confluent cells to the T25 flask. Cells are adherent, as it can be seen by its morphology, but in a small quantity.

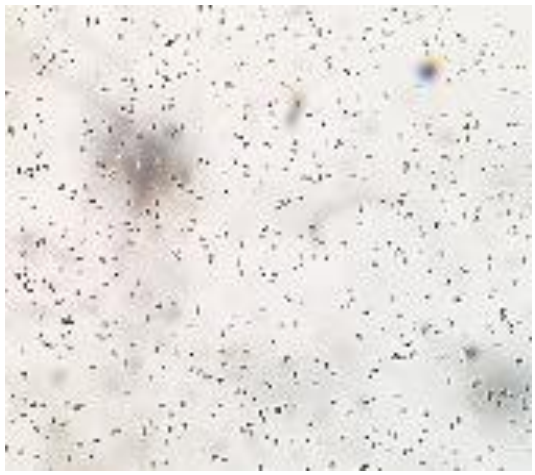


Figure 23 - Observation at the microscope of cells in suspension after a passage (trypsinization), cells are detached.

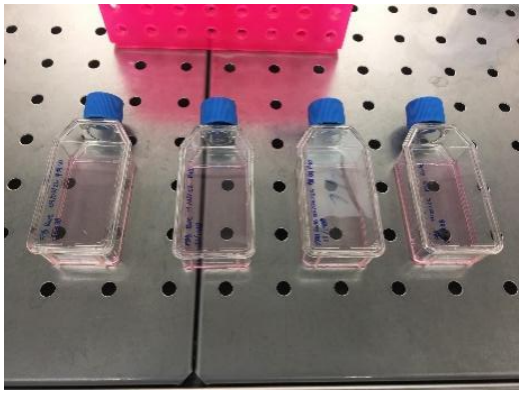


Figure 24 - Four recently divided flasks from two confluent T25 flasks, at this point cells are not adherent but rather suspended.

2.5 Generation of induced Pluripotent Stem Cells

The project's is also aimed at the generation of iPSCs from differentiated Sotos fibroblasts and healthy controls. Although this phase of the project has already started, only a few iPSCs colonies were created. After that, iPSCs will be differentiated into neural cells to constitute a cell culture model for the brain of Sotos patients. As intellectual disability is one of the most serious features of this disease, a therapy targeting the brain would be very beneficial in diagnosed Sotos patients to try to reduce the impact of Sotos *NSD1* variants in neuronal cells. This work aims to identify saRNAs

that target the NSD1 promotor, stimulating it, and reducing the unbalanced gene dosage (induced by the NSD1 haploinsufficiency).

Generally, for the generation of the iPSCs, fibroblasts cell cultures were established for each patient, and then a genetically modified viral vector, with the genetic information that encodes four transcription factors (Oct4/Sox2/cMyc/Klf4), was used to transfect the cells. The kit used was the CytoTune™-iPS 2.0 Sendai Reprogramming Kit.

Fibroblasts incorporated the modified vector into their cytoplasm and expressed the integrated transcription factors. One particular aspect of the vector used for the retroviral transfection remains in the fact that is zero footprints. Meaning that, unlike other genetically modified viruses, these ones will not integrate into the fibroblast's genome and so, do not leave traces.

These transcription factors are key genetic factors used for reprogramming somatic cells into iPSCs, by converting somatic cells from a terminally differentiated state to an embryonic state. Essentially, the difference between differentiated cells and iPSCs resides in gene activation. When cells commit to a pathway, they start to inactivate genes and be more specific in their cellular functions. That means that a differentiated cell may express fewer gene products than an iPSC because mechanisms of chromatin regulation tend to restrict gene expression. With this principle, cultured fibroblasts colonies were chosen and put to grow, based on the expression of essential transcription factors (because only cells expressing these factors are the ones who integrated the viral vector in their cytoplasm).

Cells were observed under the microscope and distinguished by the assessment of its morphology. Differentiated fibroblasts cells have a typical morphology, in a stellate structure, while iPSCs have a rounded shape. In addition, after iPSCs expansion, a PCR was performed, followed by electrophorese, to check if cells are vector free, because after viral infection it is expected that the plasmid leaves the cells. This happens because the only purpose of the plasmid was to stimulate the promotor of the key transcription factors. When cells are with no doubt iPSC, they are frozen in cryogen.

2.6 References

Ang, C. E., Trevino, A. E., & Chang, H. Y. (2020). Diverse lncRNA mechanisms in brain development and disease. *Current Opinion in Genetics & Development*, *65*, 42–46. https://doi.org/10.1016/j.gde.2020.05.006

Arcellana-Panlilio, M., & Robbins, S. M. (2002). I. Global gene expression profiling using DNA microarrays. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *282*(3), G397–G402. https://doi.org/10.1152/ajpgi.00519.2001

Bhat, A. A., Younes, S. N., Raza, S. S., Zarif, L., Nisar, S., Ahmed, I., . . . Uddin, S. (2020). Role of non-coding RNA networks in leukemia progression, metastasis and drug resistance. Molecular Cancer, 19(1). https://doi.org/10.1186/s12943-020-01175-9

Chen, J. J., Wang, S. J., Tsai, C. A., & Lin, C. J. (2006). Selection of differentially expressed genes in microarray data analysis. *The Pharmacogenomics Journal*, *7*(3), 212–220. https://doi.org/10.1038/sj.tpj.6500412

Clark, D. P., Pazdernik, N. J., & McGehee, M. R. (2019). DNA Sequencing. *Molecular Biology*, 240–269. https://doi.org/10.1016/b978-0-12-813288-3.00008-2

Comprehensive genomic characterization of head and neck squamous cell carcinomas. (2015). *Nature*, *517*(7536), 576–582. https://doi.org/10.1038/nature14129

Crossley, B. M., Bai, J., Glaser, A., Maes, R., Porter, E., Killian, M. L., . . . Toohey-Kurth, K. (2020). Guidelines for Sanger sequencing and molecular assay monitoring. *Journal of Veterinary Diagnostic Investigation*, *32*(6), 767–775. https://doi.org/10.1177/1040638720905833

Ettel, M., Zhao, L., Schechter, S., & Shi, J. (2019). Expression and prognostic value of NSD1 and SETD2 in pancreatic ductal adenocarcinoma and its precursor lesions. *Pathology*, *51*(4), 392–398. https://doi.org/10.1016/j.pathol.2019.02.005

Gabig, M., & Wegrzyn, G. (2001). An introduction to DNA chips: principles, technology, applications and analysis. *Acta Biochimica Polonica*, *48*(3), 615–622. https://doi.org/10.18388/abp.2001_3896

Garibyan, L., & Avashia, N. (2013). Polymerase Chain Reaction. *Journal of Investigative Dermatology*, *133*(3), 1–4. https://doi.org/10.1038/jid.2013.1

Hagemann, I. S. (2015). Overview of Technical Aspects and Chemistries of Next-Generation Sequencing. *Clinical Genomics*, 3–19. https://doi.org/10.1016/b978-0-12-404748-8.00001-0

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*(1), 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003

Hu, Y., Sun, H., Hu, J., & Zhang, X. (2020). LncRNA DLX6-AS1 Promotes the Progression of Neuroblastoma by Activating STAT2 via Targeting miR-506-3p. *Cancer Management and Research*, *Volume 12*, 7451–7463. https://doi.org/10.2147/cmar.s252521

Human gene NSD1. (2013). Retrieved August 3, 2022, from https://genome.ucsc.edu/cgi-bin/hgGene?hgsid=1419110693_0kpmuGRgOiVaJ46pVMP8MhgVJD3b&db=hg19&hgg_gene=uc003mfr.4&hgg_chrom=chr5&hgg_start=176560832&hgg_end=176727214

An improvement of the 2ˆ(-delta delta CT) method for quantitative Real-Time polymerase chain reaction data analysis. (2013). *Biostat Bioinforma Biomath*, 3(3):71–85.

Jaju, R. J. (2001). A novel gene, NSD1, is fused to NUP98 in the t(5;11)(q35;p15.5) in de novo childhood acute myeloid leukemia. *Blood*, *98*(4), 1264–1267. https://doi.org/10.1182/blood.v98.4.1264

Kadri, K. (2020). Polymerase Chain Reaction (PCR): Principle and Applications. *Synthetic Biology - New Interdisciplinary Science*. https://doi.org/10.5772/intechopen.86491

Kreil, D. P., Russell, R. R., & Russell, S. (2006). [4] Microarray Oligonucleotide Probes. *Methods in Enzymology*, 73–98. https://doi.org/10.1016/s0076-6879(06)10004-x

Kurotaki, N., Harada, N., Shimokawa, O., Miyake, N., Kawame, H., Uetake, K., . . . Matsumoto, N. (2003). Fifty microdeletions among 112 cases of Sotos syndrome: Low copy repeats possibly mediate the common deletion. *Human Mutation*, *22*(5), 378–387. https://doi.org/10.1002/humu.10270

Lee, P. Y., Costumbrado, J., Hsu, C. Y., & Kim, Y. H. (2012). Agarose Gel Electrophoresis for the Separation of DNA Fragments. *Journal of Visualized Experiments*, (62). https://doi.org/10.3791/3923

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Quantitative Real-Time PCR and the 2−ΔΔCT Method. *Methods*, *25*(4), 402–408. https://doi.org/10.1006/meth.2001.1262

Lucio-Eterovic, A. K., Singh, M. M., Gardner, J. E., Veerappan, C. S., Rice, J. C., & Carpenter, P. B. (2010). Role for the nuclear receptor-binding SET domain protein 1 (NSD1) methyltransferase in coordinating lysine 36 methylation at histone 3 with RNA polymerase II function. *Proceedings of the National Academy of Sciences*, *107*(39), 16952–16957. https://doi.org/10.1073/pnas.1002653107

Malewski, T., Malewska, A., & Rutkowski, R. (2003). RT-PCR technique and its applications.State-of the-art. *Journal of Animal and Feed Sciences*, *12*(3), 403–416. https://doi.org/10.22358/jafs/67719/2003

Mattick, J. S., & Makunin, I. V. (2006). Non-coding RNA. *Human Molecular Genetics*, *15*(suppl_1), R17–R29. https://doi.org/10.1093/hmg/ddl046

McPherson, R. A., & Pincus, M. R. (2021). *Henry's Clinical Diagnosis and Management by Laboratory Methods* (24th ed.). United States: Elsevier.

Metzker, M. L. (2005). Emerging technologies in DNA sequencing. *Genome Research*, *15*(12), 1767–1776. https://doi.org/10.1101/gr.3770505

Moradi Monfared, M.; Alizadeh Zarei, M.; Rafiei Dehbidi, G.; Behzad Behbahani, A.; Arabsolghar, R.; Takhshid, M.A. (2019). NDRG2 Regulates the Expression of Genes Involved in Epithelial Mesenchymal Transition of Prostate Cancer Cells. Iran. J. Med. PMCID: PMC6423436, PMID: 30936598.

Morey, J. S., Ryan, J. C., & van Dolah, F. M. (2006). Microarray validation: factors influencing correlation between oligonucleotide microarrays and Real-Time PCR. *Biological Procedures Online*, *8*(1), 175–193. https://doi.org/10.1251/bpo126

NDRG2 Regulates the Expression of Genes Involved in Epithelial Mesenchymal Transition of Prostate Cancer Cells. (2019). *Iran. J. Med. Sci*.

Novara, F., Stanzial, F., Rossi, E., Benedicenti, F., Inzana, F., di Gregorio, E., . . . Ciccone, R. (2014). Defining the phenotype associated with microduplication reciprocal to Sotos syndrome microdeletion. *American Journal of Medical Genetics Part A*, *164*(8), 2084–2090. https://doi.org/10.1002/ajmg.a.36591

Petersen, D., Chandramouli, G., Geoghegan, J., Hilburn, J., Paarlberg, J., Kim, C. H., . . . Kawasaki, E. S. (2005). Three microarray platforms: an analysis of their concordance in profiling gene expression. *BMC Genomics*, *6*(1). https://doi.org/10.1186/1471-2164-6-63

QIAGEN Company. (2010, July). Critical Factors for Successful Real-Time PCR. Retrieved from https://www.gene-quantification.de/qiagen-qpcr-sample-assay-tech-guide-2010.pdf

Rao, X., Huang, X., & Zhou, Z. (2013). An improvement of the 2ˆ(-delta delta CT) method for quantitative Real-Time polymerase chain reaction data analysis. *Biostatistics, Bioinformatics and Biomathematics*, 3(3), 71-85.

Rayasam, G. V. (2003). NSD1 is essential for early post-implantation development and has a catalytically active SET domain. *The EMBO Journal*, *22*(12), 3153–3163. https://doi.org/10.1093/emboj/cdg288

Rivera-Perez, J., Wakamiya, M., & Behringer, R. (1999). Goosecoid acts cell autonomously in mesenchyme-derived tissues during craniofacial development. *Development*, *126*(17), 3811–3821. https://doi.org/10.1242/dev.126.17.3811

Su, X., Zhang, J., Mouawad, R., Compérat, E., Rouprêt, M., Allanic, F., . . . Malouf, G. G. (2017). NSD1 Inactivation and SETD2 Mutation Drive a Convergence toward Loss of Function of H3K36 Writers in Clear Cell Renal Cell Carcinomas. *Cancer Research*, *77*(18), 4835–4845. https://doi.org/10.1158/0008-5472.can-17-0143

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., . . . Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., . . . Von Mering, C. (2020). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, *49*(D1), D605–D612. https://doi.org/10.1093/nar/gkaa1074

Tatton-Brown, K., & Rahman, N. (2006). Sotos syndrome. *European Journal of Human Genetics*, *15*(3), 264–271. https://doi.org/10.1038/sj.ejhg.5201686

Tatton-Brown, K., RP Cole, T., & Rahman, N. (2019, August 1). Sotos Syndrome. Retrieved October 15, 2021, from https://www.ncbi.nlm.nih.gov/books/NBK1479/

Tauchmann, S., & Schwaller, J. (2021). NSD1: A Lysine Methyltransferase between Developmental Disorders and Cancer. *Life*, *11*(9), 877. https://doi.org/10.3390/life11090877

Tidyman, W. E., & Rauen, K. A. (2010). Mutational and Functional Analysis in Human Ras/MAP Kinase Genetic Syndromes. *MAP Kinase Signaling Protocols*, 433–447. https://doi.org/10.1007/978-1-60761-795-2_27

Weier, H. U., & Gray, J. W. (1988). A Programmable System to Perform the Polymerase Chain Reaction. *DNA*, *7*(6), 441–447. https://doi.org/10.1089/dna.1.1988.7.441

Yehia, L., Keel, E., & Eng, C. (2020). The Clinical Spectrum of *PTEN* Mutations. *Annual Review of Medicine*, *71*(1), 103–116. https://doi.org/10.1146/annurev-med-052218-125823

Yehia, L., Ngeow, J., & Eng, C. (2019). PTEN-opathies: from biological insights to evidence-based precision medicine. *Journal of Clinical Investigation*, *129*(2), 452–464. https://doi.org/10.1172/jci121277

CHAPTER III

Laboratory of Clinical Genetics

Internship Report

Master's Degree in Laboratory Clinical Genetics

Rute Pinto

2021/2022

Chapter III

Laboratory of Clinical Genetics

3.1 Molecular Genetics

The molecular diagnostic activity in the Human Genetics Laboratory is generally focused on the detection of pathogenic single nucleotide changes, and large deletions or duplications, in prenatal as in post-natal (Harada & Korf, 2013). Each disorder has its own set of *guidelines* (instructions) regarding when to move forward to genetic testing and the appropriate testing methods that should be used to evaluate it. In this chapter, it will be reviewed the basic concepts of some genetic diseases, that were discussed, as well as the most adequate diagnostic methods for each clinical case. Additionally, the workflow of the NGS technology in clinical practice is also described. Basic concepts of several specific techniques, like PCR and Sanger sequencing, were already covered in the research chapter, therefore they are mentioned in this chapter but not described.

3.1.1 Historical Context

While cytogenetic techniques are focused on the study of large genomic changes, molecular techniques are focused on the study of the DNA sequence, particularly the single nucleotide changes that occur in it. In the past, the discovery of the double-strand DNA structure by Watson and Crick, after the laws of inheritance proposed by Mendel, raised the interest in the genetic field. The genetic code was ultimately discovered in 1966, with the understanding that a sequence of three adjacent nucleotides, which was defined as a codon, were sequence-coding for an aminoacid (Durmaz et al., 2015). The turning point in molecular genetics' history was the development of the PCR technique, in 1985, by Mullis and Smith. The generation of millions of copies from a single DNA sequence was shown to be a valuable resource for other molecular techniques. Although, only in 1989 did the scientific community recognize the advantages and utility of PCR in the molecular field. In clinical diagnosis, DNA sequences were amplified by PCR and used in technologies that required higher amounts of DNA. Initially, PCR was used together with the Restriction Fragment Length Polymorphism technique, to screen DNA

alterations (Durmaz et al., 2015). This technique follows the inheritance of a specific alteration of the DNA sequence on families, also known as genetic markers, through the generation of different size fragments by restriction enzymes ("Restriction Fragment Length Polymorphism (RFLP)," 2017). However, this method couldn't detect single-point DNA alterations, and it difficulted the clinical interpretation of genetic markers because the sequence of the target DNA was not known. Therefore, new methods were needed.

The introduction of the first method of sequencing, by Allan Maxam and Walter Gilbert, was the first step toward the determination of the DNA sequence, which would after enable the detection of single nucleotide changes in the target sequences. Although it was an effective method, the use of harmful chemicals together with its inability to read long PCR fragments, arose the necessity for a more convenient and safe method. Moreover, it was replaced by Sanger Sequencing, based on dideoxynucleotide chain termination method (Durmaz et al., 2015).

All these breakthroughs, as well as those that were still to come, accelerated the concretization of the Human Genome Project. In particular, the development of the shotgun sequencing method allowed for the end of the project before the expected time, accounting for about 20,000 to 25,000 genes in the genome covering about 93 percent of euchromatin. Shot-gun sequencing requires the break of the genome into fragments to be sequenced and, with the help of bioinformatic tools, fragments need to be overlapped and compared with a reference genome. (Durmaz et al., 2015).

Although Sanger sequencing is considered an efficient method, it is not cost or time effective. Moreover, a new methodology came to improve DNA sequencing to another level, the Next Generation Sequencing, also known as Massively Parallel Sequencing. It aims to generate large amounts of sequence data simultaneously, in a fast way, through a sequencing system. This technology allows to sequence the whole genome, exome, and RNA transcripts, and determine the methylation profile. However, NGS is more sensitive, less expensive, and more time-efficient than shot-gun sequencing. But in contrast, NGS is less accurate in comparison to Sanger Sequencing, because while NGS sequences millions of fragments at a time, Sanger only sequences a single DNA fragment at a time. Generally, Sanger sequencing keeps being a gold standard method to confirm the detection of pathogenic variants by NGS, or to target specific gene fragments for clinical diagnosis. (Durmaz et al., 2015).

With the development of new technologies and increasing knowledge, molecular genetics evolved at a high speed. Not long ago, in 2003, it was celebrated the 50th anniversary of the discovery of the double-strand DNA structure by Watson and Crick. While today, it is possible to access all personal genetic' information of an individual with the help of technologies like NGS, through whole-genome sequencing or exome sequencing, which allows the detection of many genetic factors.

3.1.2 Next Generation Sequencing on the Diagnostic Practice

The advances of technology in the last decade allowed us to use NGS to sequence the genome, known as Whole Genome Sequencing (WGS). However, in clinical practice, the target sequencing of all exons, known as Whole Exome Sequencing (WES), became a crucial diagnostic tool to characterize exon causing-variants, particularly in diseases with complex phenotypes. WES emerged as a cost and time effective method, relatively to WGS, because it enabled the reduction of the data generated, lowering the time of data interpretation as well as the time of clinical report emission (DeWitt, 2019). Considering that exons (DNA coding sequences) account for about 1% of the genome, and it is estimated that exon pathogenic variants are responsible for about 85% of genetic diseases, WES emerged as an innovative diagnostic tool (van Dijk, Auger, Jaszczyszyn, & Thermes, 2014).

Although WES can cover coding regions, exon-intron junctions, and Copy Number Variations (CNVs)[1], it cannot sequence introns. As a result, variants affecting RNA splicing might go unnoticed. Therefore, the use of genetic tests in the diagnostic practice should be considered for each case, individually, taking into account the most frequent variants causative of the suspected genetic disease. Targeting methods are used to reduce the amount of generated sequencing data, allowing to analyze the clinically relevant one.

The NGS workflow, in general, consists in library preparation, sequencing, and data analysis.

[1] Copy Number Variations are variations on the number of copies of a DNA between genomes.

3.1.2.1 Library Preparation for Targeting Sequencing Methods

First, library preparation will depend on the sequencing method, sequencing targeting methods can be by **hybridization capture** or **amplicon sequencing**. Hybridization capture allows more targets to be enriched and sequenced per panel, while amplicon sequencing has higher on-target rates, but at the cost of uniformity. Besides that, amplicon sequencing is a faster and simpler method, which requires smaller amounts of DNA (Samorodnitsky et al., 2015).

In the Human Genetics Laboratory, both methods of targeted sequencing are used. The search for associated-disease variants in a range of genetic pathologies is performed by hybridization capture. These include **custom panels** (sequence and analysis of a selected number of genes) associated with recurrent fever, genetic heart disease, and hemato-oncological diseases.

By contrast, many other conditions like short stature, gastrointestinal renal diseases, overgrowth syndromes, and ocular rare genetic diseases, are evaluated by ***in silico* panels** of WES. These *in silico* panels allow the sequencing of all genes, and its virtual selection by software analysis. An example of the NGS methodology with an *in silico* panel is described in page 59 of this report on a clinical case of Hypertrophic Cardiomyopathy.

On the other hand, the amplicon sequencing method is used in the laboratory for the identification of the most frequent variants in Thalassemia (page 181) and Cystic Fibrosis (page 72), through specific *Devyser* kits. Additionally, also the Non-Invasive Prenatal Screening (page 93) uses the amplicon sequencing method for library preparation of NGS relatively to chromosomes 21, 18, 13, X and Y. It is important to consider that the methodology for each genetic test may differ depending on the commercial kit that the company employs.

The two NGS technologies available for sequencing are the Ilumina Sequencing (by synthesis technology) and the Ion Torrent sequencing technology (by Thermo Fisher), but only Ilumina is used on the *Gaslini*'s Human Genetic Laboratory (Clark, Pazdernik, & McGehee, 2019).

➢ Targeting method – Hybridization capture

Hybridization capture, very briefly, starts with the preparation of the sequencing library. For this, genomic DNA is broken into small random fragments, followed by ligation with sequencing adapters (that function also as index[2] sequences). Index sequences allow samples from different patients to be sequenced together, while sequencing adapters allow sequences to bind to the flow cell of the Ilumina equipment. A PCR reaction is performed to amplify the ligated fragments. After that, samples are pooled together in a single tube. Then, long biotinylated probes (specifically designed for the regions of interest), are hybridized with the fragmented DNA. After hybridization, DNA is isolated through the capture of the biotinylated primers by streptavidin magnetic beads, due to the magnetic attraction between them. Non-specific molecules are washed away, leaving the targeted fragments isolated to be sequenced. Before sequencing, the pooled library is quantified by the TapeStation (Agilent) to evaluate the fragments size as well as its integrity (Targeted Sequencing Guide Handbook (RUO22-0863_001 4/22), 2022).

➢ Targeting method – Amplicon sequencing

First, genomic DNA of each patient is selectively amplified by primers that were specifically designed to flank the regions of interest. Each flanking set of primers amplifies a specific region by PCR, resulting in a group of specific fragments, named amplicons. To prepare amplicons for sequencing, a second PCR run is performed with the goal of ligating sequencing adapters, with unique index sequences in each amplicon (which allows samples to be sequenced together). Then, samples are pooled in a single tube, and purified. The purified sample pool is sequenced using sequencing by synthesis, and the resultant sequences are evaluated using a specific sequencing software.

By comparison with the hybridization capture method, amplicon sequencing is a method with less steps on its protocol, less time consuming, and cheaper. Also, while the number of targets for each panel is limited to about 10,000 amplicons, hybridization by capture can use virtual panels, which are unlimited (Targeted Sequencing Guide Handbook (RUO22-0863_001 4/22), 2022).

[2] Specific sequences that are used as barcodes to identify the genomic material of each patient

3.1.2.2 Sequencing Process

After library preparation, samples are loaded into the Ilumina Sequencing equipment (Clark, Pazdernik, & McGehee, 2019).

First, sample are loaded into a flow cell, and cluster generation begins. Fragments are on a lawn of surface-bound oligos that are complementary to the library adapters. Through bridge amplification, each fragment is converted into a cluster of identical sequences, called clonal cluster. Each cluster represents a small fragment of the genome. After cluster generation, templates are ready to be sequenced (Ilumina, 2017). The amplified sequencing libraries are sequenced through a sequence by synthesis method, where reversible terminators emit a fluorescent signal that is detected by a very sensitive optical sensor on the flow cell when are inserted into the sequencing strand. The emission is recorded and used to identify each base to create a read for each fragment, the output data is extracted in the form of FASTQ files (Ilumina, 2017), (Clark, Pazdernik, & McGehee, 2019). After that, NGS data are analyzed. A bioinformatic algorithm is used to categorize the fragments belonging to each patient.

During the bioinformatic analysis, an *in silico* panel can be selected to limit genes interpretation to a subset of genes that were reported in the literature to be relevant for a determined genetic pathology, as for example the clinical case described below of Hypertrophic Cardiomyopathy. The big advantage of the *in silico* panel method is that allows all exons to be sequenced and then, be filtered for the analysis accordingly to their relevance. This means that if no causative genes are found among the selected genes, analysis can be extended to other genes without the need to perform another sequencing run.

3.1.2.3 Analysis and Classification of Variants Detected by NGS for Report

Because the NGS technology produces a large amount of data, bioinformatic tools are needed to analyze and select nucleotide variants of interest. This variants classification will confirm or exclude a diagnosis based on the type of clinical indication, frequency of the variant in the population, and hereditary pattern. The attribution of clinical significance to a specific variant follows the most recent guidelines of the America College of Medical Genetics and Genomics (Richards et al., 2015), from 2015, which classified them in 5 classes: "benign", "probably benign", of "uncertain significance" (VUS), "probably pathogenic", and "pathogenic". The attribution of these classes is based on the presence or absence of many parameters, which considers the molecular consequences of the variant in the patient, the variant frequency on population, family segregation and its relevance in the clinical context (Tagliafico & Russo, 2021).

Parameters for variants interpretation are heavily dependent of databases, such as the Varsome, ClinGen, Intervar, Alamut Visual, among others. Particularly in the Human Genetics Laboratory, in clinical practice, diverse platforms are used to help to assist the generated data by the NGS technology.

For example, for custom and *in silico* panels by WES analysis, it is used for library preparation the kit from the SOPHiA genetics company (which can be prepared manually or automatically). This company includes the SOPHiA software platform, which uses the data extracted from Ilumina to do variant calling, annotation, prioritization, and filtering. After data, the output from SOPHiA software is manually analyzed by a biologist, accordingly to the frequency of the variant in general population, its classification through SOPHiA filters and ACMG guidelines, analysis of the coding sequence, exon-intron junctions, and CNVs of the selected genes, among others. After that, data is extracted in an excel sheet, and variants are selected by the biologist to be visualized with another program, the Alamut Visual software. The selected variants are the ones interpreted as VUS, probably pathogenic or pathogenic. Benign variants, those with high frequency on the population, are discarded because these ones are not included in the report. Through Alamut, variants can be better visualized and interpreted for reporting, with features that SOPHIA software doesn't have, such as the frequency of the variant in the population by

ethnicity, the effect of the variant on the domains of the protein (if it is on a conserved functional domain), among others.

Technically, what the biologist do in the Alamut program is to insert doubtful variants, found by SOPHiA in the genome, and do a prediction analysis. If in fact there is evidence that the variant is interpreted as pathogenic, the Alamut program allows the biologist to design specific primers that cover the exon of the pathogenic variant. However, if the Alamut identifies the variant as VUS, and all evidence suggest the same, including other databases like the ClinVar and the Human Gene Mutation Database (HGMD) database, and recent articles, this should also be included in the report (as seen in figure 25 and 26) and no primers need to be designed. The design of primers is crucial in diagnostic for confirmation of NGS results, to verify by Sanger Sequencing the presence of pathogenic variants since Sanger is considered to be a more accurate method to sequence smaller regions. After validation with Sanger sequencing, if a pathogenic variant was confirmed, this result is written in the report. The report is the final document returned to the physician who is in charge of following the patient.

COPIA CONFORME ALL'ORIGINALE

GASLINI

Area Aggregazione Servizi e Laboratori Diagnostici
U.O.C. Laboratorio di Genetica Umana
Certificazioni ISO9001:2015 e SIGU CERT
Direttore Dott. Domenico Coviello

Analisi molecolare in NGS

Genova 12/09/2022

Cognome e Nome: [redacted]
Codice Lab: [redacted]
Luogo e data di nascita: [redacted] Sesso:F

Inviato da: [redacted]
Ospedale: [redacted]
Indicazione all'indagine: Ipercolesterolemia Familiare
Data accettazione campione: [redacted]
Materiale esaminato: Sangue Periferico

Geni sequenziati:
Sono state esaminate, con profondità di analisi minima di 20X, la porzione codificante e le giunzioni esone-introne (5nt) dei seguenti geni:
APOB (NM_000384.2), APOE (NM_000041.4), LDLR (NM_000527.2), LDLRAP1 (NM_015627.2), PCSK9 (NM_174936.3) con copertura della regione di interesse di almeno 99%.

Risultati:
Varianti di significato patogenetico: presenza della variante c.1775G>A p.(Gly592Glu) nel gene LDLR allo stato eterozigote.
Varianti di significato incerto (VUS): vedere tabella a fine referto.
La porzione codificante dei geni analizzati risulta in assetto allelico corretto.

Conclusioni:
L'indagine eseguita ha evidenziato la presenza della variante c.1775G>A p.(Gly592Glu) nel gene LDLR allo stato eterozigote.
Tale variante, presente in dbSNP (rs137929307) con una frequenza nella popolazione generale europea pari a 0.012%, ricade in un residuo amminoacidico conservato, in un dominio funzionale della proteina; è classificata come patogenetica dai programmi di predizione ed è descritta come patogenetica in associazione a ipercolesterolemia familiare a trasmissione autosomico-dominante.
In base a quanto riportato si considera la variante identificata come patogenetica e la sua presenza compatibile con l'indicazione diagnostica.
Per l'interpretazione dei risultati di questo test genetico è appropriata una Consulenza genetica.
Si consigliano inoltre consulenza con lo specialista di settore e monitoraggio clinico.

Bibliografia:
Leren et al. Atherosclerosis 322 (2021) 61–66.
Meshkov et al. Genes 2021, 12, 66.
Sturm et al. JAMA Cardiol. 2021;6(8):902-909.

[redacted] [redacted] Pagina 1 di 2

Figure 25 - Part 1 of the report example; Clinical case of a patient referred for molecular testing due to a suspected Familial Hypercholesterolemia. Genetic test identified a variant (rs137929307 on dbSNP) on the *LDLR* gene in heterozygosis, classified as pathogenic, and with an autosomal dominant inheritance. Therefore, the clinical indication was coincident with the diagnosis for Familial Hypercholesterolemia. However, genetic test also identified a variant classified as VUS. It is recommended genetic counseling and monitoring by a physician specialized.

[redacted]

GASLINI

Area Aggregazione Servizi e Laboratori Diagnostici
U.O.C. Laboratorio di Genetica Umana
Certificazioni ISO9001:2015 e SIGU CERT
Direttore Dott. Domenico Coviello

Analisi molecolare in NGS

Genova 12/09/2022

Cognome e Nome: [redacted]
Codice Lab: [redacted]

Metodo di indagine:
Sequenziamento NGS mediante kit Custom Cardio Solution Sophia Genetics su piattaforma NextSeq (Illumina). Analisi dei dati mediante software Sophia.DDM (pipeline ILL1XG1G4_CNV_NextSeq_2) per l'identificazione di varianti nucleotidiche e grandi delezioni/duplicazioni (Copy Number Variations - CNV). Profondità di analisi almeno 20x.
Ai fini dell'analisi sono stati presi in esame esclusivamente i geni correlati all'indicazione clinica all'indagine.
Sequenziamento Sanger per la conferma delle eventuali varianti patogenetiche. Analisi dei dati mediante software SeqScape (v.2.6).
Sensibilità e specificità analitiche del sequenziamento Sanger >99%.
MLPA o RealTime PCR per le eventuali conferme di CNV.
Limiti del Test: l'analisi delle CNV è attendibile per delezioni/duplicazioni di più di un esone; la tecnica utilizzata non consente di identificare riarrangiamenti complessi (es. conversioni, inversioni, perdita di eterozigosi), espansioni di triplette, né mosaicismi somatici.
Database e programmi di riferimento: Human Reference Genome hg19, HGMD, Alamut Visual Plus (dbSNP151), Varsome, PanelApp, Clingen.

Note:
- Classificazione delle varianti identificate secondo linee guida (Matthijs et al., Guidelines for diagnostic next-generation sequencing. EJHG v.24 2016): (1) benigne, (2) probabilmente benigne, (3) di incerto significato (VUS), (4) probabilmente patogenetiche, (5) patogenetiche.
- Interpretazione delle varianti secondo le linee guida dell'ACMG/AMP (Richards et al. Genetics in Medicine volume 17, pages 405–423 (2015).
- Varianti classificate patogenetiche (5) o probabilmente patogenetiche (4), alla data della refertazione, vengono riportate nei Risultati e commentate nelle Conclusioni del referto.
- Varianti classificate benigne (1) o probabilmente benigne (2) alla data del referto, con frequenza uguale o superiore all'1% nella popolazione generale, non vengono riportate (disponibili su richiesta).
- Varianti VUS (3) con significato patogenetico incerto alla data della refertazione vengono riportate in tabella e non sono oggetto di commento; in particolare: per i geni a trasmissione Dominante in tabella vengono riportate Varianti eterozigoti con frequenza nella popolazione generale 0,1% o inferiore (Sinonime, Missenso, del/ins in frame e in siti splicing entro 5 nt); per i geni a trasmissione Recessiva in tabella vengono riportate Varianti con frequenza nella popolazione generale inferiore a 1% (Sinonime, Missenso, del/ins in frame e in siti splicing entro 5 nt); per i genotipi complessi in tabella vengono riportate Varianti con frequenza nella popolazione generale inferiore a 1% (Sinonime, Missenso, del/ins in frame e in siti splicing entro 5 nt).
- Le varianti potrebbero venire riclassificate nel tempo in base alle conoscenze scientifiche e a eventuali studi di segregazione.

Tabella delle varianti con significato incerto - VUS

Gene	Variante	Trasmissione	Zigosità	ID snp	Freq. nella popolazione generale (gnomAD)
APOB	c.655C>T p.(Arg219Cys)	AR/AD	Eterozigote	rs145661815	NFE 0,00077%

NOTA: Le varianti VUS non dovrebbero guidare le decisioni cliniche o la valutazione del rischio rs145661815 i membri della famiglia del paziente.

Dirigente del Settore [redacted]

Direttore [redacted]

Referto firmato digitalmente da [redacted]

[redacted]

Pagina 2 di 2

Figure 26- Part 2 of the report example; Clinical case of Familial Hypercholesterolemia identified a variant classified as VOUS (rs145661815 on dbSNP) in the *APOB* gene in heterozygosis. For the molecular study, an NGS custom panel for Familial Hypercholesterolemia was used to study the most clinically relevant genes, followed by Sanger Sequencing to confirm the presence of the pathogenic variant. However, the VOUS variant classification was supported by the ACMG classification and other evidence found in Clingen, PanelApp, HGMD, Alamut Visual Plus and Varsome. It is recommended that the VUS variant doesn't guide the clinical decision and have in mind that this classification is based on the present data.

3.1.3 Genetic Diseases and Clinical Cases

3.1.3.1 Hypertrophic Cardiomyopathy

Hypertrophic Cardiomyopathy (HCM) is a genetic disease with an autosomal dominant inheritance, caused by pathogenic variants in sarcomere-protein genes (or others) that change heart's structure and consequently, affect its contraction. Specifically, this pathology is characterized by left ventricular hypertrophy (wall thickness) unexplained by secondary causes and a nondilated left ventricle with preserved or increased ejection fraction of blood. Although HCM has a relatively benign course in most patients, sudden cardiac death is the most severe manifestation of the disease, particularly in adolescents and young adult athletes. When correctly identified in patients, manifestations of this disease can be prevented by the implantation of a cardioverter-defibrillator (Sun et al., 2017), (Basit, Brito, & Sharma, 2022), (Teekakirikul, Zhu, Huang, & Fung, 2019).

A single pathogenic variant in HCM causative genes is usually enough to induce the disease. However, a particular trace of this disease resides in its variable expressivity and incomplete penetrance that can be influenced by genetic and nongenetic factors (Basit, Brito, & Sharma, 2022). A very common example of this influence is the age-dependent expression of cardiac hypertrophy, some statistics point out that it may be a late-onset disease due to its prevalence, but then conversely, it also can account for 5% to 10% of the clinically diagnosed HCM cases in children (Ommen et al., 2020). Particularly the *MYH7* and *MYBPC3* genes, encoding the β-myosin heavy chain and myosin-binding protein C, are the two most mutated genes in HCM families, accounting for 50% of its cases. In 5% to 10% of cases, HCM is caused by variants in non-sarcomere genes (Sun et al., 2017), (Basit, Brito, & Sharma, 2022), (Teekakirikul, Zhu, Huang, & Fung, 2019).

The introduction of NGS technology have made gene-based diagnostics faster and more inexpensive, potentially shortening the diagnostic time and distinguishing HCM from phenocopies. This is crucial because HCM and its phenocopies have different approaches to disease therapy and prognosis. An HCM phenocopy is a phenotype that mimics HCM but it does not have the same cause as genetic HCM. Patients with a confirmed or suspected HCM diagnosis should undergo clinical genetic testing. If the index case tests positive, cascade screening tests in family members are recommended to

identify pathogenic or likely pathogenic variants carriers in the family and thereby, approach in a predictive way (Teekakirikul, Zhu, Huang, & Fung, 2019).

➢ Clinical Case

Case description: The present clinical case (figure 27) begins with the genetic testing of the proband (IV-1), born in 2014, that at one year of age was already showing symptoms of heart failure. The same was applied to her brother (IV-2), that was born in 2019, but none of its progenitors were manifesting symptoms. Family history showed that there were already two individuals studied in the family (II-2,3).

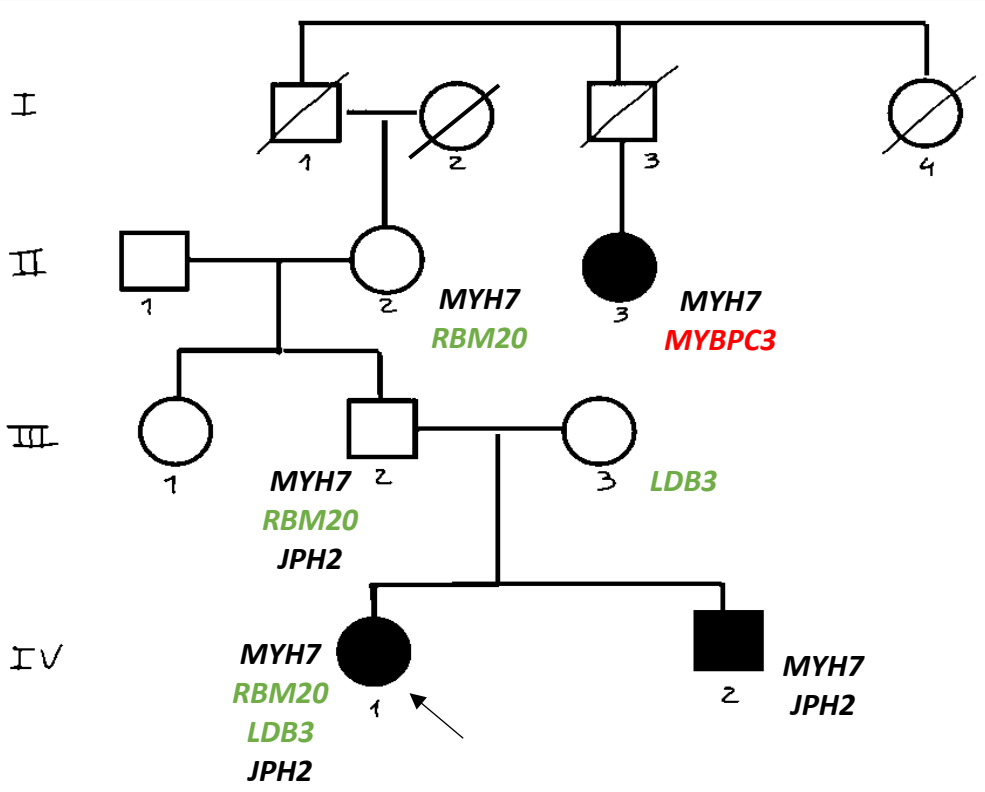


Figure 27- Pedigree of variant segregation on the P family; in **red** are highlighted genes with pathogenic, or likely pathogenic variants (*MYBPC3*); in **green** are highlighted genes with benign variants (*RBM20* and *LDB3*). In **black**, is highlighted a gene with a variant of uncertain significance (MYH7 and *JPH2*).

Family History

Genetic testing of the II-3 individual: Genetic testing of the II-3 individual, referred for genetic analysis for suspicion of HCM due to a severe HCM phenotype, identified one variant in each one of these two genes, *MYBPC3* and *MYH7*, the two most frequent genes causative of HCM. Specifically, the variant found in the *MYBPC3* gene was the c.1805C>G p.(Thr620Ser), classified as pathogenic, and in the *MYH7* gene was the c.5302G>A p.(Glu1768Lys), classified as VUS. This finding led to the study of the individual II-2, a direct-related family member of the II-3 individual. None of the individuals of the first generation (I-1,2,3,4) were studied because had already died.

Genetic testing of the II-2 individual: Genetic testing in the asymptomatic individual II-2 identified the same variant in gene *MYH7* (VUS) that was found in the proband, and an additional variant in the *RBM20* gene, the c.3373G>A p(Glu1125Lys), classified as benign. No more studies were performed in this family until the birth of the individuals IV-1,2.

Considering the family history and the IV-1,2 phenotypes, genetic testing for HCM was performed on the parents (III-2,3) and brothers (IV-1,2) for HCM.

Genetic testing of the parents (III-2,3): In individual III-2, it was identified the same two variants found in its mother (the individual II-2), a variant in the *MYH7* gene, and a variant in the *RBM20* gene. In addition, it was found a variant in the *JPH2* gene, the c.1975G>A p.(Ala659Thr), classified as VUS. In individual III-3 (the mother of IV-1,2), it was identified one variant in the *LDB3* gene, the c.352G>A p.(Val118Met), classified as benign.

Genetic testing of the individual IV-2: Individual IV-2 inherited two variants from its father (III-2), a variant in the *MYH7* gene, and other in the *JPH2* gene.

Genetic testing of the proband: Individual IV-1 inherited three variants from its father (III-2), in the *MYH7*, *RBM20,* and *JPH2* genes, and a benign variant from its mother in the *LDB3* gene.

Segregation and classification of variants in the family: Only the II-3 individual was a carrier of the pathogenic variant in the *MYBPC3* gene, meaning that there was no segregation of the mutation in the P family. Therefore, it doesn't add relevant information to the proband case, besides the need to perform screening test to family members.

On the other hand, the segregation of the VUS variant in the *MYH7* gene is observed in all offspring of carriers. This variant, whose frequency in the general population is unknown, resides in a functional domain of the protein. Therefore, this variant can impair the functionality of the protein. The *MYH7* variant c.5302G>A p.(Glu1768Lys) is classified as probably pathogenetic by computational programs and is described as probably pathogenetic in a subject with hypertrophic cardiomyopathy. However, as of the date of this report, in the ClinVar database and according to recent guidelines on the interpretation and classification of variants, the MYH7 variant gene is considered to be a VUS.

The c.1975G>A p.(Ala659Thr) variant in the *JPH2* gene, is not described in the literature. Resides in a functional domain of the protein, and pathogenicity prediction programs are conflicting regarding their classification. Based on reports and the guidelines on variant interpretation and classification (Richards et al., 2015) it should be considered a VUS.

As the *RBM20* gene variant as the *LDB3* gene variant are classified as benign because reports point out that these variants have a frequency of 1% or more in the general population.

Therefore, in overview, gene variants found among family members include:

- ***MYBPC3* - c.1805C>G p.(Thr620Ser) - pathogenic**
- ***MYH7* - c.5302G>A p.(Glu1768Lys), VOUS**
- ***JPH2* - c.1975G>A p.(Ala659Thr), VOUS**
- ***RBM20* - c.3373G>A p(Glu1125Lys) – benign**
- ***LDB3* - c.352G>A p.(Val118Met) - benign**

Case conclusion: In this family, the segregation of the *MYHC* variant is observed. Neither in individual IV-1 nor individual IV-2 was found a pathogenic variant that could explain their more severe phenotype relatively to other family members, such as heart failure at the age of one. Therefore, genetic heterogeneity was observed among family members, and these findings were not sufficient to explain the severe phenotype of individuals IV-1 and IV-2.

Recommendation: Genetic counseling is recommended for the interpretation of these findings, after appropriate cardiologic reevaluation in subjects carrying the described variants. Further testing is recommended, including exome analysis in the core family component to highlight possible disease-modifying genes that could explain the most severe phenotype.

Methodology: For this genetic investigation, the methodology used was the WES, which allows to sequence coding regions, exon-intron junctions, and identify CNVs (large deletions or duplications). An *in silico* panel of a subset of genes that were reported to be associated with HCM in literature, was analyzed. That means that, if necessary, additional genes can be studied without the need to perform another sequencing run. The following genes were evaluated from the *in silico* panel (figure 28).

ACTC1 (NM_005159.4), *ACTN2* (NM _001103.2), *ALPK3* (NM_020778.4), *BRAF* (NM_004333. 4), *CALR3* (NM_145046.4), *CAV3* (NM_001234.4), *CRYAB* (NM_001289808.1), *CSRP3* (NM_003476.4), *GLA* (NM_000169.2), *JPH2* (NM 020433. 4), *LAMA4* (NM 001105206.1), *LAMP2* (NM_013995.2), *MAP2K1* (NM_002755.3), *MYBPC3* (NM 000256.3), *MYH7* (NM 000257.3), *MYL2* (NM 000432. 3), *MYL3* (NM_000258.2), *MYLK2* (NM_033118.3), *MYOZ2* (NM_016599.4), *MYPN* (NM_032578.3), *NEXN* (NM_144573.3), *PDLIM3* (NM_014476. 5), *PLN* (NM_002667.4), *PRKAG2* (NM_016203.3), *RAF1* (NM_001354689.1), *SOS1* (NM_005633.3), *TCAP* (NM 003673.3), *TNNC1* (NM_003280. 2), *TNNI3* (NM_000363.4), *TNNT2* (NM 001276345.1), *TPM1* (NM 001018005.1), *TTR* (NM_000371.3), *VCL* (NM 014000.2).

Figure 28 – WES *in silico* panel for the molecular investigation of HCM

Other techniques are frequently used in the laboratory to confirm the detection of pathogenic variants through NGS. In this clinical case, Sanger sequencing was used to confirm the identified pathogenic variant. To confirm possible CNVs, Real-Time PCR can be used. The reference databases used in this laboratory for the analysis using NGS are the Human Reference Genome hg19, dbSNP151, HGMD, Alamut, Varsome, and PanelApp. Limitations of the NGS technique include the fact that does not allow the identification of somatic mosaicisms or complex rearrangements (e.g., inversions).

3.1.3.2 Fragile X Syndrome

Fragile X syndrome (OMIM #300624; FXS) is an X-linked dominant disorder that affects roughly about 1 in 4000 males and 1 in 5000 – 8000 females. It is known for its most common traces, like intellectual disability, and association with autism spectrum disorders (Garber, Visootsak, & Warren, 2008) The main cause of FXS is the expansion of the CGG triplets repeats in the *FMR1* gene, in its polymorphic region with abnormal hypermethylation of the promotor that leads to the silencing of transcription (Biancalana, Glaeser, McQuaid, & Steinbach, 2014).

Due to biological mechanisms, like random inactivation of chromosome X in females, the two sex are differently affected by this X-linked disease. From a clinical perspective, while all affected males can present features including IQ from normal, to mild and severe mental retardation, delay in language acquisition, and behavioral problems (often observed in about 30% of males with clinical features for autism), only 50% of affected females with FXS might develop mild to moderate mental retardation. Besides emotional and behavioral features in females being usually less severe, they can include other features like social anxiety, language deficits, and depression. On the physical aspect, features have been described but there are not concordant in all cases. Usually include mild dysmorphic characteristics (like large ears and coarse elongated face) and, particularly in males, an increase of testicular volume at least twice the normal on puberty (Garber, Visootsak, & Warren, 2008).

FMR1 gene

The *FMR1* gene is a protein-coding gene located at the chrXq27.3 and has 17 exons of length. The coding protein is an RNA-binding protein termed FMRP (Fragile X Mental Retardation Protein), involved in human development since the early stages of life. FXS is caused by the absence or deficiency of this protein, FMRP. More than 99% of cases of FXS are due to an expansion of CGG repeats in the 5' untranslated (UTR) region site on exon 1, and only in less than 1% of FXS cases there were reported single nucleotide variants including deletions and missense variants (Biancalana, Glaeser, McQuaid, & Steinbach, 2014).

CGG Repeats

Variants affecting this polymorphic sequence can change the stability of the CGG repeats, both in germ and somatic cells, leading to an expansion of the CGG repeats over generations. The CGG repeats' length are categorized in four different classes accordingly with its clinical significance. Due to the diversity of FXS cases found, sometimes clinical cases do not have a direct diagnostic, making the genetic counseling harder to practice.

Accordingly, to the 2021 revision of the American College of Medical Genetics (ACMG) (Spector et al., 2021):

- the **normal range** of CGG repeats, for the general population, is less or equal than 44 triplet repeats, with a normal phenotype (with 30 being the most frequent).

- Repeats ranging between 45 and 54 are **intermediate** (or "*grey-zone allele*") with a normal phenotype in the sense that it's not associated with disease and there are no expansions reported to a full mutation in one generation.

- **Premutation** alleles have a range of CGG repeats between 55 and 200, and are unstable and transmitted from parent to child. But typically, an eventual expansion only happens in maternal transmission.

- **Full mutation** alleles have a range of more than 200 CGG repeats and its widely variable among patients. Typically, there is partial or complete hypermethylation of the promoter, resulting in classic features of classical FXS.

In premutated alleles, usually, there is no hypermethylation of the promotor or association to FXS, although some cases have been reported with some features of FXS. Besides that, women carriers of premutation alleles should be considered since the allele is unstable and there is a risk of expansion to the next generation. In this context, it can be offered prenatal diagnosis for all pregnancies and testing of carriers for all family members at risk of having a premutation (Spector et al., 2021). Moreover, some ranges of CGG are still not well understood, like the "**indeterminate**" one. Studies pointed out that alleles in the range of 45 and 54 CGG repeats are not supposed to expand in one

generation, but a case was reported of an allele expansion of 52 repeats to a premutation of 56 repeats in one generation, and then to a full mutation in the next generation (Spector et al., 2021).

AGG Triplet Interruptions

Molecular studies showed that allele instability may not just be based on CGG expansions, but also on AGG triplet interruptions. It was observed that in normal alleles, after every 9 – 10 CGG, there was an AGG triplet with the aim to stabilize the sequence of CGGs. Also, in premutation alleles, it is observed that this range of expansion with no AGG interruptions has a higher probability of expansion to a full mutation in one generation than a sequence with a range of normal AGG interruptions. Interestingly, genetic instability is strongly affected by the sex of the transmitting parent and by the number of repeats and location of the AGG interruptions in the parental allele (Spector et al., 2021). This observation might be a crucial tool in genetic counseling since a test aimed at the AGG interruptions might help to predict the allele stability and probability of expansion in a carrier family Therefore, this information should be also included in the clinical report.

Fragil X-Related Disorders

The premutation alleles, in the range of 55 to 200 CGG repeats, contrarily to the full mutation, are not abnormally methylated. However, premutation alleles are causative of two other pathologies associated with FXS: **fragile X-associated primary ovarian insufficiency** (**FXPOI**) and **fragile X-associated tremor/ataxia syndrome** (**FXTAS**), with incomplete penetrance (Biancalana, Glaeser, McQuaid, & Steinbach, 2014).

FXPOI is defined by a widely broad group of ovarian dysfunctions that are associated with early menopause. About 20% of female carriers, of a premutation on the *FMR1* gene, will develop premature ovarian failure, and experience the end of menses before the 40 years landmark. While FXPOI affects women, **FXTAS** affects both genders. FXTAS is a late-onset neurodegenerative disease that is based on the progressive development of tremors and ataxia with progressive cognitive and behavioral deterioration. Carriers of a premutation have a higher risk to develop FXTAS with the

expansion of the CGG triplets up to 100 CGG and with increasing age (Biancalana, Glaeser, McQuaid, & Steinbach, 2014).

Size and Methylation Mosaicism

FXS patients may also exhibit diversity in their clinical manifestations due to somatic heterogeneity, also called "mosaicism". Mosaicism in X-fragile disorders can be due to the length expansion of the repeats - **size mosaicism -** as well as methylation status - **methylation mosaicism.** It is known that an allele methylated is more stable than an allele not methylated, so full mutations in adult tissues are more mitotically stable (with less probability of repeats expand to a premutation). Yet, in early embryos, the methylation process of the chromosomes is not completed until about the 13/14 weeks of gestation, making these tissues more susceptible to somatic instability during this period, allowing the arising of size mosaicisms. Individuals with size mosaicism can include mosaicism associating full mutated and premutated alleles (the most common mosaicism), mosaicisms with full mutated and normal alleles, and mosaics with the three types - full mutated, premutated, and non-mutated alleles (with normal alleles in less percentage). These last two types of mosaicism represent about 1% of FXS cases, and the phenotype can be less severe depending on the repeat length on the majority of cells (Biancalana, Glaeser, McQuaid, & Steinbach, 2014).

In addition, methylation mosaicisms can be found in large triplet expansions with incomplete or null methylation. Since there is no complete hypermethylation of the promotor, there might be an expression of the FMRP protein, leading to a phenotype less severe – the known "**high-functioning fragile X males**" - which are carriers of full mutations but not fully methylated in all cells. Meaning that they are not fully affected, may even have mild or null mental retardation (Biancalana, Glaeser, McQuaid, & Steinbach, 2014).

Diagnostic Methodologies for Genetic Testing

For testing FXS, more than just one method is necessary to evaluate all aspects of the *FMR1* full mutation. Clinical indications include the number of CGG repeats, the number of AGG interruptions, and the methylation status. Traditionally, laboratories used Southern blot analysis. PCR fragment analysis has the advantage of detecting the repeat

numbers in a fast and simple way, but its disadvantages is its difficulty on amplifying the regions rich in GC (Grasso et al., 2014). Although Southern blot has been considered the gold standard for FXS testing due to its ability to confirm repeat expansions and methylation status accurately using restriction enzymes, it was considered a very extensive and laborious technique for a diagnostic routine (Gu et al., 2021). The emergence of other innovative methods reduced the need to use Southern blots, such as the **Repeat-Primed PCR (RP-PCR)** and the **methylation-Polymerase Chain Reaction (mPCR)**. These two are the principal and more recent methods used in the laboratory of my internship, the RP-PCR for the CGG repeat sequence in the FMR1 gene promotor, and the mPCR for the evaluation of promoter methylation status of FMR1 gene.

❶ Genetic testing - Triplet Repeat-Primed PCR

RP-PCR uses gene amplification by two gene-specific primers (flanking the gene sequence), and a third CGG-specific primer (inside the repeated sequence) that binds randomly into the target sequence and helps to detect the expansion of the CGG repeats with more than 200 CGG, by comparison with a CGG ladder in an electropherogram. Also, it can detect the interrupting AGG sequences in the *FMR1* gene. The combination of the RP-PCR with high-throughput automated capillary electrophoresis results in an assay that can detect the presence or absence of an expanded *FMR1* allele with high sensitivity and specificity. Many studies already proved that indeed, RP-PCR has better results for testing FXS criteria when compared to traditional methods like Southern dot blot. There is already widespread agreement on the validity of this approach for diagnosing FXS (Gu et al., 2021). The commercial kit used in the human genetic laboratory to study *FMR1* in FXS is the "Amplidex FMR1 PCR – Asuragen – Diametra".

❷ Genetic testing - Methylation-Polymerase Chain Reaction

mPCR is used for the investigation of the methylation status of the *FMR1* gene promotor. This test is only indicated for patients in which was identified an expansion of more than 200 CGG, or intellectual disability, language delay, or autism. For patients with 200 CGGs, the methylation status of the promotor should be assessed to define whether it is a premutation (with promoter unmethylated) or a full mutation (with promoter methylated).

This method bases itself on the analysis of DNA treated with methylation-sensitive nucleases before *FMR1* gene-specific PCR. Overall, it uses a unique workflow with novel procedural controls, DNA treatment in separate control and methylation-sensitive restriction endonuclease reactions, labeled primer amplification, and two-color amplicon sizing by capillary electrophoresis. This method allows a higher resolution and analytical sensitivity for both size and methylation of mosaics when compared to Southern blot (Grasso et al., 2014). Particularly in the human genetic laboratory of my internship, DNA samples are analyzed for methylation status and CGG repeat length using the AmplideX FMR1 mPCR reagents (from Asuragen).

➢ Clinical Case

Case description: This clinical case (figure 29) arrived at the human genetics' laboratory from the Neuropsychiatry department due to a suspicion of FXS on the proband. The clinical indication was mild intellectual disability. Genetic test assessed the number of CGG triplets and AGG interruptions in the promoter region of the *FMR1* gene by RP-PCR and fragments separation by the ABI 3130XL Genetic Analyzer. Additionally, also the parents of the proband were studied.

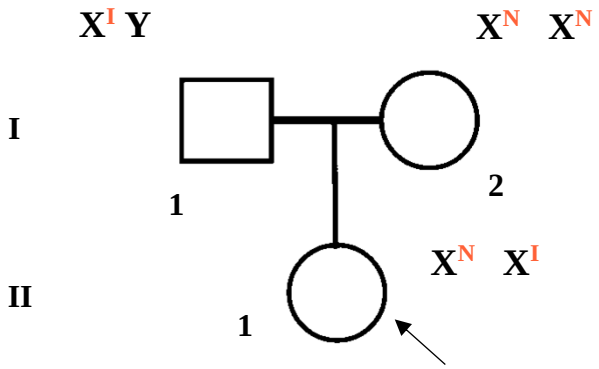


Figure 29 – Pedigree of the Z Family; I – Intermediate allele; N – Normal allele.

Diagnostic test results of the proband: The genetic amplification test on the proband showed one normal allele (with 31 CGG repeats), and an intermediate allele (with 46 CGG repeats), therefore the female patient is not a carrier of the premutation nor of the mutation responsible for the FXS. Genetic testing was performed in peripheral blood and, considering its clinical indication, an additional genetic test was required from a different tissue to discard mosaicism. Genetic test performed on saliva was concordant with the one from peripheral blood. It is known that CGG expansions within intermediate values are NOT causative of FXS, and do not have a higher risk to expand to a full mutation in the next generation. Furthermore, molecular testing found 2 AGG repeats that create 2 interruptions within the CGG region, which is a positive prognostic because it means that there is genetic stability. By contrast, the absence of AGG interruptions is associated with genetic instability, increasing the probability of CGG repeats expansion in the next generation.

Diagnostic test results from parents: Molecular testing showed that the I-2 individual is a carrier of two normal alleles, with 30 and 31 CGG repeats. However, molecular testing on the I-1 individual revealed that he is a carrier of an intermediate allele, with 46 CGG repeats. Therefore, her father passed to her his chromosome X with the same intermediate allele, and the patient's mother gave her the X chromosome with the normal allele. Having in mind that the patient's father gave her the intermediate allele and that the CGG repeats did not expanded in the next generation, we can communicate in the genetic counseling session that the prognostic is positive.

Considerations: Generally, if there are no specific information on chromosomal instability, during the genetic counseling session should be recommended to discuss reproductive options. It should be offered prenatal diagnosis if the couple wishes to get pregnant because there is the possibility that the father allele may expand from an intermediate allele to a premutation allele in one generation. Although it cannot be causative of FXS, it can higher the risk to develop FXS-associated pathologies such as primary ovarian insufficiency and tremor/ataxia syndrome. In the present clinical case, there is no need to discuss reproductive options if the couple wishes to get pregnant because chromosomic stability was already studied (the AGG interruptions), and it was found to be stable.

3.1.3.3 Cystic Fibrosis

In Caucasian populations, cystic fibrosis (MIM #219700; CF) is one of the most frequent genetic conditions with an autosomal recessive inheritance. It is caused by pathogenic variants in the Cystic Fibrosis Transmembrane Conductance Regulator (*CFTR*) gene. Classic CF features are characterized as multisystemic, with an impact on the pulmonary, digestive, and reproductive systems, as well as on the sweat glands (Deignan et al., 2020).

From a physiological perspective, the loss of function of the CFTR protein results in the production of mucus abnormally thick and viscous in the epithelial tissue. Its consequences are severe and sometimes irreversible as it can affect many organs. For example, in the lungs causes mucus build-up in the airways, limiting mucociliary clearance, and leading to chronic infection due to the hostile environment for bacterial infections. In the reproductive system, more than 95% of males with CF are infertile (Ong et al., 2017). Nonclassical features of CF can include Congenital Absence of the Vas Deferens (CAVD), a male clinical condition in which the tubes that conduct sperm from the epididymis to the outside fail to develop properly. CAVD can be also the result of males with pathogenic variants in the *CFTR* that usually do not present classical features of CF. Therefore, the genotype-phenotype correlation can be very diverse (Deignan et al., 2020), (Bareil & Bergougnoux, 2020).

CFTR Gene and Variants

CFTR is a protein-coding gene, located in chromosome 7q31.2 and is composed by 27 coding exons ("UniProt - Q20BH0 · Q20BH0_HUMAN," 2022). The protein encoded by this gene works by ATP hydrolysis and transfers negatively charged chloride ions across the membranes of cells that secrete mucus, sweat, saliva, tears, and digestive enzymes. This channel is an important regulator of fluid homeostasis due to epithelial ion and water transport. Meaning that allows the production of essential mucus, which is used to lubricate the membrane of many organs and tissues, including the lungs and pancreas (Deignan et al., 2020).

Different types of pathogenic variants in the *CFTR* gene can affect the coding protein in different ways, resulting in a broad phenotypic spectrum. Over 2000 variants

have been identified in the *CFTR* gene with diverse effects on the coding protein, including the impair of synthesis (mRNA and protein), maturation, trafficking, and channel activity. The phenotype of each CF patient is the outcome of the combination of two, or more, variants. But when these variants are combined, they can lead to more severe clinical features (Bareil & Bergougnoux, 2020).

The frequency of the CF-causing variants can diversify a lot accordingly to the geographic and ethnic origin. For example, the most common CF variant reported is the p.Phe508del. However, while in Brittany represents about 81%, in Turkey it is only responsible for 24,5% of CF variants. Of all variants, about 98% are small deletions or indels along the gene that lead to loss of function, the other 1-2% involve large rearrangements. Nonsense, frameshift, or splicing variants are classified as CF-causing because of their presumed severe consequence. While variants considered not pathogenic are the ones with higher frequency in the general population. In addition, most nucleotide alterations, on the other hand, are frequently missense variants whose molecular consequence and clinical spectrum are difficult to anticipate (Bareil & Bergougnoux, 2020).

CF assessment is carried out in many situations, including molecular diagnosis in postnatal, prenatal diagnosis, and carrier testing. A screening test was developed with the aim to identify the most frequent CF-causing variants as a first-tier screening in the national health system programs to all newborns. A second tier *CFTR* analysis can cover all coding *CFTR* regions of the gene, allowing to reveal of new variants hard to predict in the phenotype (Bareil & Bergougnoux, 2020). The ACMG in 2020 (Deignan et al., 2020), recommended to evaluate at least 23 variant-causing CF as the minimum panel to be included in the pan-ethnic carrier screening of individuals with no family history of the disease. Other CF variants should be specifically searched in specific populations for optimal screening. For example, in Italy, there is available a core panel of variants that are more frequent in the Italian population.

Indications for *CFTR* Variant Testing

Accordingly, to the American College of Medical Genetics and Genomics recommendations of 2020 (Deignan et al., 2020), indications for molecular confirmation of clinical CF diagnosis include children positive at the newborn biochemical screening, children with meconium ileus, idiopathic pancreatitis or bronchiectasis, and males with CAVD. In prenatal, testing can be performed depending on gestational age, family and personal history, ethnicity, and patient preferences, through amniocentesis or chronic villus sampling. Individuals with an alarming CF family history or partners with an alarming CF family history, males with CAVD, gamete donors, and to reproductive-age women may be offered carrier testing. Since CF is a recessive disorder, carriers of one pathogenic, or likely pathogenic, are expected to be asymptomatic. For this reason, couples that plan a pregnancy could carry out screening tests to be sure that both partners are carriers, and in this case identifying higher-risk pregnancies for CF. Tests can be carried out on the two individuals, simultaneously, or sequentially, typically first women and then male is tested only if the female is a carrier of a variant.

CF Diagnosis

Overall, the diagnosis of CF can be assessed by 3 essential observations: a) the existence of a clinical picture compatible with the disease or positive neonatal screening, b) a confirmed dysfunction of the CFTR channel, and c) the identification of two pathogenic variants in *trans* (Società Italiana Per Lo Studio Della Fibrosi Cistica, 2019).

The classical used method, and considered a gold standard in medical practice, is the sweat test (table 5), which measures the chloride concentration of the sweat. In a typical CF patient, values of chloride are above 60 mmol/L due to the loss of function of the CFTR protein in the sweat glands, leading to an increased concentration of chloride due to loss of ion homeostasis. Usually, there is no necessity to carry out more tests if the sweat test came back positive, unless the CF causing-variant determination is useful to predict risks, to other family members, or to determine the pharmacological strategy. An additional test can be used when the diagnosis is doubtful, that aims to detect differences in the electrical potential in the nasal mucosa (Società Italiana Per Lo Studio Della Fibrosi Cistica, 2019).

Sweat chloride test results	
Chloride level, mmol/L	**Result**
60 or greater	Confirmed CF diagnosis
30 to 59	Unclear CF diagnosis: further testing needed
Less than 30	Unlikely to be CF

Table 5 - Sweat chloride test results for CF diagnosis

CF Diagnosis - Molecular Testing

To identify *CFTR* gene sequence variants, a wide range of methods can be used (Dequeker et al., 2008). The choice of technique to evaluate CFTR variants needs to be considered accordingly to, for example, family history. If there is a known carrier of a variant in the family, then the method chosen should aim to look specifically for that variant.

Accordingly, to the Italian Society for the Study of Cystic Fibrosis in the 2019 consensus, there is a general consensus about the 4 levels (I-IV) of molecular tests for CF diagnosis that aim to detect CF-causing variants in a logical way.

- Level I - aim for molecular tests that search for the most frequent pathogenic variants of the *CFTR* gene. Such as a NGS panel for the most frequent CF variants.

- Level II - analyze the *CFTR* gene sequence, each exon, exon-intron junctions, promoter, and untranslated region where pathogenetic variants have been identified, to characterize and identify a higher number of pathogenic variants, whether rarer or still unknown. In these, are included the direct sequencing by Sanger method or NGS.

- Level III – search for large rearrangements, like macrodeletions and macroduplications on the *CFTR* gene, including those not detectable in level I or II. In this case, it's more used the Multiplex Ligation-Dependent Probe Amplification (MLPA), which detects gene CNVs.

- The last is the level IV, and it's not commonly used in diagnostic but manly for research use. At this level, the patient's RNA is translated into cDNA to

try to identify unknown pathogenetic variants that interfere with the mRNA splice function of the *CFTR* gene, such as intronic pathogenetic variants.

Molecular diagnosis can be performed with different technologies accordingly to the laboratory, and the number of patients analyzed. I will describe the two most used methods: the PCR Multiplex Reaction and Fragment Analysis, and the molecular diagnosis with NGS technology.

❶ PCR Multiplex Reaction and Fragment Analysis

In the Human Genetics Laboratory, one of the methods for the analysis of level I is the Devyser CFTR Core/Italy kit. The core kit is designed having in mind the most frequent CF causing-variants in the European population, while the Italian kit is designed for the most frequent variants-causing CF in the Italian population. This is an *in vitro* diagnostic test, designed for qualitative genotyping of a panel of normal and abnormal alleles in the *CFTR* gene ("Cystic fibrosis | Devyser," 2021).

The Devyser kit is based on an allele-specific multiplex PCR that generates fluorochrome-labeled fragments of different sizes for each specific variant. Following the PCR reaction, fragments are analyzed by capillary electrophoresis.

The analysis is performed with two separate amplification mixes (Core/Italy 1 and Core/Italy 2). Core/Italy 1 detects the non-mutated alleles (normal alleles) in the patient. The corresponding mutated alleles in patient samples are detected in Core/Italia 2. To verify whether the detected mutation is in homozygosity or heterozygosity, the sample should be tested with both mixes (Core 1 and 2). In addition, are used three types of controls: a known normal control to serve as an allelic ladder, a control template-free to check for possible contaminations, and a sample control with known variant to assure the quality of the run. This method is used on DNA samples extracted from peripheral blood and, in the case of prenatal diagnosis, from chorionic villi or amniotic fluid.

Analysis of the Devyser CFTR Peaks

In an individual with no *CFTR* variants, the normal alleles are detected in Core 1, because detects the unaffected sequence, while the normal alleles are absence in Core 2 because it detects the presence of *CFTR* variants. However, this method only looks for the most common variants, if the patients have a clinical suspicion of CF, a level II test should be performed to look for fewer common variants.

Variants in heterozygosity can be detected if the peak is present in both, Core 1 and Core 2 peaks. This means that one allele has that specific variant, and the other allele does not have that same specific variant. Variants in homozygosity can be detected if the peak of the normal allele in Core 1 is absent, while the corresponding allele in Core 2 is present.

A peak analysis of a clinical case is shown below (figure 30 and 31), both Core 1 and Core 2, show simultaneously the presence of a peak in the *CFTR* allele for the variant G542 (in Core 1), and G542X (in Core 2). Additionally, it's clearly observed a decrease in the peak size on Core 1, meaning that on a qualitative assessment, about 50% of the fragments were amplified for that variant, and about 50% were amplified in Core 2. Therefore, the G542 c.1624G>T variant, in exon 12, is present in heterozygosity. Accordingly, to the Devyser protocol, the fragments length was the expected.

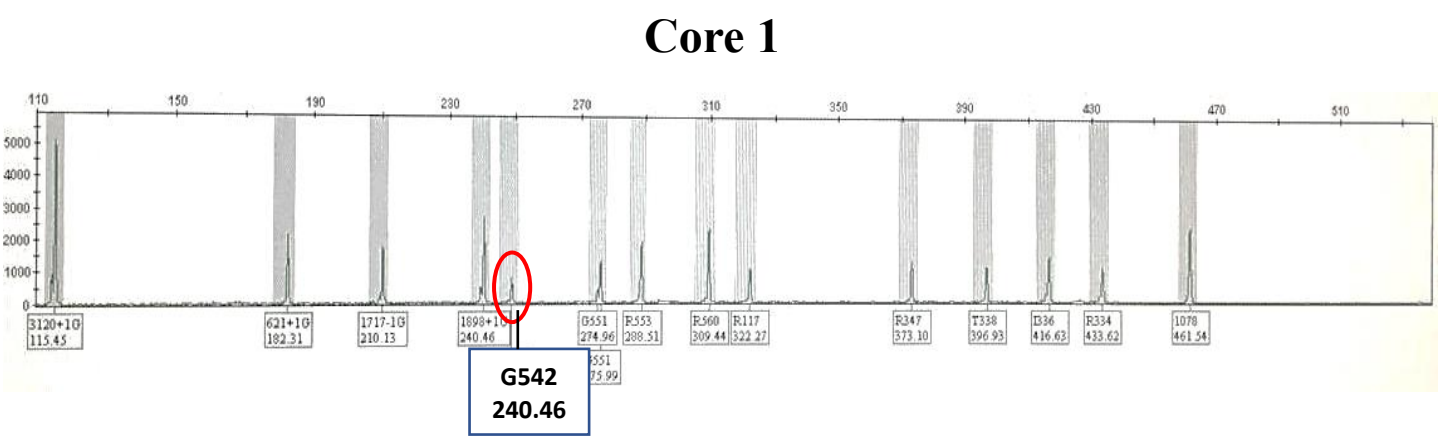


Figure 30 - In Core 1 mix, PCR fragments without frequent CFTR variants (normal). PCR fragments without frequent CFTR variants (normal alleles) are observed as electrophoretic peaks. It's highlighted with a red circle a peak that shows a significant reduction of the peak size in the G542 allele, in comparison to the reference sample. Below is indicated the fragment length (240.46).

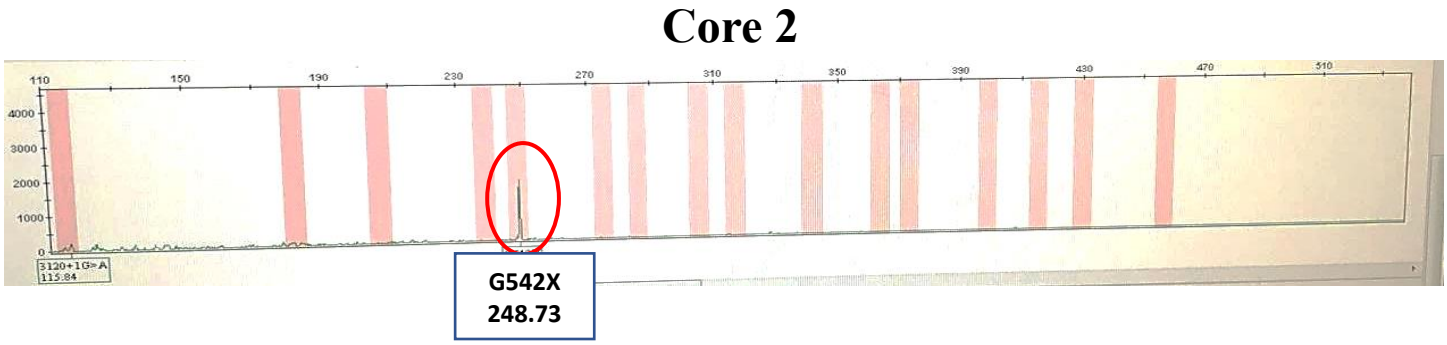


Figure 31 - In Core 2 mix, PCR fragments with specific CFTR variants can be observed as electrophoretic peaks. It's highlighted with a red circle a peak that is the result of the hybridization between the probe with the G542X mutation, and the sequence with the G542X mutation. Below is indicated the fragment length (248.73).

For the assessment of the clinical significance of this variant, the ClinVar database was consulted at 19/05/2022. The database classified the G542X variant (c.1624G>T) as nonsense pathogenic, meaning that the substitution of the guanine base in the position 1624 for a thymine result in the alteration of the sequence to a stop codon, which results in a shortened protein that is nonfunctional. Its search on the CFTR2 database indicates that when combined with another CF-causing variant causes typical CF features.

❷ Molecular Diagnosis with NGS Technology

To identify rare variants, not included in the PCR multiplex kit, it is possible to use the NGS Devyser kit for *CFTR* gene analysis ("NGS: Changing the game in CFTR analysis | Devyser," 2021). The Human Genetic Laboratory is still in a transitional period, employing variant detection for the first level of diagnosis using the multiplex PCR reaction, but at the same time is starting to integrate the Devyser kit with NGS in clinical practice.

Through the NGS technology we will get the sequence of all *CFTR* gene, enabling to identify the same frequent variants as the ones from the PCR kit but it also allows to read of the complete coding sequence of the gene. Variants can be detected from DNA of whole blood or dried blood spot (laid in spots and stored dried on Guthrie cards that come from the laboratory of " Neonatal Screening").

This method allows the choice of gene analysis according to two levels of analysis. The first level, in which the software "filters" the analysis of the 362 known variants, and

the detection of eight most common CNVs. If necessary, the analysis is extended to the second level, which allows the complete analysis of the coding sequence and exon-intron junctions of the gene, and part of the promoter region. It also offers a quantitative analysis of CNVs of the whole gene for the detection of deletions and duplications.

Only variants identified and classified as pathogenic, or probably pathogenic, are confirmed by Sanger sequencing or MLPA on the patient sample and, if available, on parental DNA.

➢ Clinical Case

Case description: A patient was referred for genetic diagnosis due to a clinical picture compatible with CF and abnormal sweat chloride levels. Molecular tests revealed that the patient was a compound heterozygous for two CFTR variants, the most common one (F508del) and the G542X variant. In the light of recent knowledge, CF patients carrying one or two F508del alleles are indicated to be treated with CFTR modulators, the Elexacaftor, Tezacaftor, and Ivacaftor. These compounds can correct, partially, CFTR defects and rescue protein function, depending on the effect of the variant in the protein biogenesis. Patient was treated with the combination of these three modulators.

Although the patient was considered eligible for the treatment with modulators, it failed to obtain any beneficial effects with the treatment, controversially to what would be expected from his genotype. In order to better understand the lack of response of the modulators on the CFTR protein, its activity, and pharmacological rescue were evaluated in the patient's nasal cells through electrophysiological techniques. The *ex vivo* experiences tested the response of the nasal cells to the combined treatment, which confirmed the lack of response to the medication. The next step was directed to the sequencing of cDNA products, from the patient's differentiated epithelial cells, spanning the two already known present CFTR variants, F508del and G542X from exon 9 to 14. Because standard molecular methods search for the most common mutations, other possible variants *in cis* are not directly searched, some of which can have an impact on the physiopathology or, as in this case, treatment response. To assess if there were variants *in cis*, the whole *CFTR* gene was sequenced. Results revealed that the patient

was carrying a third substitution, the L467F variant *in cis* with the F508del variant (figure 32), which could explain the lack of response to therapy.

The same was described on a compound heterozygous patient, for both F508del and E585X variants, also referred for modulators treatment due to the presence of the F508del variant. Patient did not responded to therapy. An additional molecular test showed the same complex allele as the above patient, with the L467F and F508del variants *in cis*.

The reason for lack of response to the therapy is not particularly described, but generally, the reason is associated with a combination of genetic factors of each patient.

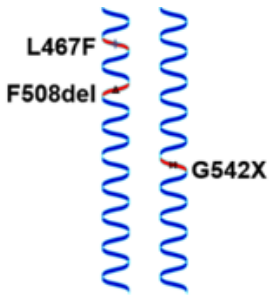


Figure 32 - Patient's genotype containing the three causative-CF mutations - two heterozygous *CFTR* variants in *trans*, F508del and the G542X variant, and a variant in *cis* with the F508del variant (the L467F variant). Retrieved from Sondo et al. (2022).

Considerations: This clinical case was retrieved from an article (Sondo et al., 2022) that was published by the Human Genetics Laboratory of the *Giannina Gaslini* Institute.

3.1.3.4 Hemoglobinopathies

Hemoglobinopathies are the result of genetic alterations affecting one or more globins that constitute the hemoglobin (Hb) tetramer, and account for one of the most frequent recessive monogenic diseases in the world (Ivaldi & Barberio, 2019), (Traeger-Synodinos et al., 2014).

Briefly, erythrocytes are crucial cellular components of blood, constituting almost 45% of total blood volume. Their main function is to carry oxygen along the body through the binding of oxygen to hemoglobin (Eleftheriou & Thalassaemia International Federation, 2003). Each hemoglobin is a tetramer that can be constituted by four types of globin chains, including the alpha-chains (α), and non-α chains - essentially beta-chains (β), gamma-chains (γ), and delta-chains (δ). In addition, each globin chain is wrapped around a heme group – a ring structure with an iron ion that allows O_2 molecules to bind. These globin chains are combined in pairs of two in each hemoglobin to form different types of hemoglobin tetramers (Ivaldi & Barberio, 2019), (Eleftheriou & Thalassaemia International Federation, 2003). The most frequent Hb in adult life is the HbA, constituted by two α-chain and two β-chain globins ($\alpha_2\beta_2$). Representing a smaller portion of the Hb in adult life, HbA_2 is composed of two α-chain and two δ-chain globins. Fetal hemoglobin (HbF) can be found in bigger proportions from the sixth week of gestation to term and represents about 1% of Hb in normal adult life. These HbF are composed by two α-chain and two γ-chain globin ($\alpha_2\gamma_2$) (Eleftheriou & Thalassaemia International Federation, 2003).

Genetic variants in genes that codify globin chains can affect hemoglobins in two ways, by interfering with the synthesis of the globin chains - resulting in thalassemia and, in the persistence of HbF into adulthood - or by affecting the structure of the globin chains, resulting in structurally abnormal variants, known as Hb variants. As, for example, the sickling of red blood cells (Traeger-Synodinos et al., 2014). The wide diversity of causing-thalassemia variants and aberrant hemoglobins result in an increasing number of defects, both in a qualitative and quantitative way, including very severe defects (Traeger-Synodinos et al., 2014). Abnormal hemoglobins are the result of the substitution of an aminoacid, on the globin chains, or other gene defects like deletions/duplications affecting the hemoglobin structure. Sometimes the resulting Hb

may not cause symptoms, other times can impair the Hb functions resulting in severe clinical pictures.

Thalassemia

Thalassemia can be categorized depending on the defective gene. For example, variants in genes that codify the α-chains may result in α-thalassemia, and the same applies to β-thalassemia. These disorders, however, can be more complicated since they can be the result of a major loss involving more than one codifying-globin gene, such as the δβ-thalassemia. This thalassemia results from a deletion in both the β and δ genes on chromosome 11. Furthermore, because of the interaction between genotypes of all different forms of thalassemia, heterozygous compounds can be formed, such as the α/β-thalassemia, which can difficult the predicted phenotype (Ivaldi & Barberio, 2019).

In thalassemia, the synthesis of globin chains can be affected depending on the genetic variant type, resulting in a more mild or severe thalassemia phenotype. In β-thalassemia, the variant's impact might be characterized as moderate, significant with some production of the globin chains, or no production at all. The complete absence of β-chains, or low production, interferes with the erythropoiesis process by creating an imbalance of α-chains compared to the β-chains. However, the severity of the β-thalassemia disease can occasionally be reduced if the number of lacking beta-chain becomes more balanced compared to the number of other chains. These situations include, for example, the low production of the α-chains or the persistence of HbF in adult life (Eleftheriou & Thalassaemia International Federation, 2003).

The two main types of thalassemia, α and β, are caused by specific variants in genes that codify the α and β chains (figure 33). β-globin chains are codified by the *HBB* gene, inserted in the β-globin gene cluster, and located in the short arm of chromosome 11 (11p15.5). α-globin chains are codified by two genes (*HBA1* and *HBA2*), inserted in the α-globin chain cluster, and located on chromosome 16 (16p13.3).

In α-thalassemia, more than 100 variants have been reported, and the most common variants include deletions in one or both *HBA* genes. In β-thalassemia, more than 280 variants were found, mostly single nucleotide variants, and in smaller percentages, deletions. Moreover, in rare cases, thalassemia can also be the result of

variants in the regulatory genes of the globin chain synthesis (Traeger-Synodinos et al., 2014).

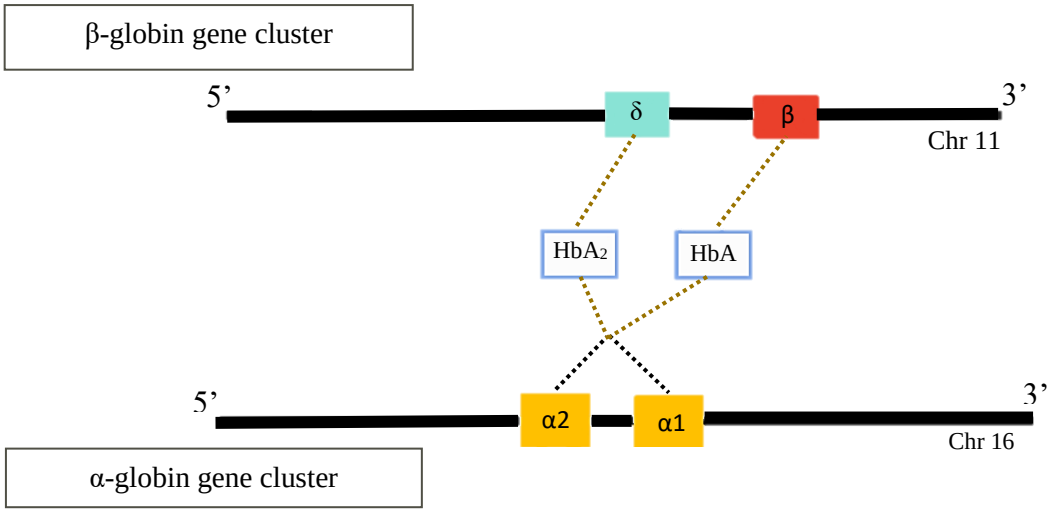


Figure 33- α and β globin gene clusters, and the respective chains that compose hemoglobins in adult life.

Carriers Detection

The detection of carriers is critical in hemoglobinopathies to identify couples with a higher risk of generating an affected child. For example, in β-thalassemia, a born child with thalassemia will not show apparent signs of the pathology due to the production of enough amounts of HbF, which ensures the balance in the number of α and γ-chains. This protects the newborn from any perturbation in the erythropoiesis process, but in adult life may be problematic as the quantity of HbF decreases. In some cases of β-thalassemia, HbF remains into adulthood to compensate for the imbalance of globin chains (Traeger-Synodinos et al., 2014).

Therefore, the identification of thalassemia-causing-variants in parents is crucial for the assessment of an affected fetus, which allows the search for inherited thalassemia-causing-variants in an early stage of life. Although molecular diagnosis allows for an accurate and easy result for prenatal diagnosis and carrier identification, this group of blood diseases is distinct from other diseases, and thus carrier identification should be performed by biochemical means rather than DNA analysis. Genetic testing should only be carried on determining the carrier status of a complex case, or to confirm hematological/biochemical results that were unclear. In addition, DNA analysis is performed in preimplantation genetic diagnosis, by known high-risk couples for a

thalassemia type. This involves the previous characterization of the parent's genotype (Traeger-Synodinos et al., 2014).

Hematological Methods for Thalassemia Diagnosis

Accordingly with the guidelines of the European Molecular Genetics Quality Network for molecular and hematology methods for carrier identification and prenatal diagnosis of the hemoglobinopathies of 2014 (Traeger-Synodinos et al., 2014), the first level of diagnostic is focused on hematological/biochemical means that are performed in blood samples. Methods include complete blood count (CBC), the determination of the hemoglobin pattern, and hemoglobin component quantification. The most used methods for the assessment of these parameters are electronic blood counts and High-Performance Liquid Chromatography (HPLC), which enables the quantitative and qualitative analysis of the hemoglobin fractions.

Complete blood count measures several blood components and features, such as mean corpuscular volume (MCV), the mean corpuscular hemoglobin (MCH), red blood cells (RBC) count, PCV indicates the RBC component of the blood in percentage, and the hemoglobin content. In addition, also sideremia (concentration of iron in the blood serum), and levels of ferritin can be observed in the blood serum. These measurements are important to discriminate different conditions. For example, if the hemoglobin levels, RBC count, or PCV values are below the normal range it is a sign of anemia, which can have many causes. The same is applied to the MCV values, which is a sign of microcytic (if its values are abnormally low).

The analysis of **hemoglobin patterns** is used to detect abnormal hemoglobins, and results should be confirmed always by two methods, including hemoglobin electrophoresis, isoelectric focusing, HPLC, and capillary electrophoresis. **For quantification of hemoglobins**, such as HbA_2, methods used include HPLC, capillary electrophoresis, and Hb electrophoresis.

Overall, these 3 measures – complete blood count, hemoglobin patterns and hemoglobin quantification – are essential for the first assessment of patients with potential thalassemia defects and are essential to discriminate thalassemia from other pathologies. In addition, second-line hematological methods can be used to identify thalassemia

carriers for cases in which the diagnosis is not clear, such as the evaluation of red cells morphology.

Accordingly, with the Italian Society of Thalassemia and Hemoglobinopathy (Ivaldi & Barberio, 2019), indications for the first-level assessment of hemoglobin defects include familiarity with hemoglobin defects, confirmation of clinical suspicion in the presence of marked anemia, raised hematocrit values, and already established microcytosis with indices in the normal range, among others.

Molecular Diagnostic Methods for Thalassemia

If hematological results from the first line tests are not clear, the diagnostic procedure should be carried on to the molecular diagnosis, which consists of genetic testing – used to confirm carriers of more complex cases or unclear biochemical/hematological results. For this effect, specific guidelines are available with recommendations on when to proceed to molecular diagnostic (Ivaldi & Barberio, 2019), (Traeger-Synodinos et al., 2014). It should be kept in mind that different countries have their own characteristic gene combinations and frequent variants, therefore genetic analysis in each patient should be assessed accordingly to the patient's origins.

☑ Molecular Diagnosis of α-thalassemia

In cases of α-thalassemia, the most frequent variants are deletions in the *HBA1* and *HBA2* genes, meaning that chosen techniques must be directed toward the identification of deletions in the *HBA* genes. PCR-based methods are frequently used, such as gap-PCR, which uses a specific set of primers to identify the most frequent *HBA* deletions. In gap-PCR, two primers are designed to be complementary to the sense and the antisense strand of the DNA region of interest, to flank the deleted region (Sabath, 2017).

Previously, other methods included Sanger sequencing and Reverse Dot Blot analysis, used for single nucleotide variants. For the detection of deletions, MLPA assay is used, which detect copy number variants through a qualitative analysis of gene dosage. Its advantage consists in detecting deletions that might not have been described yet through gene dosage imbalance (Sabath, 2017).

☑ Molecular Diagnosis of β-thalassemia

For β-thalassemia, Sanger Sequencing is the most frequent method used because the most frequent variants are single point variants. In addition, this method is easier to perform for β-thalassemia because while α-globin chains are codified by four alleles, β-chains are codified by two alleles (from one single gene). The same diagnostic methods applied to α-thalassemia can be used for β-thalassemia for the detection of deletions involving the *HBB* gene. Additionally, for the most frequent single nucleotide variants, Sanger sequencing can be used or Reverse dot-blot, which targets the most frequent single nucleotide β-variants of a specific population (Sabath, 2017).

☑ Molecular Diagnosis with NGS

The introduction of the NGS technology for the identification of thalassemia-causing-variants is very recent, and not all laboratories have already established this technique in the diagnostic practice. The *Gaslini* Institute's Human Genetics Laboratory is now in the process of transitioning from traditional technologies like Sanger sequencing and MLPA to the NGS methodology that is becoming part of the routine workflow. The kit that is in use is the *Devyser Thalassemia* kit ("Devyser Thalassemia NGS | Devyser," 2021) for the analysis of sequence variants in the *HBA1*, *HBA2*, and *HBB* genes, including the detection of single nucleotide polymorphisms, indels, and CNVs. This assay is based on a single run, with one tube and no additional workflows. The targeted NGS method used is by amplicon sequencing, therefore, first a multiplex PCR reaction is performed with the aim to amplify the genomic DNA, resulting in a library of target amplicons of each sample (the library covers the *HBA1*, *HBA2* and *HBB* genes, among others, in a partially overlapping). In a second PCR reaction, sequencing adapters, with unique index sequences, are introduced into each amplicon, allowing samples to be pooled together. After that, samples are purified and sequenced by the Ilumina technology (by synthesis). The sequencing run covers all genes involved in thalassemia, including its promoter, exon, and intron, as well as upstream and downstream regions to be able to detect the CNVs. The sequencing data is then analyzed using a suitable software for the targeted sequencing.

For a more accurate detection, deletions associated with the α- and β-thalassemia are detected by a second PCR method, the GAP-PCR, which consists of a direct method that

uses primers to align both ends of the breakpoints. As the introduction of the NGS in the diagnostic of thalassemia is still a recent event, auxiliary techniques are used to confirm potentially abnormal or unclear results generated by the NGS methodology, as the use of MLPA to confirm possible deletions in these genes.

➢ Clinical Case

Case description: In the 15th week of pregnancy, a 39-year pregnant woman of Vietnamese ethnicity was referred for clinical investigation due to an abnormal value on the hemoglobin parameters of a routine blood analysis.

Methods and Results: First, by recommendation of the Italian Society of Thalassemia and Hemoglobinopathy, hematological parameters were assessed, including indices of whole blood count, ferritin and iron concentration through electronic measurement (figure 34). The hemoglobin assessment was performed by HPLC, and it used on total four HPLCs programms: by Bio-Rad (Beta-thal DualKit and Beta-Thal Short) and by the Trinity Biotech Premier Hb9210 Resolution System in the two versions "quick scan" and "high resolution".

Very shortly, HPLC is a technique in chemistry that is useful to separate, identify and quantify each component in a mixture. It can identify and quantify several normal and abnormal hemoglobins, contributing in most cases to the clinical diagnosis of patients with thalassemia. HPLC is very useful due to its high resolution and reproducibility of the hemoglobin results. The difference between the retention times of hemoglobins of a normal and an abnormal sample is what gives the high precision of the HPLC (Khera, Singh, Khuana, Gupta, & Dubey, 2014). The use of different programs on HPLC equipment allows for better separation of mixture components and analysis of the same sample with different retention times, allowing for better separation, detection of possible hemoglobin defects, and assisting clinicians in the accurate interpretation of laboratory data quite useful considering the increasing diversity of genetics and ethnicity in the Italian population.

	Observed value	Unit of measure	Reference value	Hb pattern detected with the four HPLC instruments			
				1	2	3	4
RBC	4,87	$X10^6/\mu L$	4,0-5,4				
Hb	13,0	g/dL	12,0-14,0				
PCV	40	%	38-47				
MCV	82,2	fL	82-98				
MCH	26,7	pg	25-34				
Sideremia	93	µg/dL	37-145				
Ferritina	30	ng/mL	13-150				
HbA_2		%	2,5-3,2	3,0	1,9	1,8	1,6
HbF		%	< 1.0	0,9	0,6	-	0,1

Figure 34 - Observed results from whole blood count, sideremia, ferritin, and the assessment of the hemoglobin through four different programs of HPLC: (**1**) Biorad (BetaThal DualKit), (**2**) Bio-Rad (Beta-Thal Short), (**3**) Trinity Biotech Premier Hb9210 Resolution «quick scan», (**4**) Trinity Biotech Premier Hb9210 Resolution «high resolution».

The hemogram and hemoglobin analysis of the pregnant women are provided in figure 34. Relatively to the hemogram analysis, all values were between the reference values. Although the MCV parameter was normal (with 82.2), the value was dubious because it was placed on the borderline value (with the reference parameter between 82-98). An abnormally low value in the MCV parameter can be an indicator of microcytosis, which is defined as small red blood cells. The observation of microcytosis is usually treated as suspicion of thalassemia.

Although the hemogram parameters did not indicated abnormal values, the hemoglobin assessment performed with different HPLC systems did not agree with that. Accordingly with the findings obtained from the four HPLC runs, particularly of the HbA_2 parameter, three of the HPLC results indicated levels below the reference values. Rare genetic factors and acquired factors can impair the HbA_2 levels, therefore, further investigation was performed.

The analysis of the hemoglobin pattern through the Biorad (BetaThal DualKit) system (figure 35) showed a possible attempt of a split of the HbA_2 pick. The split could be the result of an overlapping peak caused by the presence of an α variant or caused by a δ (delta) gene variant, because HbA_2 is constituted by two α-chains and two δ-chains. Despite the peaks, the HbA_2 value (3,0%) was considered normal because it was between the reference values (2,5-3,2%). The analysis of the hemoglobin pattern through the Biorad (BetaThal Short) system (figure 36) highlighted an asymmetric HbA peak, which suggested the presence of an undefined Hb variant. Also, HbA_2 values (1,9%) were below the reference values (2,5-3,2%).

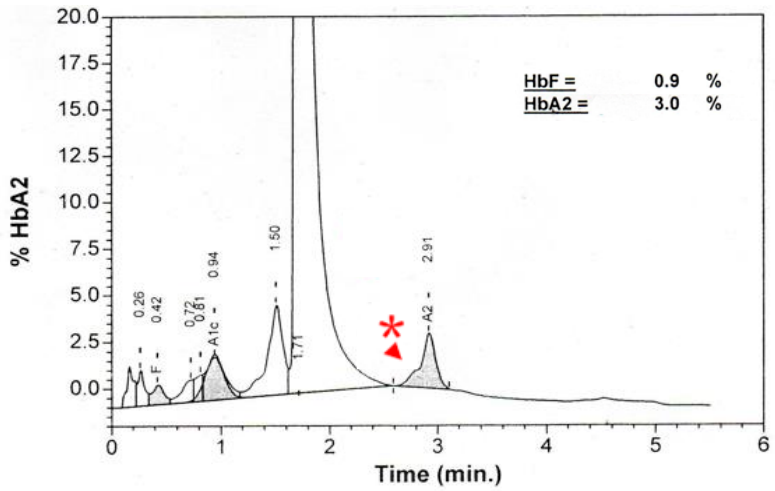


Figure 35- Hemoglobin pattern through the Bio-Rad (BetaThal DualKit) system. The red star indicates a possible division of the peak.

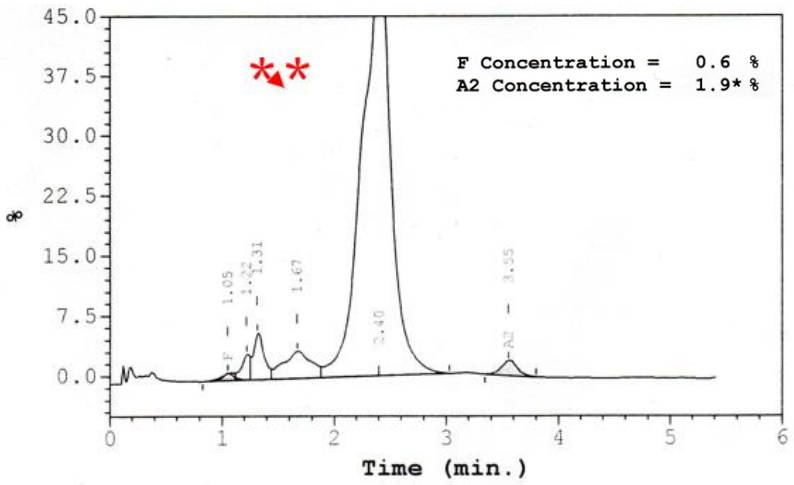


Figure 36 - Hemoglobin pattern through the Bio-Rad (BetaThal Short) system. The red stars point out an asymmetric HbA peak.

The analysis of the hemoglobin pattern through the Trinity Biotech Premier Hb9210 Resolution System by the two versions "quick scan" and "high resolution" also suggested the presence of an Hb variant. By comparison with the reference values (2,5-3,2%), HbA_2 quantification showed abnormal values (1.8% and 1.6%). Since milder forms of α-thalassemia may be present without significant alteration of blood counts, molecular studies were conducted to characterize the hypothesized defect.

First, DNA extraction was performed on the peripheral blood of the pregnant woman. Specific fragments were amplified, and sequencing was performed on the *HBA1*, *HBA2* and *HBB* genes using the BigDye™ Terminator v3.1 Cycle Sequencing Kit. Fragments were separated on ABI Prism 3130 DNA analyzer (Applied- BioSystems, Foster City, CA).

Considering that the patient has Asian origins, and that these geographic regions are particularly affected by α-thalassemia defects, the search for variants in the *HBA* genes was also performed by Reverse Dot Blot.

Molecular results: Sanger Sequencing (figure 37) identified a heterozygous variant in the *HBA2* gene. The identified variant was the substitution G>T, which results in the substitution of Glu>Asp. The HbVar (the database of the Human Hemoglobin Variants and Thalassemias) classifies this variant as Hb Hekinan (HBA2:c.84G>C). However, the resulting hemoglobin is stable and functional, classified as benign ("HbVar ID 44|Hb Hekinan," 2004). Additionally, a different α-thalassemic defect was found *in trans*, on other allele, with a deletion of about 3.7kb. In fact, the most frequent alpha globin gene deletion is the 3.7kb loss, which is caused by a 3,7kb break in the DNA sequence that encompasses the two alpha globin genes regions. Because there are two HBA genes in each chromosome, and they are very similar between themselves, after broken ends are re-joined together, it forms a DNA sequence that functions as single alfa globin gene (Borgio et al., 2017).

The two defects are therefore individually transmitted by the two parents. These variants do not produce any hematological alterations or known clinical consequences, but however is important to suggest partner testing to exclude possible interferences with more important α-thalassemic defects in the fetus.

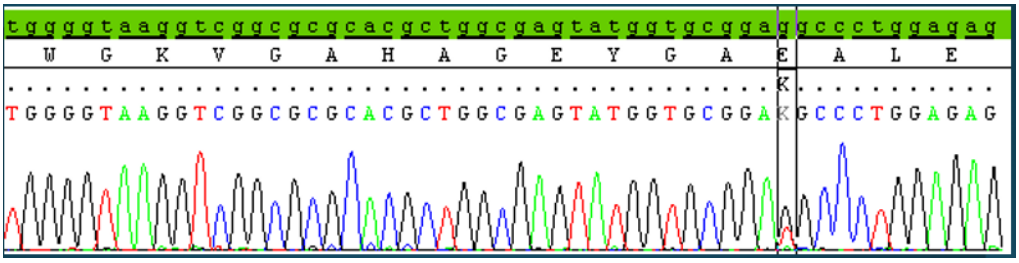


Figure 37 - Analysis of the sequencing reaction using Sanger Sequencing. It is observed the change of a G>T, which is translated in the change of Glu>Asp.

Considerations: This was a case from the laboratory archive of some years ago, therefore the methodologies used in this case were methods that are not currently used, such as the Reverse dot blot. Although HPLC is still currently used for the first diagnostic level, with Sanger sequencing being used less and less, molecular analysis today looks forward to using the NGS technology to diagnose thalassemic patients.

3.2 Molecular Cytogenetics

3.2.1. Introduction

Conventional cytogenetics has a low resolution, with only 5 to 10 megabase pairs, but it allows the global observation of all chromosomes, as well as their organization. The same cannot be said of molecular techniques, which are focused mainly on the nucleotides that constitute the DNA sequence. Therefore, the molecular cytogenetics field was born to overcome limitations derived from both fields (Speicher & Carter, 2005).

Molecular cytogenetics came to improve chromosomal analysis resolution, in molecular diagnostic and basic research, through the application of molecular techniques to chromosomes. This field emerged from the observation that complementary sequences could hybridize and form more stable complexes than non-complementary sequences. Moreover, in Fluorescent *in situ* hybridization (FISH) the development of this principle led to the use of tagged DNA probes to hybridize with chromosomes in metaphase or interphase, chromatin fibers, which was shown to be extremely useful as a discovery tool in diagnostic. A crucial advantage of this field, besides its higher resolution, is the analysis of single cells. This is an important feature as it allows a more comprehensive understanding of changes in a single cell than those at a population level, relevant in cases of mosaicism (Speicher & Carter, 2005).

A further evolution has been the technique "Comparative Genomic Hybridization" (CGH) where molecular probes were included in a microarray to evaluate possible chromosomal deletions or duplications with higher resolution. The limit of this technique is based on its impossibility to detect balanced translocations. The incorporation of molecular cytogenetics approaches into clinical practice became especially crucial for prenatal diagnosis by detection of aneuploidies and unbalanced structural rearrangements in the fetus of women with advanced maternal age.

The latest advance has been achieved with the application of NGS. In particular, NGS facilitated the sequencing of cell-free DNA in maternal circulation with the possibility to identify aneuploidies or abnormal amount of DNA of the fetus or even due to specific rearrangements or abnormal tumor DNA of the mother. Therefore, despite the increased ability to assess the fetus's DNA, this also raises many concerns for its implantation in practice (Durmaz et al., 2015).

3.2.2 Non-Invasive Prenatal Screening test

Non-Invasive Prenatal Screening (NIPS) calculates the risk of a fetus to born with an aneuploidy, through analysis of fetal cell-free DNA (cfDNA) in the maternal blood. This innovator non-invasive method was introduced to clinical practice in 2011, in Hong Kong, and in 2013, by recommendation of the ACMG, was rapidly integrated into other laboratories (Gregg et al., 2016), (Allyse et al., 2015).

NIPS is aimed to detect the most frequent aneuploidies, including the aneuploidy of chromosome 21 (Down Syndrome), chromosome 18 (Edwards Syndrome), and chromosome 13 (Patau Syndrome) as well as the detection of aneuploidies of sex chromosomes (X and Y). CNVs related to genetic diseases have been also integrated into NIPS analysis, but this field is still evolving and therefore it was not introduced in clinical practice because of its low sensitivity and specificity. Some commercial kits have already been developed and are available in the market as a useful screening test for single-gene disorders, but providers and patients must be aware of the limitations of this application (Gregg et al., 2016).

NIPS's validation studies showed that its detection rates and clinical specificity are very high compared to traditional methods (ultrasonography associated with biochemical markers), with approximately 99% for Down Syndrome, and about 80% to 100% for Edwards Syndrome and Patau Syndrome. On sex chromosomes aneuploidies, NIPS was observed to have a detection rate of more than 90% and a false-positive rate of closely 1% (Gregg et al., 2016).

Besides the high detection rates, NIPS remains as a screening test, and is not considered a diagnostic test because the possibility of false negative or false positives results. However, when results indicate an increased risk for an aneuploidy, the invasive test is still recommended before interrupting a pregnancy. Aside from that, ethical issues can also emerge from this kind of test such as the possibility of detecting incidental findings, like a genetic alteration in mother cells, when that was not the purpose of the test. Proper genetic counseling is needed to reflect NIPS test implications (Gregg et al., 2016).

Circulating cell-free DNA in maternal blood

The discovery that placental cfDNA during pregnancy was abundant in the maternal plasma allowed the development of non-invasive approaches, based on amplification of specific sequences, from maternal plasma, by PCR and sequencing (Lo et al., 1997). cfDNA are ubiquitous small fragments of double-strand DNA (dsDNA) present in maternal plasma, from mother cells but also placenta cells. In a smaller portion, it can also be found cfDNA from erythroblasts of the fetus, but its detection is harder. Circulating dsDNA fragments can result from the intracellular content released from cells undergoing apoptosis. DNA is released into the circulatory system and is broken down into tiny pieces till eventually be excreted or recycled in the body (Goldwaser & Klugman, 2018).

The percentage of cfDNA in maternal blood that is originated from placenta cells is called **fetal fraction** and must be in enough quantity to be detected by NIPS technologies. Data suggest that the lower limit of the fetal fraction that gives reliability to a NIPS result is about 4%, which happens around the tenth week of pregnancy. This aspect is useful in prenatal since the tenth week is considered an early gestational age for genetic screening. Since a NIPS result with a higher risk allows to perform a diagnostic test, pregnant women will have access to an earlier invasive prenatal diagnosis, creating more time to consider a diverse range of options, which can result in the alteration of the pregnancy's course or to prepare parents for a child with disabilities (Gregg et al., 2016).

Screening technology of cell-free DNA

The setting up of a method for extracting cfDNA from pregnant women's blood allowed the analysis also of fetal cfDNA. Historically, initial methods started with the amplification of specific sequences (that were absent in maternal genome) by PCR-based methods. These methods lead to an increasing development of the technology and open the possibility of screening tests based on placenta cfDNA analysis. Moreover, researchers directed this approach to try to develop and identify trisomy 21 in blood of pregnant women through the comparison of a relative excess of specific sequences of chromosome 21 to those from other chromosomes, as for example, the highly quantitative microfluidics digital PCR.

Fortunately, these breakthroughs converged with the development of high-throughput methods for DNA sequencing (NGS), enabling it to be possible to screen specific aneuploidies. From here, methods for screening aneuploidies in prenatal have emerged as a result of this breakthrough, and then applying a bioinformatic algorithm that take into account nucal translucency and maternal age, to estimate the risk's pregnancy for trisomy on chromosomes 21, 18, or 13 in prenatal, as well as the fetal gender (Sehnert et al., 2011).

NIPS procedure

During my internship, the screening test used in the Human Genetics laboratory was based on counting sequence tags from each chromosome, through bioinformatic analysis. The workflow starts with the separation of the plasma fraction from total peripheric blood. Red blood cells fraction is discarded, and cfDNA extraction is performed in the plasma fraction. The protocol followed for cfDNA extraction is "Purification of Circulating Nucleic Acids from 4 mL or 5 mL Serum or Plasma" from QIAGEN company and is essentially composed by four essential steps, including lyse, bind, wash and elute. After that, dsDNA present in the isolated cfDNA is quantified by the "Qubit dsDNA HS Assay" kit. The equipment makes fluorescence measures using target-selective dyes that emit fluorescence when bound to dsDNA.

For the generation of sequencing libraries in the NGS Ilumina plataform the commercial kit used is the *NIPS-GeneSGKit® Advanced*, from *Sistemas Genómicos* company (Sistemas Genómicos, 2020). The target sequencing method used is the amplicon sequencing, without the need for DNA fragmentation.

During the preparation of the sequencing library, adaptors are attached by ligation to nucleic acid fragments of each sample library. Adapters are constituted by specific sequences with different purposes on the fragment. Adapters include sequences that allow fragment recognition by the sequencer, that allow library fragments to bind to Illumina platforms' flow cell oligonucleotides. These sequence adaptors can variate their specificity along with the sequence equipment being used. Other specific sequences of the adaptor are also needed for a successful sequencing reaction, such as binding sites for the sequencing primers. Libraries preparation by amplicon, is a crucial step for a reliable analysis because the quality of sequencing data is directly affected by it.

For a faster and more efficient sequencing reaction, an index sequence, is also added to the DNA fragments. These sequences are sample-specific barcodes that allow samples from different patients to be sequenced together, in a multiplex reaction. Then, during data analysis, index sequences assign individual sequence reads to the correct sample. Depending on the number of libraries merged and the amount of accuracy needed, adapters may have single or dual sample indexes. Through the *NIPS-GeneSGKit® Advanced* kit, two index sequences are added to each fragment. After obtaining the indexed libraries using two sequences of specific sample barcodes, libraries are pooled, allowing them to be sequenced together. Samples are loaded into the NGS equipment (into the flow cell) and cluster generation begins.

Bioinformatic Analysis

Accordingly, to the index sequences, a demultiplexing algorithm sorts the readings into various files allowing to identify and separate the reads of each sample. At this point, reads are called sequence tags, strands of 30 to 60 nucleotide bases long. These ones are distinctive enough to be aligned to the reference genome sequence in order to identify which chromosome they belong to (Gregg et al., 2014).

After library sequencing, the raw data obtained is analyzed in a bioinformatics platform specific for the company that provided the kit. In practice, the files are uploaded into a database to be subjected to bioinformatic analysis. To calculate the number of copies of each chromosome or chromosomal region, a bioinformatic algorithm is used to align the sample's readings against the reference genome to calculate the number of copies of each chromosome that detects partial or total aneuploidies (Gregg et al., 2014).

For example, if there is a slight excess of sequence tags originating from chromosome 21, due to an extra chromosome 21 from a fetus, this is detectable because the number of reads from maternal plasma are compared to an expected value, a specific cutoff algorithm optimized for the detection of trisomy 21. Although this degree of sequencing is insufficient for nucleotide-level sequence analysis, it is sufficient to reliably map and count sequence tags to specific chromosomal regions (Gregg et al., 2014).

3.3 Classical Cytogenetics

In the laboratory, my interaction with the cytogenetics field was very limited. However, during the internship I followed a cytogenetic protocol with the purpose to perform a cytogenetic investigation on peripheral blood lymphocytes. In addition, I observed the set-up cultures of samples from pre- and post-natal.

3.3.1 Historical Context

Human cytogenetics was born with Walther Flemming, the first person to see human chromosomes and introduced the terms "*chromatin*" and "*mitosis*". Later, in 1988, Heinrich Von Waldeyer introduces the term "*chromosome"* to describe what Gregor Mendel had previously proposed as "*linked up groups*", the fundamental laws of inheritance. However, those who first proposed the chromosome hereditary theory were Walter Sutton and Theodor Boveri, allowing to define subjects like genetics and cytology as belonging to the cytogenetics field (Liehr, 2021).

The first human genetic analysis was executed in the cytogenetics field, by Thomas Painter in 1923, who reported the observation of 24 chromosomes in testicular cells. Suggesting that 48 was the number of chromosomes in each somatic cell. Only in 1956, Joe Hin Tijo and Albert Levan define the correct number of chromosomes as 46, due to an accidental alteration in the karyotype protocol. This was an important landmark as it allowed the field of human cytogenetics to emerge (Gartler, 2006).

With the correct number of chromosomes established, human cytogenetics started to be a field of interest. In 1959, Jérôme Lejeune and colleagues identified a genetic chromosome-related pathology, characterized by the presence of an extra chromosome in fibroblasts of patients with Down syndrome. This discovery enhanced the discovery of other pathologies linked to chromosome counting, such as Turner Syndrome, Klinefelter Syndrome, Patau Syndrome, and Edward Syndrome (Liehr, 2021). In addition, the discovery of the peripheral blood as a sample for chromosome analysis, in 1960, was fundamental for avoiding more invasive approaches, such as medulla aspiration and skin biopsy. In 1966, the discovery that amniocytes could be used to determine the fetus genotype, lead to the application of genetic procedures to prenatal investigation (Durmaz et al., 2015). After that, the discovery of a dye fluorochrome allowed chromosome

banding in plant cells. As a result, in 1970, the same dye was used to stain human chromosomes, this technique stayed known as Q-banding, with the use of *quinacrine mustard* fluorochrome. Other techniques were also developed such as the G-banding and R-banding techniques. This landmark allowed chromosomes to be more easily identified, as well as their structural chromosomal alterations, including deletions and duplications (C. Estandarte, 2012).

In 1964, an official chromosome classification was approved at the London Conference which allowed the classification and nomenclature of chromosomes based on their size, centromere position, and other morphologic features, separating chromosomes into 7 groups. This included the general agreement that the human chromosomic complement was 46 chromosomes, constituted by 1 to 22 pairs of autosomes, in descending order of length and the nomenclature of X and Y for sexual chromosomes ("The London Conference on "The Normal Human Karyotype" 1964).

Then, the identification of chromosomes through band patterns aroused the necessity of an updated general consent for its classification. Therefore, in 1978, a document was published named "An International System of Human Cytogenetics Nomenclature", and this is the document that is continuously updated when necessary for clinical practice. Its last update was on December 10 of 2020, known as ISCN 2020, and is the routine nomenclature used in cytogenetic laboratories ("ISCN Online," 2022).

Human cytogenetics achieved several discoveries in the following years, including the upgrade of its techniques and resolution. However, the introduction of molecular biology techniques that could be applied to chromosomes was a particularly significant event, as it allowed the improvement of chromosome analysis resolution and the emergence of the molecular cytogenetics field. These advances in human cytogenetics history enhanced the possibility to classify human chromosomes.

3.3.2 Karyotype

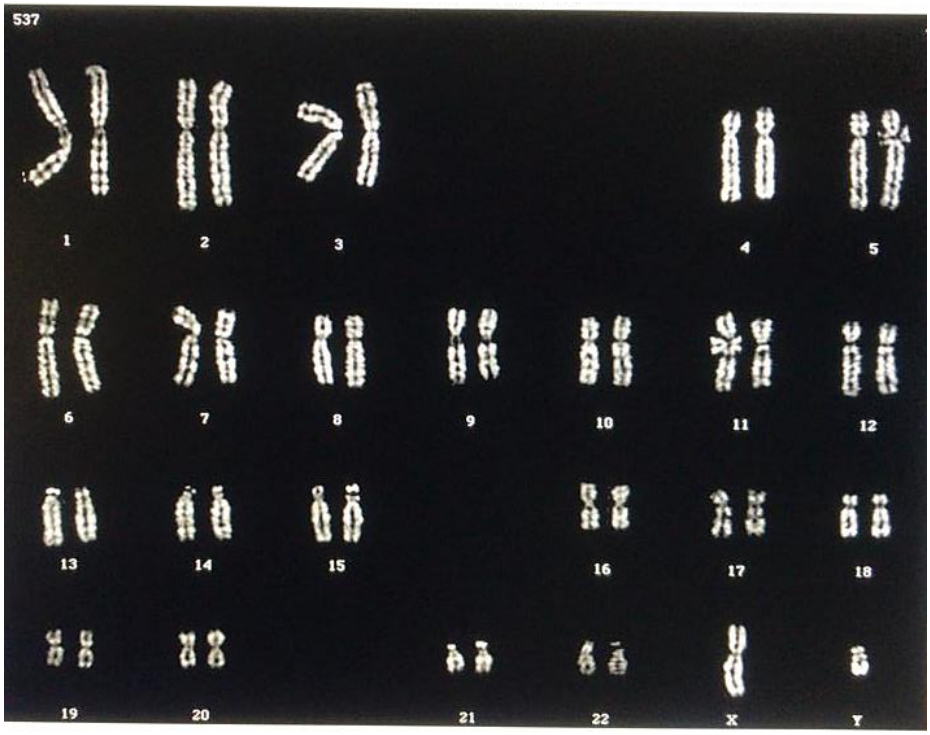


Figure 38 Normal karyotype with the chromosomic complement 46, XY. Chromosomes are arranged in pairs and designed by numbers from 1 to 22, with the last pair as X or/and Y.

Karyotype (figure 38) refers to the complete set of chromosomes of an individual. The goal is to study the numerical and structural alterations in chromosomes. To study the karyotype, samples need to be processed. Generally, samples are cultured first, and mitotic active cells are harvested at the metaphase stage (when chromosomes are maximally condensed). After that, cells are subjected to hypotonic treatment, fixation, spreading onto a glass slide, and stained through a banding technique (Korf, 2017). The observation at the microscope allows the biologist to select metaphasic cells to build a karyotype, make its clinical interpretation and write a clinical report. In clinical cytogenetics, this type of exam is commonly used in samples derived from prenatal, postnatal, and Hemato-oncology.

Although cytogenetics gives us the panoramic picture of all chromosomes, at the metaphasic stage its resolution is limited between 400 to 550 bands, which only allows the identification of rearrangements from 3 to 10 megabases (Mb). Its global resolution can depend on many parameters such as the microscope, the quality of the metaphasic cells, and the way the chromosomes are positioned. Particularly in peripheral blood, it is

possible to increase the banding resolution to as many as 800 bands due to chromosome fixed in prometaphase (because chromosomes are more extended and visible in prometaphase than in metaphase).

3.3.3 Chromosomal Abnormalities

The objective of clinical cytogenetics is to diagnose chromosome abnormalities, either numerical or structural. Numerical abnormalities correspond to changes in the number of chromosomes either by a complete set of chromosomes or, only by one or more chromosomes, known as ploidy or aneuploidy, respectively. Structural alterations include changes in chromosome structure, including translocations, inversions, duplications, deletions, isochromosomes, ring chromosomes, and marker chromosomes. The detection of these anomalies is dependent on the technique used. Alterations involving chromosome rearrangements of size greater than 5 Mb can be diagnosed by conventional cytogenetic techniques, while the detection of smaller alterations, with only a few kilobases, can be diagnosed by other techniques of molecular cytogenetics, those include the use of FISH, multicolor FISH, or array-CGH.

3.3.4 Conventional Cytogenetics in Diagnostic Practice

In the *Gaslini's* Human Genetics Laboratory, cytogenetic testing is used to search for germline and somatic chromosomal abnormalities in post- and prenatal samples. In prenatal, samples are from amniotic liquid (amniocytes) or chorionic villi (citotrofoblast cells), while in post-natal are from peripheral blood (lymphocytes). It can also be studied products of abortion with culture of the fetus tissue.

In cytogenetics, I was able to execute all steps of the cytogenetic protocol in order to build a karyotype, using as sample my own peripheral blood. Steps included cell culture in suspension, preparation of sample slides, and analysis at the microscope. However, for diagnostic practice, samples are first accepted and registered in the laboratory system, then a cell culture is established. After that, samples are processed for slide preparation and in the end, slides are stained, evaluated by a biologist at the microscope, and a report is made.

Both for prenatal and postnatal, there are available specific recommendations from SIGU (Italian Society of Human Genetics - https://sigu.net/), compiled in the document “"Guidelines for Cytogenetic Diagnostics of the Italian Society of Human Genetics - 2013”, “Constitutional Cytogenetics Guidelines – 2013” and “Constitutional Cytogenetics Operating Notes – 2013”. These should be followed in order to keep the high-quality standards of the cytogenetic investigations.

3.3.4.1 Postnatal Diagnosis

The postnatal karyotype is a useful exam used to search for aneuploidies of autosomes and sex chromosomes, search cellular mosaicisms, and to identify and characterize small supernumerary chromosomes (makers), and structural chromosomal rearrangements. The structural rearrangements include reciprocal translocations, Robertsonian translocations, insertions, inversions, duplications, deletions, isochromosomes, and ring chromosomes.

Karyotype is done in metaphasic cells cause in metaphase, chromosomes are individualized, condensed, and aligned in the center of the cell, which enables a more accurate observation at the microscope. The cytogenetic protocol for cytogenetic investigation on peripheral blood was provided by the operative instructions of the laboratory.

Peripheral Blood

Blood is essentially composed by plasma with cells in suspension, such as platelets, erythrocytes, and lymphocytes. Only lymphocytes contain DNA in their nuclei, making them ideal for use in cytogenetic analyses. Peripheral blood is collected with an anticoagulant, like heparin, for cytogenetic analysis. The wide use of peripheral blood is mainly due to the simplicity of obtaining it and the low cost of the culture techniques and reagents employed on them. The minimum amount of peripheral blood for the execution of the karyotype must be at least 1,5 ml.

Establishment of Lymphocytes Cell Culture

Sterile *Clarinet* beak tubes, with 5 mL of suspension medium are used to prepare suspension cell cultures. One lymphocyte culture is prepared for each peripheral blood sample, in which is added sterile heparinized peripherical blood to the 5mL of complete medium. This is applied to adults and children with more than 3 months.

Cellular Synchronization

Clarinet tubes with the culture medium and peripheral blood, are placed horizontally in a thermostat at 37ºC for 48 hours, without the need for 5% CO_2. To make sure that all lymphocytes are at the same phase of the cell cycle, Syncroset A solution and Syncroset B solution are added to the cell culture. Syncroset A includes methotrexate, a drug that through a cascade of chemical reactions, inhibits the synthesis of thymidine and consequently inhibits the continuation of the cell cycle in the S phase. After that, Syncroset B solution is added, containing thymidine, to resume the continuation of the cell cycle. At this point, lymphocytes theoretically start dividing from the same point.

The step of cell synchronization is crucial in blood samples because it is not possible to follow the morphology of lymphocytes under the microscope like in an adherent culture, so this step will ensure that the lymphocytes are synchronized.

Blocking Cells in Metaphase

To obtain a high mitotic index (a high number of cells in metaphase), a mitotic inhibitor such as colcemid is used. This mitotic inhibitor is the most used in cytogenetics and its function is to inhibit the formation of the achromatic fibers, or even degrade the achromatic fibers already formed. These prevent the separation of chromatids, do not progress to anaphase, resulting in the interruption of the cell cycle in metaphase. Particularly in blood samples, for cell synchronization, lymphocytes can block its cell division in pro-metaphase (when the chromosomes are still long, but already condensed).

Hypotonic Treatment

This step is performed after exposure to colcemid and consists of the entrance of water into the cell by osmosis. A hypotonic solution is used for this purpose, with a saline concentration lower than that of the cytoplasm of the cells, which allows the passage of water from the hypotonic solution to the interior of the cell (the hypertonic solution). As a result, cells become turgid, which is necessary for the following stage, which is a good spread of chromosomes across the slide. The hypotonic solution used is potassium chloride (KCl). The exposure time to KCl is decisive for a good spreading, if the exposure time exceeds the optimum time, it can cause cell lysis before spreading, but if the exposure time is less than the optimum time the cells may not become turgid and the spreading will be of poor quality.

Fixation of Preparations

The purpose of fixations is to stop the action of the hypotonic solution, to fix the turgid cells, and, particularly in blood, to cause the cellular lysis of the erythrocytes, as can be seen by the change in color from the first fixation to the last.

The addition of the first fixation solution is a crucial step for obtaining preparations with high-quality chromosomes. Therefore, the fixative can be added “drop by drop” with continuous agitation, otherwise, blood clots can be formed leading to a "dirty" preparation, with low quality and mitotic index. However, the fixation approach can depend on the laboratory and the technician that is performing it. Afterward, the following fixations serve more as washing steps.

Chromosomes spread

A well-performed spreading is the basis of good chromosome analysis, providing a considerable number of pro-metaphases, with chromosomes well spread over the glass slide, with little chromosome overlap but not too much dispersed. At least two slides are prepared for each specimen and all slides are, previously, marked with the patient’s first and last name, the laboratory code corresponding to the test subject, and the date.

Variables such as temperature and humidity affect the evaporation rate of the fixative, which influences the quality of the spread. Meaning that, when the turgid cells meet the

glass slide, the fixative starts to spread and evaporates, putting pressure on cells that are compressed between the slide and the fixative. As the pressure increases, also the cells' membrane is compressed and stretched occupying a larger area on the slide. Thus, the longer the evaporation time, the more spread out the cells and chromosomes become. That said, low temperatures and high humidity decrease the speed of evaporation and therefore favor spreading. To control these elements, the spreading should be performed quickly and then, slides are placed to dry in a machine that controls temperature and humidity.

Many approaches can be utilized to do the spreading, each laboratory does it in its own way but essentially is the addition of drops to the slide, promoting cells lysis and spread. After that, slides are placed to dry in the "Optichrome apparatus" machine, with 28ºC temperature and 36ºC humidity. Slides are evaluated in the microscope for the quality of the spread, if it is poor, the spread should be repeated on another slide.

Chromosome Staining and Banding

Staining techniques are aimed to form a chromosomic pattern, which is commonly known by the formation of light and dark "bands" because of differential staining of DNA in metaphase, along with chromosomes' length. Bands produced by different banding techniques reflect a specific region or substructure of a chromosome (C. Estandarte, 2012).

The approach used by the cytogenetic laboratory for karyotype, during my internship, was the Q-banding. It treats metaphase slides with the chemical *quinacrine mustard*, a selective fluorescent stain. Quinacrine binds uniformly throughout chromosomes' length and produces its band patterns when subjected to an ultraviolet light microscope. It's a more simple and versatile technique because doesn´t need pre-treatment like the G-banding (trypsin), but it has a tendency to fade away when exposed for a long period of time to the microscope

For a better understanding of the staining techniques is important to have in mind that chromatin can be divided into euchromatin, or heterochromatin. While constitutive heterochromatin consists of repetitive sequences of DNA found in telomeres and centromeres (satellite DNAs), facultative heterochromatin is like constitutive heterochromatin, but it can be triggered by several factors to decondense into

euchromatin. Giemsa-banding and Q-banding, which are specific to facultative heterochromatin, are the most used banding techniques.

In practice, slides are immersed in a solution of *quinacrine mustard* for 20 minutes in the dark, approximately, then slides are washed through its submersion in water and mounted with filtered solution, to reduce the background of fluorescence. This is usually executed just right before the analysis at the microscope.

Image Acquisition, Karyotype Construction, Interpretation of Results and Reporting

After the chromosome staining step, it is necessary to view the chromosomes under the ultraviolet light microscope to see if there are any numerical or structural abnormalities in patients' cells. A careful and methodical analysis is done, cell by cell, for each slide in order to find and count metaphase cells in the best possible state, cells with individualized/ non overlapping chromosomes. Each microscope is connected to a computer, for the processing of capture metaphases.

Generally, first metaphases are visualized and chosen by image capture of microscope associated to the computer. The metaphases are analyzed, and a karyotype is constructed. For each sample, at least 16 metaphases are captured and analyzed, and at least 4 karyotypes are reconstructed by homologous chromosome pairing, including at least two karyotypes with minimum 400-band resolution. If a mosaic aneuploidy (excluding monosomy of an autosome) is detected during the analysis of the first 16 cells, the count is extended to 30 cells. If a second cell with the same abnormality is found, the count is extended to 50 cells. Each karyotype (figure 39) is analyzed, band to band, and interpreted. The senior biologist is responsible for report results, writing it and case filing.

Figure 39 - My karyotype (46,XX).

3.3.4.2 Prenatal Diagnosis

If a prenatal screening test identifies an increased risk for a particular pregnancy, an invasive test is carried out with the aim to identify chromosomic abnormalities in the fetus. To obtain a karyotype from a fetus tissue, invasive testing is based on the examination of fetal cells obtained by amniocentesis, chorionic villus sampling, or fetal blood sampling. Chromosome abnormalities are observed in about 50% and 20% of spontaneous abortions, in the first and second trimesters respectively, and are the leading cause of mental retardation and pregnancy loss (Skirton, Goldsmith, Jackson, Lewis, & Chitty, 2013).

Amniotic fluid

Set up of Amniocytes Cell Cultures for karyotyping

Amniotic fluid is obtained by amniocentesis. This is an invasive diagnostic approach performed at 15 to 20 weeks of gestation, on the second trimester (Jindal, Sharma, & Chaudhary, 2022). The cell culture is setup using specific cell culture flasks (amnioflask slide z) which allows the direct analysis of adherent amniotic cells without the need for trypsinization. Cells are incubated at 37ºC, monitored for growth at the inverted microscope, and the medium is changed when necessary. Once enough independent colonies have been reached for diagnosis, cells can be prepared for karyotyping. The culture is done *in situ* and slides are suitable for chromosome banding and FISH analysis. For each amniotic fluid, three culture flasks are prepared for quality control.

Slide Preparation for Karyotyping

When enough cells are present in culture, cells can be blocked in metaphase with colcemid. After colcemid treatment, hypotonic solution is added. For the fixation of chromosome preparations, a pre-fixing phase is done by introducing, with a insulin syringe, a few drops of fixative at the time. After removing pre-fixing solution, three changes with fixative are done. After removing the last fixative, the chamber is removed, and each slide is dried by placing it in the Optichrome equipment, at controlled temperature and humidity for 5 minutes. Staining is done by immersing the slides in a solution of Quinacrine Mustard. The staining time is determined based on the date the solution was prepared. Stained slides are rinsed with water, allowed to dry, and stored in the dark until reading under a microscope. Prior to reading, slides are mounted with filtered McIlvaine solution. The slides are read using fluorescence microscopes to identify the most suitable metaphases for cytogenetic analysis. Metaphases are analyzed by the McType and Cytovision computerized image analysis systems. Metaphase capture is performed by both technical staff and the biologist, but analysis of metaphases and derived karyotypes is performed by the biologist. As karyotype analysis is going to be performed in at least 2 primary cultures, for each fetal cytogenetic investigation, at least 10 metaphases obtained from at least 10 independent colonies are captured and analyzed by at least 4 karyotypes for each clinical case.

Chorionic Villi

Set up of Chorionic Villi Cell Cultures for Karyotyping

Chorionic villi sampling is an invasive diagnostic approach and consists of a biopsy of fetal trophoblastic tissue between the 10 and 13 week of gestation, for prenatal genetic testing (Jones & Montero, 2021). Chorionic villi sampling has the advantage of offering a diagnosis at a very early stage of pregnancy, in the first trimester. However, in 1 to 2% of cases of chorionic villi analysis, there is the possibility to detect confined placental mosaicism, meaning that cells from placenta are genetically different from those of the fetus, which might induce a false result. This is a disadvantage that should always be considered when offering this type of procedure.

The fetal cytogenetic investigation of chorionic villus cells involves the splitting of the sample to use two types of culture, the short-term culture (direct method) and the long-term culture. A short-term culture is used due to the urgent demand for diagnostic results, cause in this phase of gestation it is critical to know whether a fetus is affected or carrier of a genetic abnormality for decision making, to decide if the pregnancy should be carried on. The long-term culture can take about 7 days in culture till it's ready for karyotype, but it will have a better visualization of the chromosomes (a better resolution) relatively to the direct method. It serves as a control for the short-term method.

Slide preparation for karyotyping

First, it is necessary to disaggregate chronic villus in smaller pieces and check for maternal tissue contamination. It's important to separate maternal tissue from fetal tissue to minimize contamination of the samples with maternal cells because cells are put to grow, and later we could not know what is from the fetus and what is from the mother (if they are of the same gender), which could affect the correct cytogenetic diagnosis. Maternal tissue is distinguished from fetal tissue through its more whitish and undifferentiated appearance, while fetal tissue is more transparent and branched. In short-term culture there is no need to observe at the microscope for maternal contamination because it's such a fast method that maternal cells don't have the time to originate metaphases, while only trophoblast cells are fast diving cells able to multiply fast enough to produce metaphases.

The protocol for slide preparation of long-term culture to count metaphases and reconstruct a karyotype consists in generally the same steps as in amniocytes, meaning that is based on: blocking cells in metaphase, hypotonic solution, fixation, and chromosomal band staining. While short-term culture has a very short incubation time and use essentially the same steps for slide preparation.

Through the combination of long and direct methods, it is analyzed at least 6 metaphases obtained from the direct method, and 10 from the long-term method. Karyotypes need to be reconstructed from at least 3 metaphases (with 1 from the direct method) at a minimum resolution of 300 bands for the direct method, and a minimal resolution of 400 bands for the long-term culture method. In the presence of a mosaic, it is necessary to perform at least one karyotype for each cell line and compare the results obtained with the two methods. If necessary, chromosomal analysis can be integrated with an additional molecular technique, the QF-PCR to quickly search for aneuploidies of chromosomes 13, 18, 21, X, and Y.

3.4 References

Allyse, M., Minear, M., Rote, M., Hung, A., Chandrasekharan, S., Berson, E., & Sridhar, S. (2015). Non-invasive prenatal testing: a review of international implementation and challenges. *International Journal of Women's Health*, 113. https://doi.org/10.2147/ijwh.s67124

Bareil, C., & Bergougnoux, A. (2020). CFTR gene variants, epidemiology and molecular pathology. *Archives de Pédiatrie*, *27*, eS8–eS12. https://doi.org/10.1016/s0929-693x(20)30044-0

Basit, H., Brito, D., & Sharma, S. (2022, June 7). Hypertrophic Cardiomyopathy. Retrieved July 10, 2022, from https://www.ncbi.nlm.nih.gov/books/NBK430788/

Biancalana, V., Glaeser, D., McQuaid, S., & Steinbach, P. (2014). EMQN best practice guidelines for the molecular genetic testing and reporting of fragile X syndrome and other fragile X-associated disorders. *European Journal of Human Genetics*, *23*(4), 417–425. https://doi.org/10.1038/ejhg.2014.185

Borgio, J., Abdulazeez, S., Almandil, N., Naserullah, Z., Al-Jarrash, S., Al-Suliman, A., . . . Al-Ali, A. (2017). The -α3.7 deletion in α-globin genes increases the concentration of fetal hemoglobin and hemoglobin A2 in a Saudi Arabian population. *Molecular Medicine Reports*. https://doi.org/10.3892/mmr.2017.8033

C. Estandarte, A. K. (2012, February). *A Review of the Different Staining Techniques for Human Metaphase Chromosomes* (Master's dissertation). London, United Kingdom: University of London.

Clark, D. P., Pazdernik, N. J., & McGehee, M. R. (2019). DNA Sequencing. *Molecular Biology*, 240–269. https://doi.org/10.1016/b978-0-12-813288-3.00008-2

Cystic fibrosis | Devyser. (2021). Retrieved August 14, 2022, from https://www.devyser.com/product-categories/cystic-fibrosis

Deignan, J. L., Astbury, C., Cutting, G. R., del Gaudio, D., Gregg, A. R., Grody, W. W., . . . Richards, S. (2020). CFTR variant testing: a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, *22*(8), 1288–1295. https://doi.org/10.1038/s41436-020-0822-5

Dequeker, E., Stuhrmann, M., Morris, M. A., Casals, T., Castellani, C., Claustres, M., . . . Girodon, E. (2008). Best practice guidelines for molecular genetic diagnosis of cystic fibrosis and CFTR-related disorders – updated European recommendations. *European Journal of Human Genetics*, *17*(1), 51–65. https://doi.org/10.1038/ejhg.2008.136

Devyser Thalassemia NGS | Devyser. (2021). Retrieved August 14, 2022, from https://www.devyser.com/products/devyser-thalassemia-ngs

DeWitt, J., PhD. (2019, November 21). Targeted sequencing: Hybridization capture or amplicon sequencing? Retrieved July 2022, from https://eu.idtdna.com/pages/education/decoded/article/targeted-sequencing-hybridization-capture-or-amplicon-sequencing

Durmaz, A. A., Karaca, E., Demkow, U., Toruner, G., Schoumans, J., & Cogulu, O. (2015). Evolution of Genetic Techniques: Past, Present, and Beyond. *BioMed Research International*, *2015*, 1–7. https://doi.org/10.1155/2015/461524

Eleftheriou, A. & Thalassaemia International Federation. (2003). *About Thalassaemia*. Nicosia, Cyprus: Thalassaemia International Federation.

Garber, K. B., Visootsak, J., & Warren, S. T. (2008). Fragile X syndrome. *European Journal of Human Genetics*, *16*(6), 666–672. https://doi.org/10.1038/ejhg.2008.61

Gartler, S. M. (2006). The chromosome number in humans: a brief history. *Nature Reviews Genetics*, *7*(8), 655–660. https://doi.org/10.1038/nrg1917

Goldwaser, T., & Klugman, S. (2018). Cell-free DNA for the detection of fetal aneuploidy. *Fertility and Sterility*, *109*(2), 195–200. https://doi.org/10.1016/j.fertnstert.2017.12.019

Grasso, M., Boon, E. M., Filipovic-Sadic, S., van Bunderen, P. A., Gennaro, E., Cao, R., . . . Coviello, D. A. (2014). A Novel Methylation PCR that Offers Standardized Determination of FMR1 Methylation and CGG Repeat Length without Southern Blot Analysis. *The Journal of Molecular Diagnostics*, *16*(1), 23–31. https://doi.org/10.1016/j.jmoldx.2013.09.004

Gregg, A. R., Skotko, B. G., Benkendorf, J. L., Monaghan, K. G., Bajaj, K., Best, R. G., . . . Watson, M. S. (2016). Noninvasive prenatal screening for fetal aneuploidy, 2016 update: a position statement of the American College of Medical Genetics and Genomics. *Genetics in Medicine*, *18*(10), 1056–1065. https://doi.org/10.1038/gim.2016.97

Gregg, A. R., van den Veyver, I. B., Gross, S. J., Madankumar, R., Rink, B. D., & Norton, M. E. (2014). Noninvasive Prenatal Screening by Next-Generation Sequencing. *Annual Review of Genomics and Human Genetics*, *15*(1), 327–347. https://doi.org/10.1146/annurev-genom-090413-025341

Gu, H., Kim, M. J., Yang, D., Song, J. Y., Cho, S. I., Park, S. S., & Seong, M. W. (2021). Accuracy and Performance Evaluation of Triplet Repeat Primed PCR as an Alternative to Conventional

Diagnostic Methods for Fragile X Syndrome. *Annals of Laboratory Medicine*, *41*(4), 394–400. https://doi.org/10.3343/alm.2021.41.4.394

Harada, S., & Korf, B. R. (2013). Overview of Molecular Genetic Diagnosis. *Current Protocols in Human Genetics*, *76*(1). https://doi.org/10.1002/0471142905.hg0901s76

HbVar ID 44|Hb Hekinan. (2004). Retrieved August 14, 2022, from https://globin.bx.psu.edu/cgi-bin/hbvar/query_vars3?mode=output&display_format=page&i=44

Ilumina. (2017). *An introduction to Next-Generation Sequencing Technology* (Pub. No. 770–2012-008-B ed.).

ISCN Online. (2022). Retrieved August 18, 2022, from https://iscn.karger.com/

Italian Society of Human Genetics. (2013). *Cytogenetic Diagnosis Guidelines*. Retrieved from http://sigu.net/wp-content/uploads/2020/11/Linee_Guida_14_febbraio-2014.pdf

Ivaldi, G., & Barberio, G. (2019, April 16). Linee Guida per le Emoglobinopatie – SITE Società Italiana Talassemie ed Emoglobinopatie. Retrieved from https://www.unitedonlus.org/linee-guida-emoglobinopatie-site/

Jindal, A., Sharma, M., & Chaudhary, C. (2022, March 16). Amniocentesis. Retrieved August 14, 2022, from https://www.ncbi.nlm.nih.gov/books/NBK559247/

Jones, T. M., & Montero, F. J. (2021, December 19). Chorionic Villus Sampling. Retrieved August 14, 2022, from https://www.ncbi.nlm.nih.gov/books/NBK563301/

Khera, R., Singh, T., Khuana, N., Gupta, N., & Dubey, A. P. (2014). HPLC in Characterization of Hemoglobin Profile in Thalassemia Syndromes and Hemoglobinopathies: A Clinicohematological Correlation. *Indian Journal of Hematology and Blood Transfusion*, *31*(1), 110–115. https://doi.org/10.1007/s12288-014-0409-x

Korf, B. R. (2017). Introduction to Human Genetics ∗. *Clinical and Translational Science*, 281–311. https://doi.org/10.1016/b978-0-12-802101-9.00016-8

Liehr, T. (2021). Molecular Cytogenetics in the Era of Chromosomics and Cytogenomic Approaches. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.720507

Lo, Y. M. D., Corbetta, N., Chamberlain, P. F., Rai, V., Sargent, I. L., Redman, C. W., & Wainscoat, J. S. (1997). Presence of fetal DNA in maternal plasma and serum. *The Lancet*, *350*(9076), 485–487. https://doi.org/10.1016/s0140-6736(97)02174-0

The London Conference on ‘ The Normal Human Karyotype.’ (1964). *Annals of Human Genetics*, *27*(3), 295–298. https://doi.org/10.1111/j.1469-1809.1963.tb00798.x

NGS: Changing the game in CFTR analysis | Devyser. (2021). Retrieved August 14, 2022, from https://www.devyser.com/whitepapers/ngs-changing-the-game-in-cftr-analysis

Ommen, S. R., Mital, S., Burke, M. A., Day, S. M., Deswal, A., Elliott, P., . . . Sorajja, P. (2020). 2020 AHA/ACC Guideline for the Diagnosis and Treatment of Patients With Hypertrophic Cardiomyopathy. *Circulation*, *142*(25). https://doi.org/10.1161/cir.0000000000000937

Ong, T., Marshall, S. G., Karczeski, B. A., Sternen, D. L., Cheng, E., & Cutting, G. R. (2017). Cystic Fibrosis and Congenital Absence of the Vas Deferens. Retrieved 2022, from https://www.ncbi.nlm.nih.gov/books/NBK1250/

Restriction Fragment Length Polymorphism (RFLP). (2017). Retrieved July 2022, from https://www.ncbi.nlm.nih.gov/probe/docs/techrflp/

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, *17*(5), 405–424. https://doi.org/10.1038/gim.2015.30

Sabath, D. E. (2017). Molecular Diagnosis of Thalassemias and Hemoglobinopathies. *American Journal of Clinical Pathology*, *148*(1), 6–15. https://doi.org/10.1093/ajcp/aqx047

Samorodnitsky, E., Jewell, B. M., Hagopian, R., Miya, J., Wing, M. R., Lyon, E., . . . Roychowdhury, S. (2015). Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. *Human Mutation*, *36*(9), 903–914. https://doi.org/10.1002/humu.22825

Sehnert, A. J., Rhees, B., Comstock, D., de Feo, E., Heilek, G., Burke, J., & Rava, R. P. (2011). Optimal Detection of Fetal Chromosomal Abnormalities by Massively Parallel DNA Sequencing of Cell-Free Fetal DNA from Maternal Blood. *Clinical Chemistry*, *57*(7), 1042–1049. https://doi.org/10.1373/clinchem.2011.165910

Sistemas Genómicos. (2020, September 30). GeneSystems© Platform. Retrieved August 14, 2022, from https://www.sistemasgenomicos.com/en/specialists/diagnosis-kits/genesystems-platform/

Skirton, H., Goldsmith, L., Jackson, L., Lewis, C., & Chitty, L. (2013). Offering prenatal diagnostic tests: European guidelines for clinical practice. *European Journal of Human Genetics*, *22*(5), 580–586. https://doi.org/10.1038/ejhg.2013.205

Società Italiana Per Lo Studio Della Fibrosi Cistica. (2019). *Analisi genetica in Fibrosi Cistica – Consensus 2019 (rev. 1.0 – 21 Dicembre 2019)*.

Sondo, E., Cresta, F., Pastorino, C., Tomati, V., Capurro, V., Pesce, E., . . . Pedemonte, N. (2022). The L467F-F508del Complex Allele Hampers Pharmacological Rescue of Mutant CFTR by Elexacaftor/Tezacaftor/Ivacaftor in Cystic Fibrosis Patients: The Value of the Ex Vivo Nasal Epithelial Model to Address Non-Responders to CFTR-Modulating Drugs. *International Journal of Molecular Sciences*, *23*(6), 3175. https://doi.org/10.3390/ijms23063175

Spector, E., Behlmann, A., Kronquist, K., Rose, N. C., Lyon, E., & Reddi, H. V. (2021). Laboratory testing for fragile X, 2021 revision: a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine*, *23*(5), 799–812. https://doi.org/10.1038/s41436-021-01115-y

Speicher, M. R., & Carter, N. P. (2005). The new cytogenetics: blurring the boundaries with molecular biology. *Nature Reviews Genetics*, *6*(10), 782–792. https://doi.org/10.1038/nrg1692

Sun, Y., Yang, Z., Zheng, B., Zhang, X. H., Zhang, M. L., Zhao, X. S., . . . Wen, J. K. (2017). A Novel Regulatory Mechanism of Smooth Muscle α-Actin Expression by NRG-1/circACTA2/miR-548f-5p Axis. *Circulation Research*, *121*(6), 628–635. https://doi.org/10.1161/circresaha.117.311441

Tagliafico, E., & Russo, S. (2021). *Indicazioni per la refertazione di analisi genetiche eseguite mediante metodica Next-Generation Sequencing (NGS)* (Società Italiana di Genetica Umana Italian Society of Human Genetic ed.).

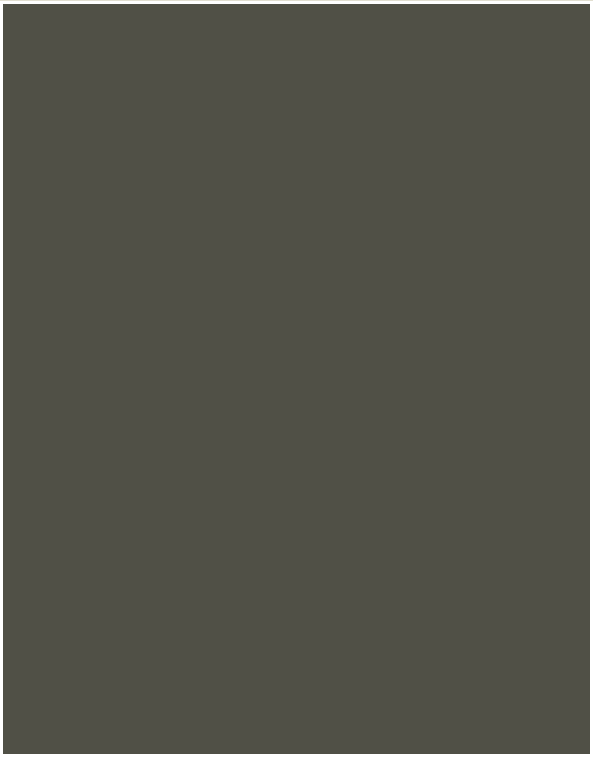
Targeted Sequencing Guide Handbook (RUO22-0863_001 4/22). (2022) (Doc ID: RUO22-0863_001 ed.). Retrieved from https://go.idtdna.com/rs/400-UEU-432/images/IDT%20Targeted%20sequencing%20guide%20%281%29.pdf?aliId=eyJpIjoiT2drdFpQdnllVXFtc09iYyIsInQiOiJoUjZGYkllUm1VXC81S3ZFTVBIK3F5Zz09In0%253D

Teekakirikul, P., Zhu, W., Huang, H. C., & Fung, E. (2019). Hypertrophic Cardiomyopathy: An Overview of Genetics and Management. *Biomolecules*, *9*(12), 878. https://doi.org/10.3390/biom9120878

Traeger-Synodinos, J., Harteveld, C. L., Old, J. M., Petrou, M., Galanello, R., Giordano, P., . . . May, A. (2014). EMQN Best Practice Guidelines for molecular and haematology methods for carrier identification and prenatal diagnosis of the haemoglobinopathies. *European Journal of Human Genetics*, *23*(4), 426–437. https://doi.org/10.1038/ejhg.2014.131

UniProt - Q20BH0 · Q20BH0_HUMAN. (2022). Retrieved from https://www.uniprot.org/uniprotkb/Q20BH0/entry

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics*, *30*(9), 418–426. https://doi.org/10.1016/j.tig.2014.07.001



CHAPTER IV

Genetic Biobank

Internship Report

Master's Degree in Laboratory Clinical Genetics

Rute Pinto

2021/2022

Chapter IV

Genetic Biobank

During the internship, I visited the Genetic Biobank of the Human Genetics Laboratory in the *Giannina Gaslini Istitute*, where I was allowed to assist the technical protocols for isolation and conservation of blood cells, as well as all the documentation necessary for the acceptance of sample in this biobank. A set of papers need to be filled in by the person who gives the biological sample, including an informed consensus explaining all implications of this process. All information is provided by the protocols of quality control of the laboratory.

The Genetic Biobank is a service unit aimed at collecting, preserving, and distributing biological materials. They operate according to high quality standards and offer support to diagnosis and research services to people affected by genetic diseases, their relative ones, and the scientific community. Accordingly with the privacy law, information regarding the person who gave the biological sample are protected by keeping separated information about the entity of the person (one coded file) and the additional medical information (like diagnosis) in a separated file. Personal data are essential in cases of important results are achieved and need to be returned to those who patients where an actionable treatment is possible.

This genetic biobank is part of the "Telethon Network of Genetic Biobanks" (TNGB). The network coordinate already existing non-profit genetic biobanks with the goal to promote biomedical research on genetic diseases. This goal is reached through the sharing of operating methods of each individual biobank, the possibility to consult the list of samples available in each biobank and monitoring the use of this samples by the scientific community. In addition, many Italian genetics biobanks collaborate in this project with the Patients Association with Rare Diseases, to contribute to research projects that might be able to identify new causative genes of diseases or associated molecular mechanisms and contribute to the development of new therapeutic strategies.

To guarantee the confidentiality of personal data, the Biobank applies requirements, including the assignment of an alphanumeric code to each sample, the connection between the code and the identity of the depositor and related information

(clinical, genetic, genealogical, etc.) may only be view by the Director of the Biobank, or authorized people. In addition, the access to the information systems, where data recording takes place, is controlled by means of suitable security measures to prevent their dissemination or use by unauthorized persons. Any research results published in journals or communicated at scientific conferences do not contain information that would allow the identification of the person concerned.

All the information was provided by the instructive operations of the genetic biobank.

- **Method for isolation and preservation of mononuclear cells from peripheral blood**

Peripheral blood samples must be handled with the goal of isolating and preserving lymphocytes, mononucleate cells that contain the patient's DNA. Thereby, the procedure includes gradient separation that is focused on the elimination of erythrocytes and dead cells and collect the remaining cells (lymphocytes and monocytes).

For quality standards, samples that arrive to the laboratory came in duplicated for each patient to ensure that there is no need to repeat the blood draw. The starting volume of peripheral blood is 3 to 5 mL, in sodium-heparin, lithium-heparin, sodium-citrate, or EDTA. The procedures that the samples are subjected to derive from the internal protocols established for the biobank in this laboratory.

- **Isolation of Peripherical Blood Mononucleate Cells**

To start the procedure, first is added 3 mL of lymphocyte separation medium in a specific type of tube, a tube with a circular barrier, and a hole in the center. Then, 3 to 5 mL of blood is added, and sterile saline solution is transferred to the tube until the desired volume is reached. This barrier in the tube is designed to help differentiate the cell extracts produced by cell separation through a density gradient, allowing a more efficient extraction of the lymphocyte layer. If the starting volume is less than 2.5 mL, separation should be performed using a falcon tube, without barrier, using a volume of lymphocyte separation medium equal to the volume of starting blood. The starting blood must be diluted 1:2 with sterile saline and transferred very slowly over the medium.

For a successful gradient separation, accordingly to the degrees of cell density, centrifugation is performed at 2500 rpm without pause for 15 min at room temperature. This step is crucial as the division of the extracts is very sensible, and perturbations in the movement of the tube can mix the different phases leading to the diffusion of the lymphocytes with other layers.

The result is a tube (figure 40) with phases including plasma (a), a ring of lymphocytes (b), a separation density gradient medium (c), and a pellet, constituted by red cells and dead cells (d). The extraction of the lymphocyte ring is made using a sterile Pasteur glass with a disposable filter/pipet. The ring is removed through circle movements and transferred to a Falcon tube with 10 ml of sterile medium or sterile saline solution. In this step, the barrier in the tube allows a more effective extraction of the lymphocyte ring, without touching in other phases.

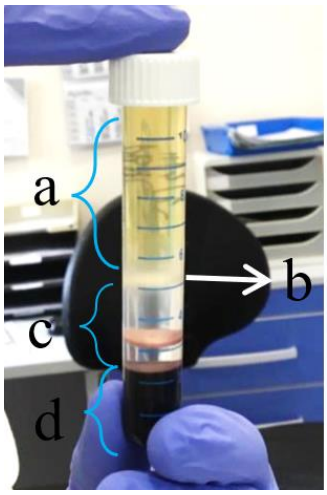


Figure 40 - Tube of diluted blood with separation medium after centrifugation; a - Blood plasma; b - Lymphocytes ring; c - Separation medium, and in the middle the circular barrier; d - Pellet constituted by red cells and dead cells.

○ Conservation of Peripherical Blood Mononucleate Cells

To conserve the cell's DNA, cells are frozen. First cells are centrifuged at 1600 rpm for 10 minutes at room temperature. Then, the supernatant is discarded, and the pellet is resuspended in 0.8 mL of sterile PBMC freezing medium, constituted by Fetal Bovine Serum and Dimethyl sulfoxide (DMSO). Finally, the liquid is transferred to a cryotube, previously labeled with the date, sample code, sample type, and pathology if known. To store, the cryotube is put into a freezing container and stored at -80ºC. After one night, it can be put into a cryogenic container with liquid nitrogen.

CHAPTER V

Final Considerations

- Conclusion
- Annexes

Internship Report

Master's Degree in Laboratory Clinical Genetics

Rute Pinto

2021/2022

Conclusion

The traineeship in the Human Genetics Laboratory allowed me to apply the theory learned during the first year of the Laboratory Clinical Genetics master's degree into practice. During this period, I was able to observe the clinical diagnosis workflow in the laboratory in all of its components, including molecular genetics, molecular cytogenetics, and conventional cytogenetics. The most useful skills learned, during the internship, in the clinical diagnosis context included the observation and discussion of different laboratory techniques to obtain the patients diagnosis, as well to interpretate and observe the emission of clinical reports, and to learn more about the strategy used for approach clinical cases with different backgrounds.

Another crucial component of the laboratory work was focused on the research on Sotos Syndrome, as it allowed me to better understand the molecular and genetic mechanisms involved in the *NSD1* pathology, as well as to understand all the laboratory work and strategy behind the writing of a research article. Research techniques are transversal to the ones of genetic diagnosis, therefore, although my laboratory experience was focused mainly on research, the execution of different laboratory techniques was useful for my own independence in the laboratory, either for research or genetic diagnosis. This research work allowed me to build my critical perspective, as I was able to observe, execute and manage different techniques. Also, my participation in this project strengthened my line of thought, as now I am able to do better decisions on the laboratory on what technique to choose and how.

Protocols used during the internship in the *Giannina Gaslini* Institute

Rute Pinto

Laboratory Clinical Genetics master's degree

2021/2022

PROTOCOL – EXTRACTION OF DNA FROM CULTURED CELLS

Sample: **Frozen pellet** with staminal cells extract

Previously:

- Prepare in the laminar flow chamber PBS = NaCl
- Check if the incubator is working at 56ºC

Procedure:

1 **WASH THE PELLET**

1- Thaw; vortex and spin the pellet
2- Add **1 mL of PBS** (=NaCl);
3- Vortex
4- Spin (1 min; 14 000 rpm);
5- Discard supernatant with a micropipette (without taking the pellet)
6- Vortex (to leave pellet more dispersed)

2

7- Resuspend the pellet in **200µL of PBS** (=NaCl) → don´t touch with the tip
8- Vortex
9- Add **20µL of Proteinase K** to each sample
10- Vortex

Optional: **If RNA-free genomic DNA is required**:

- Add **4µL of RNase A** (100mg/mL);
- Mix by vortexing;
- Incubate for 5 minutes at room temperature

3

11- Add **200µL of Buffer AL** (without added ethanol) to each sample → don't touch the tip!
12- Mix with a micropipette and by vortexing for 15 seconds
13- Incubate at 56ºC for 15/20 minutes

(while waiting, it can be prepared 2 eppendorfs for each sample – 1 normal eppendorf and 1 collection tube with spin-column – and mark them with a pen)

4

14- Add **200µL of ethanol** (96-100%) → mix with a micropipette
15- Vortex for 15 seconds

5

16- Transfer (pipetar/caricar) the mix of the eppendorf to the spin-column with the collection tube
17- Centrifuge (spin) at 8000rpm for 1 min
18- Discard the flow-through and collection tube of each sample → put another collection tube

6

19- Add **500 µL Buffer AW1**
20- Centrifuge (spin) at 8000 rpm for 1 min
21- Discard the flow-through and collection tube of each sample → put another collection tube

7

This step ensures that no residual ethanol will be carried over during the following elution

22- Add **500 µL Buffer AW2**
23- Centrifuge at 14 000 rpm for 3 min → to dry the DNeasy membrane
24- Discard flow-through, and **reuse the collection tube**

Remove DNeasy Mini spin column carefully so that the column does not come into contact with the flow-through since this will result in carryover of ethanol

25- Do another centrifugation for 1 min at 14 000 rpm
26- Discard the collection tube, and **place the spin column into an Eppendorf** (microcentrifuge tube)

8

27- Add **30 µL Buffer AE** directly onto the DNeasy membrane

Elution with 30 µL (instead of the 200 µL from the original protocol**) increases the final DNA concentration** in the eluate but also decreases the overall DNA yield

28- Incubate at room temperature for 5 min
29- Centrifuge (spin) at 8 000 rpm for 1 min to elute

9 **TO INCREASE DNA CONCENTRATION for mL of sample**

30- With micropipette, pipette the filtrated into the eppendorf again in the DNeasy membrane column
31- Centrifuge 1 min at 8 000 rpm
32- Discard spin-column

NEXT STEP: DNA Quantification

→ Through Nanodrop: calculate the DNA concentration for mL of sample

PROTOCOL – POLYMERASE CHAIN REACTION (PCR)

Goal: Amplification of DNA fragments

Steps:

1. Keep the reagents on ice
2. Thaw the previously prepared eppendorf(s) by rubbing hands;
3. Spin on vortex (to homogenize the solution);
4. Put back on the ice the eppendorf(s) while pipetting the reagents;
5. Mark/label the PCR reaction eppendorfs;
6. Add the reagents to the eppendorf creating a reagent mix (except the cDNA) → leaving 48 µL of the mix;
7. Vortex;
8. Use a PCR tube to pipette the 48 µL of mix and 2 µL of cDNA template → which will make up 50 µL;
9. Vortex
10. Calibrate the thermal cycler and insert the program, place the samples and when the program is finished, store them.

Mix preparation:

*	Reagent	Quant.
⑧	cDNA template	2 µL
②	10X PCR Buffer	5 µL
③	Mg++ 1,5mM	1,5 µL
④	dNTPs 0,2 mM	0,8 µL
⑤	Primers Forward 0,6 µM	3 µL
⑥	Primers Reverse 0,6 µM	3 µL
①	Sterile water add to a final volume of 47 µL	33,7 µL
⑦	Taq Polymerase (1 unit/µL)	1 µL

Total Volume: 50 µL

*Order in which the reagents are added to the mix

Programm (pcr5830cicli):

1) Desnaturation → 1 cycle 95ºC 10 min
2) Amplification 40 cycle:
   - 95ºC 30 sec
   - 57ºC 30 sec
   - 72ºC 1 min
3) Extension → 1 cycle 72ºC 1 min

Preparation of the concentration of the reagents for the mix:

$$Volume_{initial} = \frac{Concentration_{final} \times Volume_{final}}{Concentration_{initial}}$$

For example:

Magnesium (Mg++) → if the concentration of the commercial Mg is 50mM (Ci), we will dilute it to a Vf of 50 µL:

Vi= (Cf x Vf)/Ci ⇔ Vi= (1,5mM x 50µL)/50 mM

Application to an example

The goal was to amplify two DNA fragments (exons) x2, for this it was necessary to prepare 2 mixes with double the amount to amplify each fragment:

A. exon 1F to exon 9R
B. exon 2F to exon 9R

} *NSD1* gene in chromosome 5

F – for which we will use the forward primer

R – for which we will use the reverse primer

Then we prepare 2 mixes with double the quantity:

Product:	Quant. x2
cDNA template	4 µL
10X PCR Buffer	10 µL
Mg++ 1,5mM	3 µL
dNTPs 0,2 mM	1,6 µL
Primers Forward 0,6 µM	6 µL
Primers Reverse 0,6 µM	6 µL
Sterile water add to a final volume of 47 µL	67,4 µL
Taq Polymerase (1 unit/µL)	2 µL

Note: only 2 µL of cDNA is added to each tube in which a PCR reaction is going to be carried out, as we are going to do 4 amplifications (2 for each set) we use 4 tubes (2 µL for each one).

PROTOCOL – REVERSE TRASCRIPTASE PCR (RT-PCR)

mRNA ⟶ cDNA

Equipment:

- Kit CLONTECH Advantage RT-for-PCR
- Thermal cycler
- Centrifuge

Procedure:

1. Spin the kit reagents;
2. Prepare the 2 mixes for the number of RNA samples we have (keep the reagents cold):

1º mix:

RNA	3 µL
Oligo dT	1 µL
Random Primer	1 µL
H_2O DEPC	5 µL

V_f = 10 µL

2º Mix

Buffer 5X	4 µL
RT	1 µL
RNase inhibitor	0,5 µL
dNTPs	1 µL
H_2O DEPC	3,5 µL

V_f = 10 µL

3. Spin rapidly the 2 mixs
4. Pipette 7µL of the 1st mix + 3µL of RNA on a PCR tube
5. Spin quickly (if there are bubbles it is not necessary)
6. 2 min at 70ºC in thermal cycler (denaturation) rtden70 program
7. Pipet 10µL of the 2nd mix to the PCR tube
8. Spin quickly (if bubbles are present, not necessary)
9. Thermal cycler (incubate at 42°C for 90 minutes; 2 min at 90°C) program rt42º90min
10. At the end of the program add 80µL of H2O to the PCR tube
11. Store in refrigerator (4ºC or -80ºC).

Annex 4

PROTOCOL – PREPARATION OF THE AGAROSE GEL

For the preparation of an Agarose gel of 2%

1. Turn on the laminar flow chamber at 2sd speed and prepare platform
2. **3g of agarose** + **150 mL of Buffer*** (liquid solution)
3. 3 min in the microwave
4. Turn on the hotte at 3rd speed
5. Add the liquid to the platform where it cools and takes shape
6. When cooled down slightly, add **15 µL** of binding nucleic acid (allows you to see the DNA on the gel with radiation) - wait a few seconds
7. Check for bubbles (if there are, remove them with the pipette tip)
8. Wait ≈30 min
9. Store the gel in the cold at -4°C, it can last up to 3 weeks in the cold (in cling film and foil with the respective indication of the agarose percentage and day it was made).

***How to prepare the buffer?** 500mL H20 + 10mL TAE 50X

PROTOCOL – REAL-TIME PCR with SYBR GREEN

Protocol:

- ✓ 1 Mix for the target
- ✓ 1 Mix for the housekeeping gene (β-actine)

Reaction Mix:

②	SYBR Green mix*	10 µL
③	Primer Forward	2 µL
	Primer Reverse	2 µL
①	H_2O	4µL

→ Mix (18 µL) + ④ 2µL cDNA (template) → V_f= 20 µL

*includes TaqPolimerase

The 18µL + 2µL are pipetted onto the qPCR plate;
The cDNA is the last thing to be pipetted

Steps:

1- Prepare the mix for the number of samples (usually done for 1 or more);

2- In the qPCR plate, pipette 18µL of the mix of the gene we are studying, and 2µL of cDNA from each sample;

3- Seal the plate with optical adhesive;

4- Centrifuge the plate for 2 min at 4ºC at 1500 rpm;

5- Turn on the computer →open the LightCycler program → turn on the equipment → press "on and eject" → place the plate on the real-time PCR equipment → close it and press the button again → choose the option "New experiment from template" and choose the giusy program to start the race

6- After the end of the run (≈1:30h / 2h), check for the:

- ***Melting peaks*** → if there is only 1 peak it means that there was a good specificity in the amplification, if there are 2 peaks it means that there was not a good specificity (there were 2 amplicons);

- Write down the **Cp values**→ the Cp values are the quantity of times that the target gene was expressed
  ↑ Cp → less expressed gene
  ↓Cp → more expressed gene

7- Record the data on an excel sheet at the end, remove the plate from the qPCR and discard it.

Real-Time PCR plate

PROTOCOL – SANGER SEQUENCING FOR cDNA

1- Thaw and spin the reagents

2- Dilution of the primers (1:10)

3- Preparation of a mix for each purified PCR product (and for each primer– forward and reverse).

Mix:

			x3*
②	**Buffer**	2µL	6 µL
④	**Big Dye**	0,4 µL	1,2 µL
①	**H_2O**	3,6 µL	10,8 µL
	DNA	3 µL	
③	**Primer Forward**	1 µL	3 µL

V_f= 10 µL

		x3
Buffer	2µL	6 µL
Big Dye	0,4 µL	1,2 µL
H_2O	3,6 µL	10,8 µL
DNA	3 µL	
Primer Reverse	1 µL	3 µL

V_f= 10 µL

* x3 because we are doing in duplicate (x2) plus one (x3) to make sure that our quantity is sufficient, since due to pipetting errors µL are lost

After preparing a mix for each primer,

4- Prepare 8 PCR tubes (for example):

5-

1) CD19 Forward 1 2) CD19 Forward 2 3) CD19 Reverse 1 4) CD19 Reverse 2

5) RNU Forward 1 6) RNU Forward 2 7) RNU Reverse 1 8) RNU Reverse 2

6- Add to each PCR tube: 7µL of mix + 3µL of cDNA

7- Spin

8- Thermo Cycler (programm: folder – seq → seq std)

PROTOCOL – PURIFICATION OF PCR PRODUCTS FOR SEQUENCING

1- Thaw and centrifuge the reagents

2- Dilute the PCR product with 40µL of H2O

3- Mix with the micropipette the PCR product and the water

4- Transfer the mix to a plate of sequencing products

5- 3 minutes in the machine that sucks/filters the water

6- Dry (on paper) and add 30µL of H2O to the wells

7- Put in shaker for 10 minutes

8- Transfer to PCR tubes

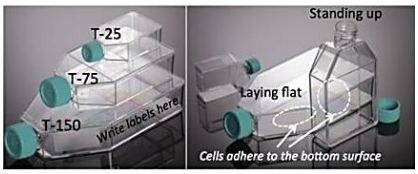
PROTOCOL – FIBROBLASTS CELL CULTURES

Composition of complete medium for cell line (V_f=50mL)

- 1 mL L-glutamine 100x
- 1mL Penicillin-Streptomycin (antibiotic) 100x
- 5 mL Fetal Bovine Serum (10%)
- 43 mL DMEM (Dulbecco's Modified Eagle Medium) or RPMI 1640 Medium

Composition of freezing medium for cell lines (V_f=10mL)

- 1mL DMSO (dimethyl sulfoxide) (10%)
- 9mL complete medium



Maintenance of cell lines in culture

- 4-5 days after set up, change the medium
- Monitor cell growth and change medium every 2-3 days
- After about 2 weeks the number of cell growth areas should be sufficient to trypsinize the flask *in situ*.
- When the trypsinized *in situ* flask reaches 80-90% confluence we proceed with a new trypsinization and division into 2 flasks.

Trypsinization procedure

This procedure is used to expand the initial number of cells and allows to obtain a cell suspension that, properly diluted, can be transferred into new flasks and allow the growth of new cells.

- Remove the culture medium completely from the flasks
- Wash with sterile physiological solution (PBS or NaCl) in order to remove all the medium
- Remove the physiological solution
- Add to the flask 0.5 ml of trypsin for T25 flasks or 1 ml in case of T75 flasks
- Leave in thermostat at 37°C for 5 minutes to allow all the cells to detach from the adhesion surface
- Slightly shake the flask so that all adherent cells are well dislodged
- Add 3 ml of culture medium for T25 flasks or 10 ml for T75 flasks, pipetting repeatedly in order to block the action of trypsin and to allow the cells to distribute evenly on the surface
- Make a medium change when cells are fully adhered (at least after 2-3 hours or overnight)

Freezing Procedure

From a confluent T25 flask, freeze one cryotube; for each cell line freeze a minimum of 6 cryotubes on at least two different days according to the procedure below:

- Trypsinize the flask and blot with culture medium
- Transfer cells to be frozen in a falcon and centrifuge at 1200 rpm for 10'
- Discard the supernatant and resuspend the pellet in 1 ml of cold freezing medium
- Shake well with a pasteur, transfer to a cryotube and place immediately in an ice cube tray
- Transfer cryotubes into a freezing container (Cool Cell) and place in the ultrafreezer-80°C for about 10-12 hours
- Transfer the cryotube into the dewar

Rute Luísa Cabrita Pinto

INTERNSHIP REPORT AT THE *GIANNINA GASLINI* INSTITUTE

1290 UNIVERSIDADE D COIMBRA

1 2 9 0

UNIVERSIDADE D
COIMBRA

Rute Luísa Cabrita Pinto

INTERNSHIP REPORT AT THE *GIANNINA GASLINI* INSTITUTE

MASTER'S DEGREE IN LABORATORY CLINICAL GENETICS

Curricular Internship Report in the context of the Master's Degree in Laboratory Clinical Genetics, oriented by Dr. Domenico Coviello M.D. Ph.D. and Prof. Isabel Maria Marques Carreira, M.D. Ph.D., presented to the Faculty of Medicine of the University of Coimbra.

August of 2022