1 2 9 0

UNIVERSIDADE D
COIMBRA

Carolina Pereira Teixeira

MOLECULAR DIAGNOSIS OF NEUROFIBROMATOSIS TYPE I AND OTHER RASOPATHIES

Relatório de Estágio no âmbito do Mestrado em Genética Clínica Laboratorial, orientado pela Doutora Teresa de Jesus Semedo Fidalgo e pela Professora Doutora Maria Joana Lima Barbosa de Melo e apresentado à Faculdade de Medicina da Universidade de Coimbra.

Setembro de 2023

MOLECULAR DIAGNOSIS OF NEUROFIBROMATOSIS TYPE I AND OTHER RASOPATHIES

Carolina Pereira Teixeira

Relatório de Estágio no âmbito do Mestrado em Genética Clínica Laboratorial, orientado pela Doutora Teresa de Jesus Semedo Fidalgo e pela Professora Doutora Maria Joana Lima Barbosa de Melo e apresentado à Faculdade de Medicina da Universidade de Coimbra.

Setembro de 2023

1 2 9 0

FACULDADE DE MEDICINA
UNIVERSIDADE D
COIMBRA

Agradecimentos

Um obrigada à Professora Doutora Isabel Carreira e à Professora Doutora Joana Barbosa de Melo, também orientadora deste trabalho, pela criação do Mestrado em Genética Clínica Laboratorial que me permitiu adquirir as bases para as minhas perspetivas profissionais futuras.

Um agradecimento à Doutora Teresa Fidalgo por me orientar neste trabalho, pelo desafio, pela confiança e pela oportunidade de realizar o meu projeto de estágio na Unidade Funcional de Hematologia Molecular do CHUC.

Um agradecimento especial à Patrícia Martinho por coorientar o meu trabalho, por todos os ensinamentos na parte prática do laboratório, todas as discussões e dores de cabeça partilhadas e pela paciência. Obrigada!

Ao Doutor Mário Laço um especial obrigada por toda a ajuda, todo o suporte na parte clínica, disponibilidade e por ter trazido o estudo de neurofibromatose tipo I para a UFHM, que levou à possibilidade de realização deste trabalho.

A toda a equipa da Unidade Funcional de Hematologia Molecular, do CHUC, por toda a aprendizagem e convívio ao longo deste ano: Ana Catarina Oliveira, Ana Teresa Simões, Catarina Silva Pinto, Celeste Bento, Elizabete Cunha, Janet Pereira, Luís Relvas, Margarida Coucelo, Patrícia Martinho e Teresa Fidalgo. E às minhas companheiras de estágio: Ana, Cristiana e Mariana.

Aos meus amigos, em especial aos que me acompanharam ao longo destes últimos 5 anos académicos: Ana, Cláudia e Eliana. E aos meus colegas de mestrado, em especial à Daniela, Ema e Inês. Obrigada pela vossa amizade!

Um agradecimento especial a toda a minha família, irmão, tios, primos, avós, e ao meu namorado pelo vosso apoio incondicional.

Aos meus pais, Janet e Carlos, um agradecimento muito especial, por estarem sempre presentes, pelo vosso apoio em todos os momentos da minha vida, e pela oportunidade de frequentar este mestrado. À minha mãe, com quem partilho esta paixão pela genética, obrigada pela partilha do teu conhecimento, experiência e por todo o teu suporte ao longo deste trabalho. Ao meu pai, obrigada pelo teu incentivo e curiosidade, mesmo não percebendo nada de biologia molecular. Sem vocês não seria possível! São um grande exemplo para mim! Obrigada por tudo!

Index

Resumo

Introdução: As RASopatias são um grupo de síndromes genéticas, causadas por variantes patogénicas na linha germinativa num dos constituintes da via de sinalização *RAS/mitogen-activated protein kinase* (RAS/MAPK). Até à data, cerca de 20 genes foram associados a RASopatias, que incluem a neurofibromatose tipo I (NF1), e é um dos maiores grupos de síndromes de malformação com características clínicas comuns. A neurofibromatose tipo I tem uma incidência entre 1 em 2500 a 1 em 3000 indivíduos e é uma doença autossómica dominante. Caracteriza-se principalmente por múltiplas manchas "café-com-leite" (CALMs), neurofibromas cutâneos (cNFs) e subcutâneos (sNFs), harmatomas da iris (nódulos de *Lisch*), sardas na região axilar e inguinal e neurofibromas plexiformes. NF1 é causada por variantes de perda de função no gene *neurofibromin 1*, um gene supressor tumoral. Mais de 4000 variantes patogénicas estão reportadas no HGMD® Profissional e cerca de 30% das variantes patogénicas encontradas no *NF1* afetam o splicing. *NF1* codifica a neurofibromina, uma proteína ativadora de *Ras-guanosine triphosphatase* (GTPase), que regula negativamente a proteína RAS. O estudo molecular de *NF1* não é fácil, uma vez que, é um gene grande, sem "hotspot" e com um elevado número de variantes que alteram o *splicing* e que não são facilmente detetadas. Esta abordagem deve incluir a análise do DNA genómico (gDNA) e do DNA complementar (cDNA). NF1 tem inúmeras manifestações clínicas que afetam diferentes partes do corpo, sendo a remoção por cirurgia o principal tratamento.

Objetivos: 1) Aprendizagem, familiarização e aquisição de competência técnica na utilização das técnicas de biologia molecular. 2) Aprender e adquirir competências técnicas na classificação de variantes genéticas. 3) Completar o processo de implementação de um protocolo para o estudo molecular de *NF1* na Unidade Funcional de Hematologia Molecular - Centro Hospitalar e Universitário de Coimbra (UFHM). 4) Realização de diagnóstico molecular em doentes com suspeita clínica de NF1. 5)Desenvolvimento de um painel de sequenciação de nova geração (NGS) para RASopatias.

Materiais e Métodos: Neste trabalho, foram estudados 40 indivíduos, de 34 famílias diferentes, com suspeita clínica de NF1. O estudo sequencial de *NF1* desenvolvido na

UFHM inicia-se com o estudo da região codificante por NGS, seguindo para o estudo de *copy number variation* (CNV) por *multiplex ligand probe amplification* (MLPA) e finalizando com a análise do cDNA. Os doentes com resultados negativos para NF1 (gDNA e cDNA) devem ser incluídos no estudo das RASopatias para diagnóstico diferencial de NF1.

Resultados e Discussão: Foram identificadas 19 variantes diferentes, em 20 de 34 probandos, quatro das quais *novel* (c.5692dup, c.2614G>T, c.1392+751T>G, c.(3974+10_3982)_(6250_6387)del). Sessenta e um por cento dos doentes foram positivos para NF1 e 39% foram negativos. O estudo NGS revelou 16 variantes diferentes, uma grande deleção foi encontrada por MLPA e a sequenciação de Sanger do cDNA permitiu a identificação de uma variante *deep* intrónica e a confirmação da variante de splice previamente identificada (foi detetado o *skipping* do exão 31). Seis variantes, incluindo duas das novas variantes (c.2614G>T and c.1392+751T>G), foram classificadas segundo o sistema de classificação de variantes de *splicing* de *NF1*, com base no seu efeito no evento de *splicing*. Não foram encontradas novas correlações fenótipo-genótipo. Dos 14 doentes negativos para o estudo de *NF1*, seis puderam ser estudados para o diagnóstico diferencial de RASopatias por sequenciação completa do exoma (WES) aplicando um painel virtual de RASopatias. Foram encontradas anomalias moleculares em quatro destes seis doentes: em dois doentes, foram detetadas duas novas variantes provavelmente patogénicas, c.1769A>G no *SOS1*, e c.410C>A no *LZTR1*, genes associados à síndrome de Noonan; nos outros dois doentes, foram identificadas duas grandes deleções: no *MAP2K2* e no *HRAS*.

Conclusão: Esta estratégia de duas etapas revelou ser altamente eficiente, uma vez que a combinação de três métodos diferentes e complementares (sequenciação de gDNA e de cDNA e MLPA) permitiram identificar 19 variantes diferentes, quatro das quais não tinham sido previamente descritas. A inclusão da sequenciação do cDNA no algoritmo de diagnóstico foi fundamental para compreender o impacto das variantes do DNA no mRNA do *NF1*. Para uma melhor e correta avaliação da sensibilidade deste estudo, devem ser excluídos os diagnósticos diferenciais de NF1 em todos os doentes negativos e mais indivíduos devem ser estudados.

Palavras-chave: Neurofibromatose Tipo I; RASopatias; NGS

Abstract

Introduction: The RASopathies are a group of genetic disorders, caused by germline pathogenic variants in one of the constituents of the RAS/mitogen-activated protein kinase (RAS/MAPK) signalling pathway. To date, approximately 20 genes have been associated with RASopathies, including neurofibromatosis type I (NF1), and it is one of the largest groups of malformation syndromes with overlapping features. Neurofibromatosis type I has an incidence of between 1 in 2500 to 1 in 3000 individuals and is an autosomal dominant disorder. It's mainly characterized by multiple café-au-lait macules (CALMs), cutaneous and subcutaneous neurofibromas (cNFs and sNFs), iris hamartomas (Lisch nodules), freckling of axillary and inguinal regions and plexiform neurofibromas (pNFs). NF1 is a disease caused by loss-of-function variants in the neurofibromin 1 gene, a tumor suppressor gene. More than 4000 pathogenic variants are reported in the HGMD® Professional and around 30% of the pathogenic *NF1* DNA variants affect the splicing event. *NF1* encodes neurofibromin a Ras-guanosine triphosphatase (GTPase) activating protein (RAS-GAP) that acts as a negative regulator of RAS protein. Molecular testing of *NF1* is difficult because it's a large gene, there is no mutation hotspot, and a significant number that of variants that affects splicing are not easily detected. This approach should include analysis of both genomic DNA (gDNA) and complementary DNA (cDNA). NF1 has many clinical features that affect different parts of the body and surgical removal is the main treatment.

Objectives: 1) Learning, familiarization, and acquisition of technical skills in the molecular techniques. 2) To learn and acquire of technical skills in the classification of genetic variants. 3) To complete the process implementing a protocol for the molecular study of *NF1* at the Unidade Funcional de Hematologia Molecular – Centro Hospitalar e Universitário de Coimbra (UFHM). 4) Performing molecular diagnosis in patients with clinical suspicion of NF1. 5) Development of a next-generation sequencing panel (NGS) for RASopathies.

Materials and Methods: In this study, 40 individuals, from 34 unrelated families with NF1 clinical suspicion were studied. The sequential study of *NF1* developed at UFHM

starts with the study of the coding region by NGS, proceeds to the study of copy number variations (CNV) by multiplex ligand probe amplification (MLPA) assay and ends with cDNA analysis. Patients with negative results for NF1 (gDNA and cDNA) should be included in the RASopathies study for differential diagnosis of NF1.

Results and Discussion: Nineteen different variants have been identified in *NF1,* in 20 of 34 probands, four of which were *novel* (c.5692dup, c.2614G>T, c.1392+751T>G, c.(3974+10_3982)_(6250_6387)del). Sixty-one percent of the patients were positive for *NF1* and 39% were negative. The NGS study revealed 16 different variants, one large deletion was found by MLPA, and cDNA Sanger sequencing allowed the identification of one deep intronic variant and the confirmation of the splice variant previously identified (skipping of exon 31 was detected). Six variants, including two of the *novel* variants (c.2614G>T and c.1392+751T>G), were classified according to the *NF1* splicing variant classification system, based on their effect on the splicing event. No new phenotype-genotype correlations were found. Of the 14 NF1-negative patients, six could be investigated for the differential diagnosis of RASopathies by whole exome sequencing (WES) using a virtual RASopathies panel. Among of the six patients studied by WES, in four cases gene defects were found: in two patients, two *novel* likely pathogenic variants were identified, c.1769A>G in *SOS1* and c.410C>A in *LZTR1*, genes associated with Noonan syndrome; in the other two patients, were identified two gross deletions in *MAP2K2* and in *HRAS*.

Conclusion: This two-step strategy proved to be highly efficient, as the combination of three different and complementary methods (gDNA and cDNA sequencing, and MLPA) it was able to identify 19 different variants, four of which had not been previously described. The addition of cDNA sequencing to the diagnostic algorithm was fundamental to understanding the impact of DNA variants on *NF1* mRNA. For a better and correct evaluation of the sensitivity of this study, the differential diagnoses of NF1 in all the negative patients should be excluded and more individuals should be studied.

Key-words: Neurofibromatosis Type I; RASopahties; NGS

List of Abbreviations

ACMG/AMP - American College Medical Genetics/Association for Molecular Pathology	**mRNA –** Messenger RNA
CALMs - Café-au-lait macules	**NF1 -** Neurofibromatosis Type I
CAs - Choroidal abnormalities	**NF2 –** Neurofibromatosis Type II
cDNA - Complementary DNA	**NGS -** Next Generation Sequencing
CFC - Cardio-facio-cutaneous syndrome	**NIH -** National Institutes of Health
CM-AVM - Capillary malformation arteriovenous malformation syndrome	**NLS -** Nuclear localization signal
CMMRD - Constitutive Mismatch Repair Deficiency	**NS -** Noonan syndrome
cNFs - Cutaneous neurofibromas	**NSML -** Noonan syndrome with multiple lentigines
CNV - Copy number variation	**OMIM –** Online Mendelian Inheritance in Man
CS - Costello syndrome	**OPG –** Optic pathway glioma
CSRD - Cysteine-serine-rich domain	**PCR-** Polymerase Chain Reaction
CTD - C-terminal domain	**PH -** Pleckstrin homologous domain
DNA - Deoxyribonucleic acid	**PND -** Prenatal diagnosis
FD/MAS - Fibrous dysplasia/McCune-Albright syndrome	**pNFs -** Plexiform neurofibromas
GAP - Guanosine triphosphatase activating protein	**QC** – Quality Control
gDNA - Genomic DNA	**RAS-GAP -** Ras-guanosine triphosphatase activating protein
gnomAD - Genome Aggregation Database	**RAS/MAPK -** RAS/mitogen-activated protein kinase
GRD - GAP-related domain	**RNA -** Ribonucleic acid
GTPase - Guanosine triphosphatase	**ROI** – Region of interest
HGMD® - Human Gene Mutation Database	**SBR -** Syndecan-binding regions
HLR - Heat-like repeats	**Sec -** Sec14 homologous domain
LOVD - Leiden Open Variation Database	**sNFs -** Subcutaneous neurofibromas
LRD - Leucine-rich domain	**SNV -** Single-nucleotide variation
LS - Legius syndrome	**TBD -** Tubulin-binding domain
MLPA - Multiplex Ligand Probe Amplification	**UFHM -** Molecular Hematology Functional Unit
MPNST - Malignant peripheral nerve sheath tumors	**WES –** Whole Exome Sequencing

List of Figures

List of Tables

Tables

Supplementary Tables

Chapter 1

Curricular Internship Contextualization

Chapter 1. Curricular internship contextualization

The Molecular Hematology Functional Unit (UFHM) of the Clinical Haematology Service is part of the Oncology Department of the Hospital and University Centre of Coimbra, E.P.E. (CHUC).

The UFHM's main area of expertise is in the following diseases: hemato-oncology, erythropathology and iron metabolism, thrombosis and hemostasis, microangiopathies (thrombotic thrombocytopenic purpura and haemolytic-uremic syndrome).

Clinical-laboratory correlation is the basis for the diagnostic efficiency of a multidisciplinary team at the UFHM, which is proficient in the different technological areas of molecular biology required for the diagnosis and monitoring of hematological diseases and other molecular genetic studies. The integration between clinical, diagnostic, research and continuous training is very important to guarantee the quality and excellence of the services provided.

The laboratory is ISO 9001 certified and integrates a quality management system with External Quality Control activities through the ECAT Foundation, NEQAS and GenQA.

Well equipped with the latest technologies, the UFHM team is dedicated to the diagnosis of these disorders and develops applied research in collaboration with other national and international scientific groups.

This curricular internship was part of the Master in Clinical Laboratory Genetics of the Faculty of Medicine of the University of Coimbra and took place at the UFHM in collaboration with the Medical Genetics Service of the CHUC. It lasted eleven months, from 19 September 2022 to 31 August 2023, under technical supervision of Dr Teresa Fidalgo and Dr Patrícia Martinho.

At the UFHM, first I had the opportunity to learn the main techniques of molecular biology - DNA and RNA extraction, PCR, DNA gel electrophoresis, DNA sequencing (Sanger and Next Generation) and MLPA - until I was actively involved in the laboratory routine.

In a second phase, my internship focused on the molecular diagnosis of patients with suspected or clinically diagnosed neurofibromatosis type I (NF1), under the clinical supervision of geneticist Dr Mário Laço. The aim of this study was to implement the two-step strategy for NF1 testing, gDNA and cDNA, and the applicability of the most recent software and databases for the interpretation of results and the preparation of clinical reports.

Chapter 2

Introduction

Chapter 2. Introduction

2.1. RASophaties

The RASopathies are a group of genetic disorders, caused by germline pathogenic variants in one of the constituents of the RAS/mitogen-activated protein kinase (RAS/MAPK) pathway (Figure 1)[1–3]. This pathway has an important role in cell development processes, such as proliferation, survival, differentiation, and metabolism[1,2]. Dysregulation of this pathway can occur by various mechanisms, most of which culminate in its increased activation. Consequently, this increase in RAS/MAPK pathway signal alters normal cell development processes, stimulates cell proliferation and cell survival, and predisposes to the occurrence of benign and/or malignant tumors[1,2].

To date, about 20 genes have been associated with RASopathies[3], making them one of the largest groups of malformation syndromes[1,4]. Commonly, neurofibromatosis type I (NF1), Noonan syndrome (NS), Noonan syndrome like syndrome, Noonan syndrome-like disorder with loose anagen hair, Noonan syndrome with multiple lentigines (NSML; LEOPARD syndrome), cardio-facio-cutaneous syndrome (CFC), Costello syndrome (CS), Legius syndrome (LS) and capillary malformation arteriovenous malformation syndrome (CM-AVM) are all recognized as RASopathies (Figure 1)[1,2].

Each of these syndromes results from germline variants in different genes related to the RAS/MAPK pathway[1]. All have in common the fact that they disrupt this pathway, although they may affect it through different mechanisms[1,4]. The involvement of a common signaling pathway (RAS/MAPK pathway) explains the overlap of features between several of these diseases, while the association of each one to specific genes might be responsible for their own unique features[1,4]. Common features include cardiac malformations, dysmorphic facial features, growth problems, intellectual disability, cancer risk, bone alterations and skin, eyes, and muscles abnormalities[1,4].

Noonan syndrome and neurofibromatosis type I are the most frequent RASopathies[3].

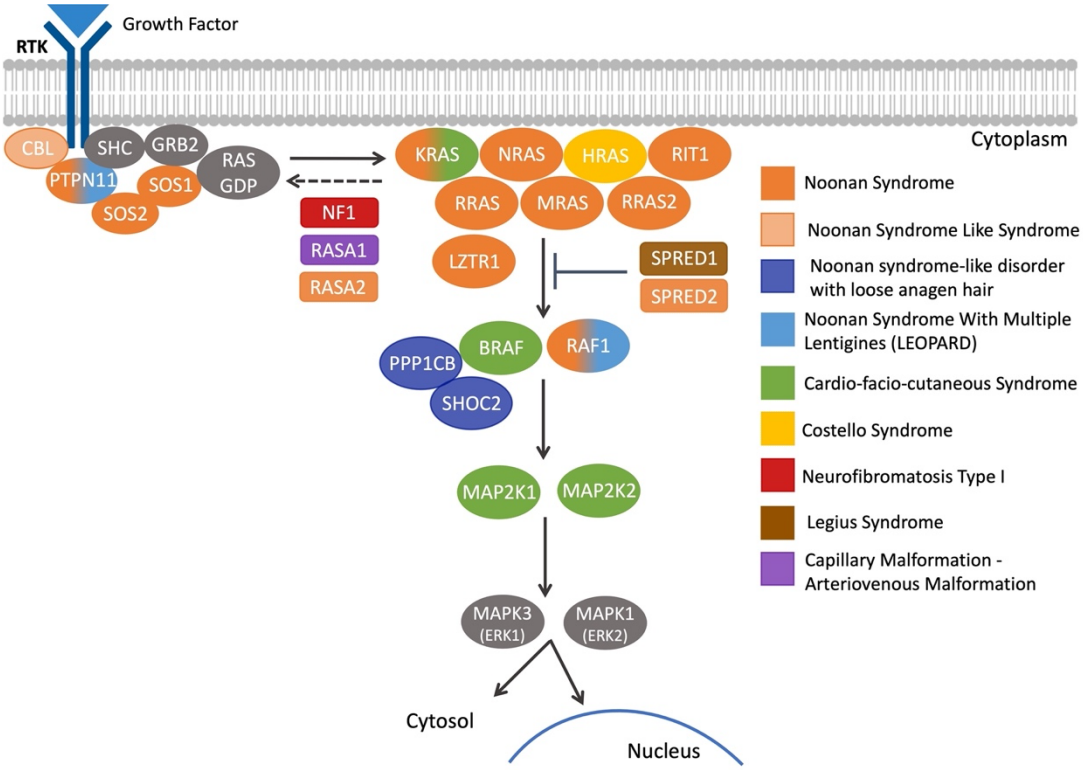

Figure 1: Schematic representation of RAS/MAPK pathway and associated genetic syndromes. Adapted from: Rauen et al.[1] and Aoki et al.[48]

2.2. Neurofibromatosis Type I

Neurofibromatosis type I (OMIM #162200), a RASopathy described by Friedrich Daniel von Recklinghausen in 1882, is one of the most frequent genetic disorders, with an incidence of 1 in 2500 to 1 in 3000 individuals[5,6]. NF1 is an autosomal dominant disorder, and its penetrance is almost complete by adulthood[6,7]. The exception is a small number of individuals, in whom molecular testing has shown incomplete penetrance of pathogenic *NF1* variants[8].

This Mendelian disease is mainly characterized by multiple café-au-lait macules (CALMs), cutaneous and subcutaneous neurofibromas (cNFs and sNFs), iris hamartomas (Lisch nodules), freckling of axillary and inguinal regions and plexiform neurofibromas (pNFs)[5–7,9,10]. Other features such learning disabilities, optic gliomas, bone abnormalities, an increased risk of particular malignancies (e.g. malignant peripheral nerve sheath tumors - MPNST), scoliosis, pseudoarthrosis, short stature, macrocephaly and cardiovascular disease may also be seen[6,7,9,10]. The diagnosis of NF1 is based on

clinical criteria defined in a National Institutes of Health (NIH) Consensus Development Conference, in 1987 and further revised in 2020[9,11]. NF1 should be suspected in an individual who presents two or more of the following specific features: six or more café-au-lait macules greater than 5 mm in its greatest diameter in prepubertal individuals or greater than 15 mm in its greatest diameter in post pubertal individuals; axillary or inguinal freckling; two or more neurofibromas of any type or one plexiform neurofibroma, optic pathway glioma (OPG); two or more iris Lisch nodules or two or more choroidal abnormalities (CAs); a distinctive osseous lesion such as sphenoid dysplasia, anterolateral bowing of the tibia, or pseudarthrosis of a long bone; a heterozygous pathogenic NF1 variant; and parent who meets the diagnostic criteria[8,12]. However, the full clinical manifestations appear progressively over time, usually from the age of eight years, and in the absence of a positive family history, the clinical signs alone have low sensitivity, making a clinical diagnosis difficult to establish during childhood[11–14]. In a child under eight years of age, molecular diagnosis could be considered to confirm the diagnosis of NF1 in individuals with only CALMs and skinfold freckling (with or without a family history) and no other clinical criteria[8,12].

2.2.1. Causes of Neurofibromatosis Type I

NF1 is a disease caused by loss-of-function variants in the neurofibromin 1 gene (*NF1*; OMIM *613113; NM_000267.3)[7,9,15,16]. *NF1* is a tumor suppressor gene, located on chromosome 17q11.2, with 57 exons (Figure 2)[7,9,17]. The same gene can generate different mRNAs transcripts, which leads to an increased coding capacity. These different mRNAs transcripts result from an ingenious mechanism known as alternative splicing, which has an important role in the developmental and cell-type specific control of gene expression[6,7,18]. Around 30% of pathogenic *NF1* DNA variants affects this process, causing an aberrant splicing event. Apart from that, alternative splicing events can also occur naturally and there are a few normal alternative transcripts have been described for *NF1*[6,7,15]. This gene has at least four alternatively spliced exons (Figure 2), which extend the sequence but do not change the reading frame, namely: between exons 11 and 12, 30 bp can be added (11alt12, formerly 9a); between exons 12 and 13, 45 bp can be included (12alt13, formerly 10a-2); between exons 30 and 31, additional 63 bp can be present (30alt31, formerly 23a); and between exons 56 and 57, more 54

bp can be in the reading frame (56alt57, formerly 48a). Additionally, it can occur skipping of exons 37 and/or 38 (formerly 29 e 30, respectively) and skipping of exon 51 (formerly 43)[5,6,17,19–21].

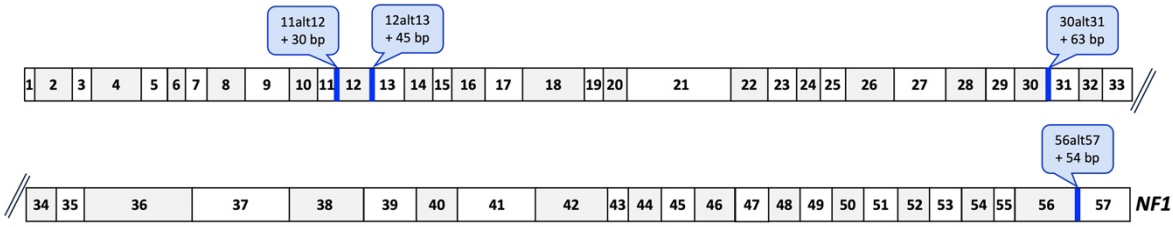

Figure 2: Schematic representation of the 57 exons of *NF1* and the position of the alternatively spliced exons (11alt12, 12alt13, 30alt31, 56alt57). The first row shows exons 1 to 33 and the second row shows exons 34 to 57. Adapted from: Anastasaki et al.[17]

Currently, over 4000 pathogenic variants are reported in the Human Gene Mutation Database (HGMD® Professional 2023.2). These variants include missense, nonsense, frameshift, insertions, deletions, and splice site variants. Half of cases of NF1 are caused by *de novo* variants, with the other half being inherited[19,22].

NF1 encodes neurofibromin, a protein containing 2818 amino acids (Figure 3), involved in the regulation of the RAS/MAPK pathway (Figure 1)[23]. Neurofibromin is a Ras-guanosine triphosphatase (GTPase) activating protein (RAS-GAP), which works as a negative regulator of RAS protein[7,9,24]. Thus, loss of neurofibromin function, such as that caused by NF1-associated variants, results in a decrease in neurofibromin GAP function and a consequent increase in overall RAS signal, which affects cell proliferation and cell differentiation[7,24,25]. Neurofibromin is composed of various domains (Figure 3): an cysteine-serine-rich domain (CSRD), a tubulin-binding domain (TBD), a central GAP-related domain (GRD), two syndecan-binding regions (SBR), a Sec14 homologous domain (Sec), a pleckstrin homologous domain (PH), a leucine-rich domain (LRD), a heat-like repeats (HLR), and a C-terminal domain (CTD) with a nuclear localization signal (NLS)[10,19,23,25–27].

Figure 3: Schematic representation of domains of the neurofibromin protein. CSRD between 543-909, TBD between 1095-1198, GRD between 1198-1530, SBR between 1357-1473 and 2619-2719, Sec between 1550-1698, PH between 1715-1816, LRD between 1579-1971, HLR between 1825-2428 and CTD between 2260-2818 with NLS. Adapted from: Bergoug et al.[19] and Mo et al[27].

The GRD is the most highly conserved region and the best studied domain of the protein. It is located between exons 27 and 35[10,19,26]. Its main function is hydrolysis of RAS-GTP, active form, in RAS-GDP, inactive form, inhibiting downstream RAS signaling[10,19,23,26]. The previously described alternative splicing of *NF1* pre-mRNA results in different isoforms of neurofibromin[19].

2.2.2. Genotype-Phenotype Correlation

The most severe forms, large deletions, which include deletion of the entire *NF1* gene, are related with more severe, more frequent, and atypical manifestations with an earlier onset[8,28]. Plexiform or spinal neurofibromas, optic pathway gliomas, malignant neoplasms, and skeletal abnormalities, which is an uncommon critical phenotype, have been reported in individuals with missense variants affecting one of five codons between 844 and 848[8].

Other missense variants and a small deletion have been associated with specific phenotypes, such as the missense variants affecting amino acids p.Arg1809, p.Met1149, and p.Met992del, which are associated with the absence of dermal neurofibromas[8,29]. On the other hand, p.Arg1038Gly, in addition to the absence of neurofibromas, has been associated with mild pigmentary features[8]. Cardiovascular malformations, especially pulmonary artery stenosis, and the NF1-Noonan phenotype have been observed with higher frequency with in missense variants affecting p.Arg1276 and p.Lys1423, two highly conserved residues[8,30]. The p.Arg1276 residue is also called the arginine “finger” and it is an important catalytic element for RAS-GAP activity. Several studies have shown that variants affecting these residues severely reduce GAP activity[19,30].

However, this genetic disorder has a highly variable clinical expression between families as well as within the members of the same family[31]. In fact, in most cases of NF1, it is difficult to predict the phenotype based on the genetic variants[29], because there is no a clear genotype-phenotype correlation[31].

2.2.3. Molecular Diagnosis

Molecular testing of *NF1* might not be straightforward because: *NF1* is a large gene; there is no clear mutation hotspot (which makes it necessary to evaluate all exons); and a significant number of variants affecting alternative splicing are not easily detected by regular genomic DNA (gDNA) analysis or not predicted by bioinformatics tools, such as in the case of deep intronic variants[7,24]. For these reasons, a multistep protocol has been recommended, in order to detect the different types of variants present in NF1 patients. This approach should include analyses of both gDNA and complementary DNA (cDNA)[11].

In this proposed protocol, the first step consists in the regular analysis of *NF1* coding region through gDNA sequencing. The fact that *NF1* is a large gene and that all exons must be evaluated, makes it very laborious and time consuming to approach by conventional PCR followed by Sanger sequencing. For this reason, next generation sequencing (NGS) should be the technique of choice[9]. NGS is a sequencing technology that revolutionized molecular diagnosis[32]. It consists in a method that sequences a huge number of DNA fragments, in parallel, at the same time[33]. NGS methodology is fundamental for the contemporary need of rapid genetic diagnosis, since it allows the sequencing of different gene regions simultaneously[34]. In addition, for rare diseases with high genetic heterogeneity, it allows the simultaneous study of multiple genes in a single test[34]. Besides permitting the analysis of different samples simultaneously, NGS has a higher capacity to detect uncommon variants and increased sensitivity compared to Sanger sequencing[32]. However, NGS has also its limitations when it comes to detecting copy number variation (CNV). Therefore, a multiplex ligand probe amplification (MLPA) is thoroughly evaluated CNV in this region and complete the study of *NF1*.

MLPA is commonly used to identify whole or partial gene deletions or duplications[9,11]. MLPA is a multiplex PCR that amplifies different probes, specifically

designed to hybridize with the region of interest. This process quantifies the DNA sequences of interest, identifying deletions and duplications[35]. It is a highly sensitive technology, with no time-consuming sample preparation and low cost[35]. In conclusion, sequencing the entire coding region of the *NF1* gene by NGS, complemented with MLPA to detect CNV, is the preferred strategy to analyze gDNA [9,11,15,24,36].

A second step is applied when the *NF1* gDNA results are negative. In this additional step, *NF1* cDNA is amplified by conventional PCR and subsequently sequenced by Sanger sequencing. This second step allows the identification of variants not previously detected variants by gDNA analysis, namely intronic and/or splicing variants. This two-step strategy seems to increase the sensitivity of molecular diagnosis of NF1 patients, reaching values over 95%[24].

2.2.4. Treatment and Surveillance

NF1 has a plethora of clinical features, affecting different parts of the body and, therefore, in need of specific follow-up[29]. As previously mentioned, the onset of manifestations is progressive throughout life, there is a high variability of clinical signs between affected individuals and, in most cases, they cannot be predicted in advance[8,22,29]. According to the American College Medical Genetics (ACMG) guidelines for care of adults and children with NF1 (published in 2018 and 2019, respectively), surgical removal is the main treatment, and these patients should have regular surveillance[8,37,38].

The surgical treatment remains the main treatment for high-grade MPNST[22,37]. It is also possible for small plexiform neurofibromas[8]. For cNFs surgical excision, laser removal, or electrodesiccation are available[8,37]. In all cases, an experienced medical team in NF1 must be required[22].

However, chemotherapy and radiotherapy have shown some promising results in MPNST but their benefits need to be further investigation[8,37]. For progressive OPG, chemotherapy is available with mixed results[8]. In other tumors, such as plexiform neurofibromas, brain tumors and OPG, radiotherapy is contraindicated due to the risk of secondary tumor development in the irradiated area[8,37].

Specific medication and other therapies are under development, such as MAP2K1 and MAP2K2 inhibitors, immunotherapy and/or radiotherapy inhibitors[8,37].

NF1 adults have been reported to exhibit low bone mineral density and osteoporosis[37]. Vitamin D supplementation has been indicated for individuals with NF1, although there are no clear data to support its efficacy[8,37].

For neurocognitive and psychiatric problems, the approach is similar to that recommended for non-NF1 individuals[8,37].

2.2.5. Genetic Counseling

The children of a patient with NF1 have a 50% chance of inheriting the disease. Prenatal diagnosis (PND) and preimplantation genetic diagnosis are possible as long as the pathogenic variant has been identified in the family member[8]. In the case of a positive prenatal diagnosis, medical termination of pregnancy may be proposed, subject to evaluation by an ethics committee.

2.2.6. Differential Diagnosis

In the study of NF1, other conditions with similar features, such as CALMs, should be considered in the differential diagnosis. Distinguishing it from other RASopathies, namely with Noonan syndrome, Noonan syndrome with multiple lentigines and Legius syndrome can be challenging due to the occurrence of identical manifestations, such as macrocephaly, Noonan-like facial features, short stature, learning disabilities and the presence of CALMs[8,39]. In young children, distinguishing Legius syndrome from NF1 based on clinical signs is not an easy task, since LS is also characterized by the presence of CALMs and sometimes freckles, although in the absence of other NF1 criteria[8,12]. Distinctive NF1 manifestations, such as neurofibromas and Lisch nodules, usually do not appear until later in life[8], therefore, children suffering from NF1, or LS may have similar manifestations. In these cases, family history and parental observation may be extremely helpful in the diagnosis.

Besides RASopathies, another differential diagnosis to consider is Constitutive Mismatch Repair Deficiency (CMMRD). CMMRD is a rare autosomal recessive disease caused by homozygous or compound variants in *MLH1, MSH2, MSH6* or *PMS2*

(mismatch repair genes)[8,22]. It is a predisposition syndrome for childhood cancer with pigmentary features identical to those of NF1 and LS[8,22]. Since it is an autosomal recessive disorder, it should be suspected in a child of consanguineous parents, in the absence of other NF1 manifestations besides CALMs and in the presence of family history of Lynch syndrome[8,22].

The Piebald trait, an autosomal dominant disease, is caused by heterozygous variants in the proto-oncogene *KIT* or in *SNAI2,* a zinc finger transcription factor[8,22]. Clinical manifestations include white hair, areas with and without pigmentation, hyperpigmentation zones within or at the border of the depigmented areas, which can be mistaken for CALMs, and sometimes, can even include true CALMs and intertriginous freckles[8,22].

Neurofibromatosis type II (NF2) is an autosomal dominant disease that results from variants in *NF2*. Bilateral vestibular Schwannomas, schwannomas of other cranial and peripheral nerves, cutaneous schwannomas, meningiomas and ependymomas, and juvenile posterior subcapsular cataract are all typical features of NF2. Children can present CALMs, although they are usually fewer in number, smaller, paler and with more irregular borders than in NF1[8,22].

Schwannomatosis is a rare autosomal dominant condition, with patients predisposed to have multiple schwannomas, meningiomas and some individuals have at least one CALM and at most four. It can be caused by variants in *SMARCB1* or *LZTR1*[8,22].

Fibrous dysplasia/McCune-Albright syndrome (FD/MAS) results from an early embryonic postzygotic somatic activating variant of *GNAS*[8]. It is characterized by large CALMs, usually the first manifestation, with irregular boundaries, associated to polyostotic fibrous dysplasia[8,22].

Proteus syndrome results from somatic mosaic *AKT1* variant, with progressive overgrowth of multiple tissues specifically the skeleton, skin, adipose tissue, and central nervous system[8,22].

Infantile myofibromatosis is a rare autosomal dominant condition caused by variants in *PDGFRB*, and the main features are multiple tumors of the skin, subcutaneous tissues, skeletal muscle, bone, and viscera[8,22].

Chapter 3

Aims

Chapter 3. Aims

The overall aim of this internship was to improve the diagnosis and characterization of neurofibromatosis type I in Portuguese patients.

3.1. Specific aims

(i) Learn, become familiar with, and acquire technical skills in DNA and RNA extraction, cDNA reverse transcription, Sanger sequencing, NGS, and MLPA.

(ii) Acquire and develop skills in the classification of genetic variants according to the criteria recommended by the American College Medical Genetics/Association for Molecular Pathology (ACMG/AMP, 2015) and interpretation of their clinical significance.

(iii) Complete the process of implementing a protocol with high efficiency and increased sensitivity for the molecular study of the *NF1*.

(iv) Molecular diagnosis of suspected or clinically diagnosed NF1.

(v) Development of an NGS panel for RASopathies where *NF1* is included.

Chapter 4

Materials and Methods

Chapter 4. Material and methods

4.1. Patients

This study included a group of 40 patients of Portuguese origin from 34 unrelated families with suspected or clinically diagnosed NF1, referred by genetics and genetic counseling appointments. The mean age was 26 years with a gender distribution of 26 (65%) females and 14 (35%) males.

The clinical diagnosis of NF1 was based on the clinical criteria defined by the NIH Consensus Development Conference, in 1987 and revised in 2020. The study was approved by the Ethics Committee of the Faculty of Medicine of the University of Coimbra. All patients signed an informed written consent.

Two of the 34 probands had a previous *NF1* gDNA negative result, and 1 had a positive result in a previous gDNA study and was included as a control for the cDNA study.

Only 2 of the 6 relatives included in this study had clinical suspicion of NF1 (Table 1).

Table 1: Clinical data of probands and relatives with suspected or clinically diagnosed NF1.

Patient ID	Gender	Age	Clinical Information	Family History	Clinical Features
P1[a]	Male	14 y	NF1	No	Multiple CALMs; Neurofibromas; Scoliosis; Facial dysmorphia
P2	Male	2 y	NF1 suspect	No	Multiple CALMs; Blue nevus; Scalp xanthomas
P3	Female	5 y	NF1 suspect	No	Multiple CALMs; Inguinal freckles; Hypopigmented spot; Macrocephaly
P4	Female	33 y	NF1 suspect	Yes	Multiple CALMs; Lisch nodules
F4.1	Female	27 y	NF1 suspect	Yes	Multiple CALMs
P5	Male	12 y	NF1 suspect	No	Multiple CALMs; Axillary freckles; Pectus excavatum
P6	Female	9 months	NF1 suspect	No	Multiple CALMs; Tibial pseudoarthrosis
P7	Male	17 y	NF1 suspect	No	Multiple CALMs; Inguinal freckles; Macrocephaly; Lisch nodules; Hypochromic spot; Subcutaneous neurofibroma
P8	Female	28 y	NF1 suspect	Yes	Multiple CALMs; Axillary and inguinal freckles; Cutaneous and subcutaneous neurofibromas; Short stature; Learning disabilities
P9	Female	8 y	NF1 suspect	No	Multiple CALMs; Axillary freckles; Lisch nodules
P10	Female	27 y	NF1 suspect	Yes	Multiple CALMs; Multiple cutaneous and subcutaneous neurofibromas; Lisch nodules; Glioma
P11	Female	5 y	NF1 suspect	Yes	Multiple CALMs; Plexiform neurofibromas
F11.1	Male	37 y	NF1	Yes	Multiple CALMs; Multiple cutaneous neurofibromas

CALMs - café-au-lait macules; UBOs - unidentified bright objects; P - Proband; F - Relative; Y – Years; NA – Not Available

*With previous gDNA study, only cDNA was performed; [a] NF1 previously study performed

Table 1 (Continued)

P12	Male	6 y	NF1 suspect	No	Multiple CALMs with > 0,5 cm of diameter; Axillary and inguinal freckles; Xanthogranulomas
P13	Female	58 y	NF1 suspect	No	Multiple CALMs; Axillary freckles; Multiple cutaneous and subcutaneous neurofibromas; Lisch nodules
P14	Female	49 y	NF1 suspect	Yes	Multiple CALMs; Axillary and inguinal freckles; Lisch nodules; Neurofibromas; Pheochromocytoma
P15	Male	30 y	NF1 suspect	No	Multiple CALMs; Multiple cutaneous and subcutaneous neurofibromas; Plexiform neurofibroma
P16	Female	6 months	NF1 suspect	No	Multiple CALMs
P17	Male	48 y	NF1 suspect	Yes	Multiple CALMs; Axillary and inguinal freckles; Neurofibromas
P18*	Male	38 y	NF1 suspect	No	Multiple CALMs; Axillary freckles; Cutaneous and subcutaneous neurofibromas; Lisch nodules; Abdominal paraganglioma
P19	Male	44y	NF1 suspect	Yes	Multiple CALMs; Axilliary freckles; Neurofibromas; Lisch nodules; Mild cognitive impairment; Kyphoscoliosis
P20	Female	1y	NF1 suspect	No	6 CALMs; Freckles
P21	Female	24 y	NF1 suspect	No	Multiple CALMs; Lisch nodules
P22	Female	45 y	NF1 suspect	Yes	Multiple CALMs; Axillary freckles; Lisch nodules
P23	Female	22 y	NF1 suspect	No	6 CALMs with > 1,5 cm of diameter; Axillary freckles
P24	Female	22 y	NF1 suspect	Yes	Multiple CALMs; Axillary and inguinal freckles
P25	Female	57 y	NF1 suspect	No	Multiple CALMs; Cutaneous freckles; Cutaneous lipoma; Facial dysmorphisms; Hand anomalies
P26	Female	25 y	NF1 suspect	No	6 CALMs with > 1,5 cm of diameter; 1 Lisch nodule
P27	Female	40 y	NF1 suspect	Yes	Multiple CALMs; Freckles; Neurofibromas; Divergent strabismus
P28	Female	22 y	NF1 suspect	No	Multiple CALMs; Neurofibromas
P29	Female	8 y	Segmental NF1 suspect (mosaicism?)	No	4/5 CALMs; 1 large CALM ; Nevus achromicus; Nevus anemicus
P30	Female	35 y	NF1 suspect	No	CALMs right cervical, right thigh and calf region; Axillary freckles; Lisch nodules; Optic glioma; Cerebral glioma; Multiple cerebral hamartomatous lesions; Epilepsy
P31	Male	50 y	NF1 suspect	Yes	Freckles; Neurofibromas
P32	Female	47 y	NF1 suspect	No	Multiple CALMs and freckles left cervical, hemiface, shoulder and part of upper trunk on left; Left axillary freckles; Right breast cancer
P33	Female	8 y	NF1 suspect	No	Multiple CALMs; Freckles; Pectus excavatum; Scoliosis
P34*	Male	2 y	NF1 suspect	No	2 CALMs; Congenital tibial dysplasia

CALMs - café-au-lait macules; UBOs - unidentified bright objects; P - Proband; F - Relative; Y – Years; NA – Not Available

*With previous gDNA study, only cDNA was performed; [a] NF1 study previously performed

Nearly all patients, 35/36 (97%), presented CALMs, 21/36 patients (58%) had axillary and/or inguinal freckles, at least one type of neurofibroma was observed in 15/36 patients (42%) patients and in 12/36 patients (33%) Lisch nodules were identified (Table 2). Scoliosis, macrocephaly, tibial pseudoarthrosis, learning disabilities, facial dysmorphism, and hand anomalies, other less common clinical manifestations, were also found in at least one patient (Table 2).

Table 2: Clinical features in NF1 patients.

Features	N=36
CALMs	35
Freckling	21
Lisch Nodules	12
Neurofibromas (cNFs, sNFs, pNFs)	15
Optic Glioma	1
Skeletal Anomalies	7
Scoliosis	3
Tibial Dysplasia	1
Tibial Pseudarthrosis	1
Pectus Excavatum	2
Short starue	1
Macrocephaly	2

4.2. Strategy for *NF1* molecular study

The strategy developed at UFHM, is a two-step approach. It starts with regular analysis of the *NF1* coding region by gDNA sequencing using NGS. In case of a negative result, the evaluation of deletions/duplications in *NF1* should be performed by MLPA assay. If the gDNA result is negative, cDNA analysis is performed (Figure 4).

Finally, as long as the initial gDNA study does not include the differential diagnosis of NF1, patients with negative results for *NF1* (gDNA and cDNA) should proceed to the study of RASopathies (Figure 4).

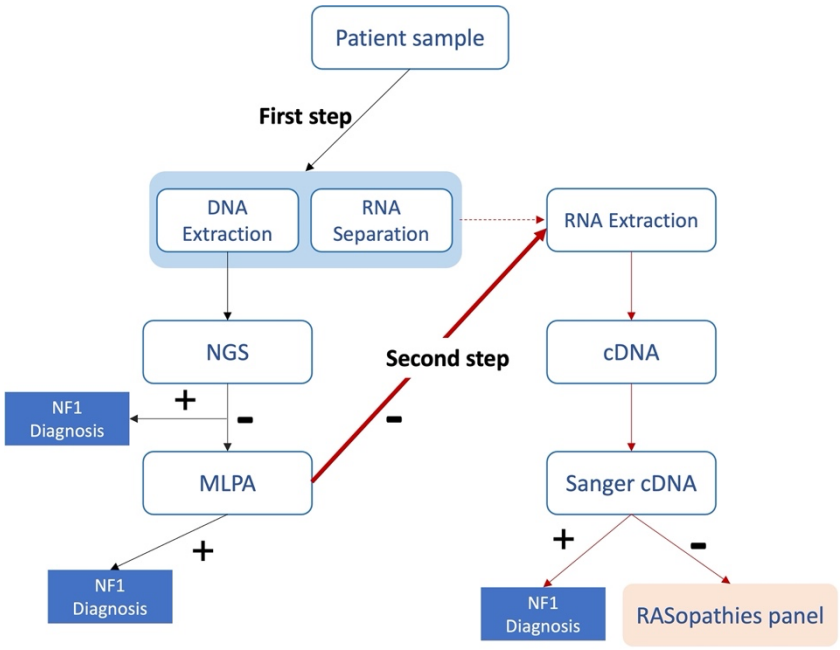

Figure 4: Schematic representation of *NF1* study in UFHM. The black arrows represent the first and the red arrows the second strategy steps.

4.3. DNA and RNA Extraction

Genomic DNA was extracted from peripheral blood lymphocytes by automated isolation with the QIAsymphonySP instrument using the QIAsymphony DSP DNA mini kit (Qiagen, Hilden, Germany).

Total RNA was isolated from peripheral blood lymphocytes and stored in RLT buffer at -80ºC. RNA was extracted using QIAsymphony RNA Kit and QIAsymphony SP (Qiagen, Hilden, Germany) and reverse transcribed using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems™ by Thermo Fisher Scientific, Waltham, MA, USA).

4.4. Next Generation Sequencing

The first step was the analysis of *NF1* coding region through gDNA sequencing by NGS. *NF1* (57 coding exons, NM_000267.3) is included in a gene panel designed using the AmpliSeq Designer software (Thermo Fisher Scientific, Waltham, MA, USA). The *NF1* target region (ROI) included the coding exons, their intron boundaries, and the 5' and 3' untranslated regions (UTRs). The 88 amplicons were divided into two primer pools. DNA was quantified in Qubit® 2.0 Fluorometer (Invitrogen by Thermo Fisher Scientific,

Waltham, MA, USA) using Qubit® dsDNA HS Assay Kits (Invitrogen by Thermo Fisher Scientific, Waltham, MA, USA). The NGS library was prepared using the Ion AmpliSeq™ Library Kit Plus (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's instructions and purified using Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA). The amplified libraries were quantified using the Ion Library TaqMan™ Quantitation Kit (Thermo Fisher Scientific, Waltham, MA, USA) in QuantStudio (Applied Biosystems™ by Thermo Fisher Scientific, Waltham, MA, USA), diluted to a concentration of 30 pM, and then mixed to create a combined library (Figure 5).

Emulsion PCR and chip loading were performed in the ION CHEF™ System (Thermo Fisher Scientific, Waltham, MA, USA), and then the sequencing reaction was performed in the Ion GeneStudio™ S5 System (Thermo Fisher Scientific, Waltham, MA, USA), all using the Ion 510™ & Ion 520™ & Ion 530™ Kit - Chef (Thermo Fisher Scientific, Waltham, MA, USA), according to the manufacturer's instructions (Figure 5).

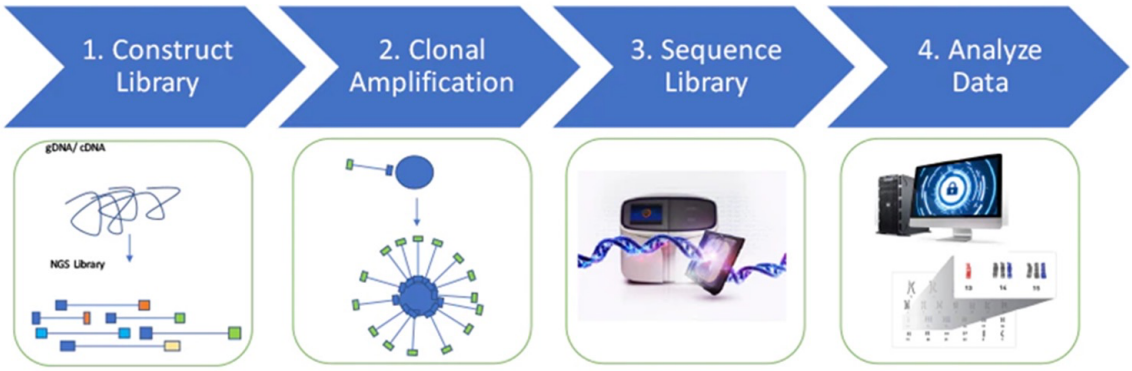

Figure 5: Schematic representation of NGS steps. Source: Thermo Fisher Scientific.

4.4.1. Bioinformatic Analysis

The NGS pipeline output, after Quality Control/Quality Assurance (QC/QA) filtering, was analyzed using Torrent Suite™ Software v.5.18, and sequences were aligned to the human genome version 19 (hg19). Variants were called using Torrent VariantCaller 5.18, and Ion Reporter™ Software v.5.20. The Coverage Analysis plugin v5.18 was run in Torrent Suite™ software before data were transferred to and analyzed

in Ion Reporter™ software. These metrics report the molecular coverage of ROI of the reference genome.

The minimum coverage depth required for each nucleotide in the ROI to be identified/considered a candidate pathogenic variant was >=100x. However, the average coverage for all ROIs was at least 500-fold coverage for each base in all patients.

4.5. Multiplex Ligation-dependent Probe Amplification

DNA samples from patients with no pathogenic variants identified by NGS were tested using the SALSA® MLPA® Probemix P081-D1 NF1 mix 1 & P082-C2 NF1 mix 2, version D1/C2-05, following the manufacturer's protocol (MRC Holland, Amsterdam, The Netherlands) to detect deletions/duplications in *NF1* (Figure 6). The reference sequence used in this kit is NM_001042492.3 and represents the longest transcript of NF1, including the in-frame coding exon 30alt31. Data normalization was performed using four healthy control samples. Analysis of PCR products was performed on an ABI 3130 Genetic Analyzer (Applied Biosystems™ by Thermo Fisher Scientific, Waltham, MA, USA) using GeneScan™ 500 LIZ™ (Applied Biosystems™ by Thermo Fisher Scientific, Waltham, MA, USA). Data analysis was performed using Coffalyser.Net software, version 220513.1739 (MRC Holland, Amsterdam, The Netherlands).

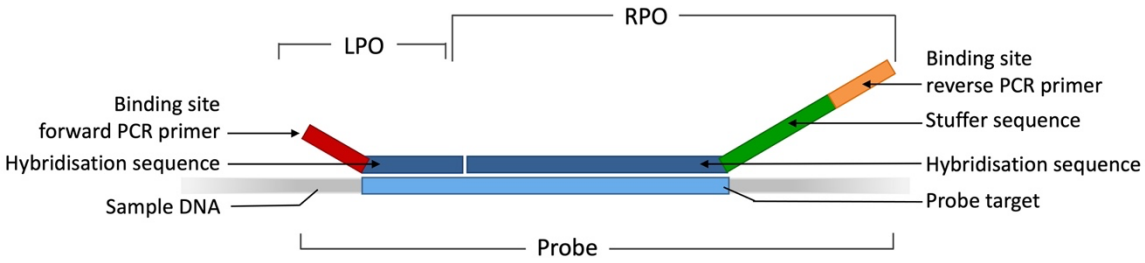

Figure 6: Schematic representation of MLPA technique. Source: MRC Holland.
LPO - left probe oligonucleotide; RPO – right probe oligonucleotide

4.6. Sanger Sequencing

Sanger sequencing was used in 4 different situations: cDNA analysis, family studies, deep intronic variant search and the *SPRED1* study.

4.6.1. cDNA Sequencing

As described by Audrey Sabbagh et al., the cDNA was amplified in eight overlapping fragments using conventional PCR[16]. The 8 fragments were divided as follows: fragment 1 - exon 1 to 9; fragment 2 - exon 8 to 17; fragment 3 - exon 12 to 22; fragment 4 - exon 21 to 33; fragment 5 - exon 32 to 37; fragment 6 - exon 36 to 43; fragment 7 - exon 42 to 52; and fragment 8 - exon 48 to 57 (Figure 7). For PCR amplification specific primers were used, some previously described in literature and others designed in UFHM. The primer sequences and polymerase chain reaction (PCR) conditions used in this study are detailed in Supplementary Material (Table S1, Table S4 and Table S5).

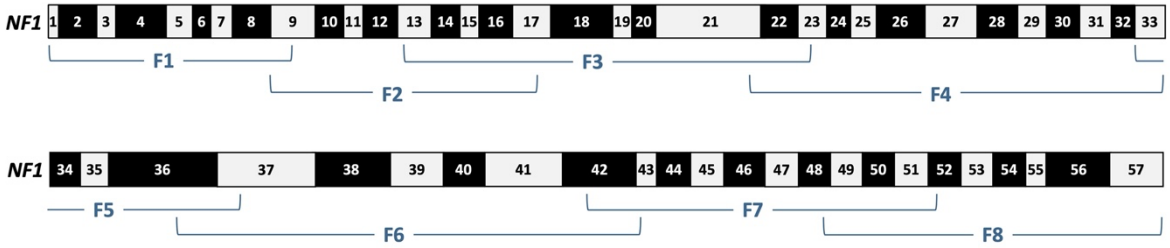

Figure 7: Representation of the 8 fragments (F1 to F8) used for the sequencing of cDNA of *NF1*.

To confirm PCR amplification, PCR products were run on agarose gel electrophoresis (Figure 8), using a 2% agarose gel with Tris-Borate-EDTA Buffer (TBE) and stained with SYBR™ Safe (Thermo Fisher Scientific, Waltham, MA, USA).

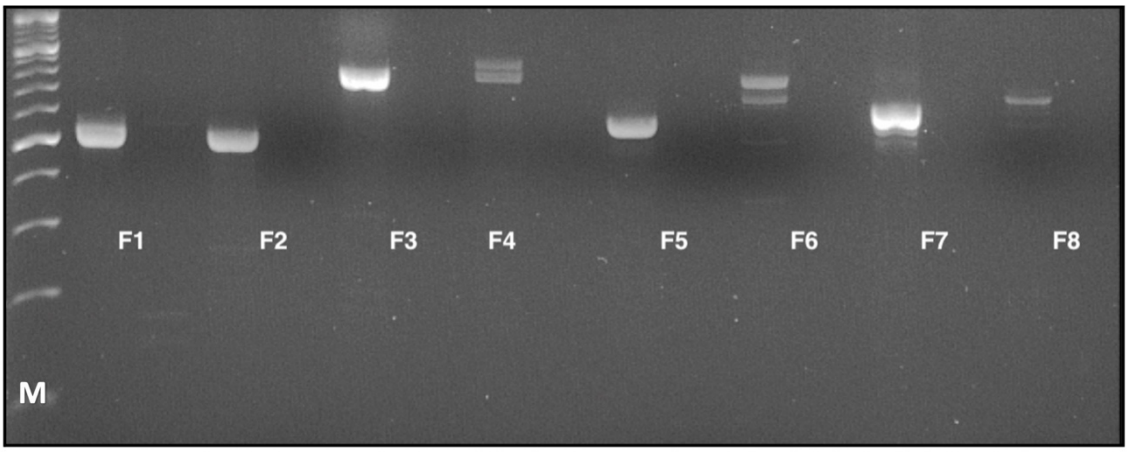

Figure 8: Quality control, by an agarose gel electrophoresis, of PCR amplification of the 8 fragments (F1 to F8). M – 200 bp ladder-size standard; F - Fragment

Sequencing of the PCR products was performed in three steps: ExoSAP-IT™Express (Applied Biosystems™ by Thermo Fisher Scientific, Waltham, MA, USA) purification, sequencing reaction, and resin column purification. Sequencing was carried out using the ABI BigDye Terminator Sequencing Kit v1.1 (Applied Biosystems™ by Thermo Fisher Scientific, Waltham, MA, USA) and specific primers (Table S1). Sequenced PCR products were run on the SeqStudio™ Genetic Analyzer (Applied Biosystems™ by Thermo Fisher Scientific, Waltham, MA, USA). Sequence analysis was performed using Seqscape analysis software v.4.0 (Applied Biosystems™ by Thermo Fisher Scientific, Waltham, MA, USA) with the cDNA reference sequence NM_000267.3.

4.6.2. Family Studies and deep intronic variant search

In family studies, the search for the family variant was carried out by gDNA Sanger sequencing as in cases where the search for deep intronic variants was necessary.

The procedure was the same as previously described. PCR conditions and specific primers used are detailed in Supplementary Material (Table S2 and Table S6).

4.7. *SPRED1* Study

Due to the clinical overlap between NF1 and Legius syndrome (OMIM #611431)[7,12], the molecular study of sprouty related EVH1 domain containing 1 gene (OMIM *609291; *SPRED1*) was performed in one patient with a negative result in the *NF1* study. As a first step, the coding region of *SPRED1* was investigated by Sanger sequencing. The PCR conditions and primers used are detailed in Supplementary Material (Table S3 and Table S7). In the absence of a positive result, MLPA was performed using SALSA® MLPA® Probemix P295 SPRED1, version B3 (MRC Holland, Amsterdam, The Netherlands) according to the manufacturer's instructions. Analysis was performed as described for *NF1*.

4.8. Whole Exome Sequencing - Virtual RASopathies Panel

Since distinguishing NF1 from other RASopathies can be challenging and there are some overlapping features, the ideal study should start with a RASopathies panel that includes *NF1*. For this reason, a virtual panel of 23 RASopathies-associated genes (Table 3), all belonging to the RAS/MAPK pathway, was developed through literature review. This panel was applied to the large amounts of data generated by whole exome sequencing.

Table 3: RASopathies Panel.

RASopathies Panel			
BRAF NM_004333.6	*CBL* NM_005188.4	*HRAS* NM_176795	*KRAS* NM_033360
LZTR1 NM_006767.4	*MAP2K1* NM_001411065	*MAP2K2* NM_030662.4	*MAPK1* NM_002745.5
MRAS NM_001085049.3	*NF1* NM_000267.3	*NRAS* NM_002524.5	*PPP1CB* NM_002709.3
PTPN11 NM_001330437	*RAF1* NM_002880.4	*RASA1* NM_002890.3	*RASA2* NM_006506.5
RIT1 NM_006912.6	*RRAS2* NM_012250.6	*SHOC2* NM_007373.4	*SOS1* NM_005633.4
SOS2 NM_001411020	*SPRED1* NM_152594.3	*SPRED2* NM_001128210	

Whole Exome Sequencing (WES) was performed with Twist Comprehensive Exome Panel.

DNA was quantified in Qubit® 2.0 Fluorometer (Invitrogen by Thermo Fisher Scientific, Waltham, MA, USA), using Qubit® dsDNA HS Assay kits (Invitrogen by Thermo Fisher Scientific, Waltham, MA, USA). Genomic DNA libraries, starting with the enzymatic fragmentation of DNA, were prepared using Twist Library Preparation Kit (Twist Bioscience, South San Francisco, CA, USA), according to the manufacturer's instructions. A quality control (QC) analysis was carried out to quantify and evaluate the size of each library. Libraries were quantified in Qubit® 2.0 Fluorometer, using Qubit® dsDNA HS Assay kits, and concentration values should be ≥80 ng/µl. The fragment size was evaluated in 4150 TapeStation (Agilent Technologies, Santa Clara, CA, USA) and the average fragment length should be 420-470 bp. All the steps of library preparation were performed according to the manufacturer's protocol.

The next step is to generate enriched DNA libraries using Twist Target Enrichment Kit (Twist Bioscience, South San Francisco, CA, USA), according to the manufacturer's instructions. The enriched libraries were purified using DNA purification beads and it was performed a QC analysis to quantify and to evaluate the size of each enriched library. The fragments size was evaluated as described above. The final concentration was evaluated in Qubit® 2.0 Fluorometer, with Qubit® dsDNA HS Assay kits, and concentration values should be 7-15 ng/µl. The sequencing reaction of enriched libraries was realized on an Illumina platform.

Data analysis was performed by SOPHiA DDM™ platform, and the Virtual RASopathies Panel applied.

4.9. Variants Classification

All the variants were analyzed using the bioinformatics tools Varsome and Franklin, which include in silico algorithms to predict the effect of variants and provide ACMG classification of genetic variants. SpliceAI software, an in silico splicing prediction tool, was also used to study splicing variants.

Databases such as dbSNP, Genome Aggregation Database (gnomAD), ClinVar, Leiden Open Variation Database (LOVD) and HGMD® were consulted to check whether variants have been previously described and associated with disease.

The splicing variants identified in this study were divided into five categories according to the *NF1* splicing variants system classification, proposed by Wimmer et al[40].

All genetic variants were classified according to the ACMG/AMP guidelines[41,42].

Chapter 5

Results

Chapter 5. Results

5.1. Identification of variants in NF1 patients

The present study included a cohort of 40 individuals from 34 unrelated families with suspected or clinically diagnosed NF1. Family studies were performed on six individuals, four healthy and two with clinical disease. As expected, no healthy individuals had the familial variant, and the other two had it.

The parental study of P6 and P12 allowed the conclusion that the variant found in both patients was *de novo*. In P11, the family study concluded that the variant was inherited from the father, and in P4 the variant was also detected in her sibling, but the parental study was not possible.

According to the two-step strategy, in the first approach, 31 of which started with *NF1* analysis by NGS that revealed 16 different variants: three small deletions, two small duplications, one splice site variant, seven nonsense, and three missense changes (Figure 9 and Figure 10). In one patient, a large deletion was found by MLPA (Figure 9 and Figure 10).

In three cases (P1, P18, and P34), the gDNA had been previously studied, and in this study only cDNA analysis was performed. The cDNA Sanger sequencing allowed the identification of one deep intronic variant, which was subsequently confirmed by intron sequencing, and the confirmation of the splice variant previously identified in P1 (Figure 9 and Figure 10).

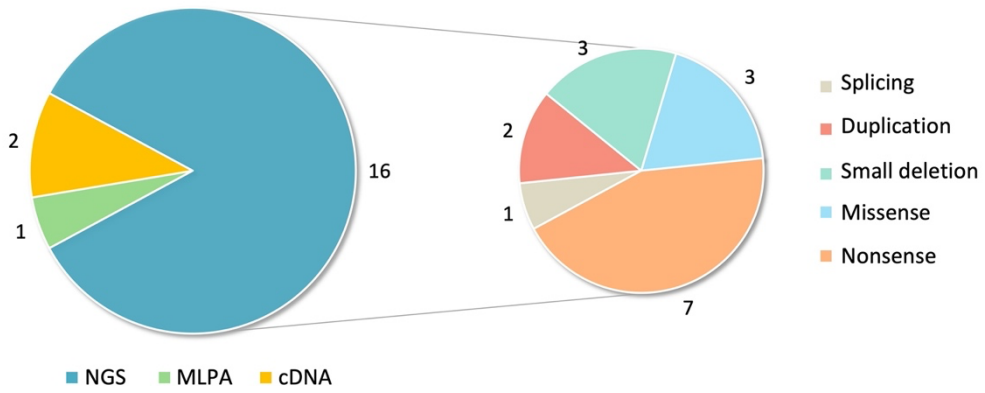

Figure 9: Type of variants identified in NF1 patients.

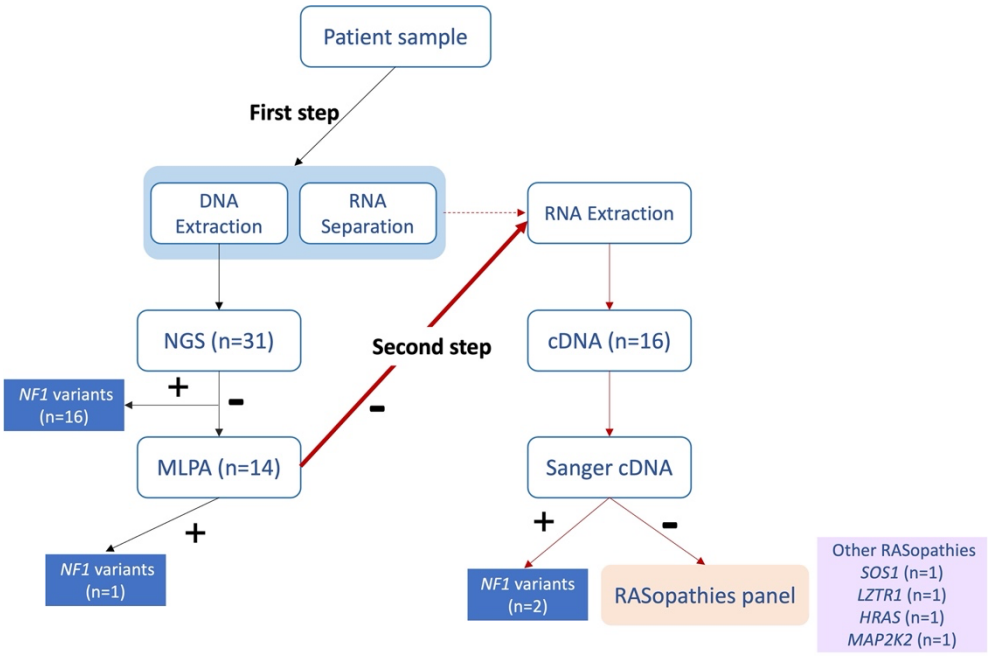

Figure 10: Schematic representation of *NF1* study and the number of studied patients by each method. The black arrows represent the first and the red arrows the second strategy steps.

Nineteen different pathogenic (18) and likely pathogenic (1) variants were identified in *NF1* in 20/34 unrelated patients (Table 4). In the remaining 14 probands, no variant was detected in *NF1*. Among the 19 variants, four (21%) had never been reported in the population and in international disease databases (Table 4).

The results showed a considerable spectrum of *NF1* variants, distributed between exons 5 and 53, including intron 12, 15, and 31 of *NF1* (Figure 11). These outcomes and the fact that only one recurrent variant was detected in two probands confirm that there are no hotspots for variants in this gene. Fourteen of the variants identified, in this cohort, affect functional domains of neurofibromin: four variants in CSRD, four in GRD, one involving PH and LRD, one in LRD and HLR, and three in CTD, one of which also involves HLR. The large deletion involves GRD, SBR, Sec, PH, LRD, and HLR (Figure 11).

Table 4: *NF1* variants detected in our cohort.

Patient ID	Gender	Location	Variant	mRNA Effect	Amino Acid Change	Reported	ACMG Classification	ACMG Criteria
P1	Male	IVS 31	c.4269+1G>A	r.4111_4269del	p.(Val1371_Lys1423)	Yes - [40]	Pat	PVS1, PM2, PP5
P2	Male	E 28	c.3850del	r.3850del	p.(Ile1284*)	Yes - [49]	Pat	PVS1, PM2, PP5
P3	Female	E 28	c.3827G>A	r.3827g>a	p.(Arg1276Gln)	Yes - [50]	Pat	PM1, PP2, PM2, PP3, PP5, PM5
P4	Female	E 45	c.6792C>A	r.6757_6858del; r.6792c>a	p.(Ala2253_Lys2286del); p.(Tyr2264*)	Yes - [51]	Pat	PVS1, PM2, PP5
F4.1	Female	E 45	c.6792C>A	r.6757_6858del; r.6757c>a	p.(Ala2253_Lys2286del); p.(Tyr2264*)	Yes - [51]	Pat	PVS1, PM2, PP5
P5	**Male**	**E 38**	**c.5692dup**	**r.5692dup**	**p.(Glu1898Glyfs*17)**	***Novel***	**LPat**	**PVS1, PM2**
P6	Female	E 50	c.7486C>T	r.7486c>u	p.(Arg2496*)	Yes - [52]	Pat	PVS1, PM2, PP5
P7	Male	E 13	c.1466A>G	r.1466_1527del; r.1466a>g	p.(Tyr489*); p.(Tyr489Cys)	Yes - [53]	Pat	PM1, PP2, PM2, PP5
P8	Female	IVS 15	c.1722-2A>G	r.1721_1722ins1722-43_1722-1	p.(Ser574fs)	Yes - [40]	Pat	PVS1, PM2, PP5
P9	Female	E 28	c.3826C>T	r.3826c>u	p.(Arg1276*)	Yes - [54]	Pat	PVS1, PM2, PP5
P10	Female	E 14	c.1541_1542del	r.1541_1542del	p.(Gln514Argfs*43)	Yes - [55]	Pat	PVS1, PM2, PS4, PP5
P11	**Female**	**E 21**	**c.2614G>T**	**r.2410_2638del; r.2614g>u**	**p.(Ala804*); p.(Glu872*)**	***Novel***	**Pat**	**PVS1, PM2, PP1, PP4, PP5**
F11.1	**Male**	**E 21**	**c.2614G>T**	**r.2410_2638del; r.2614g>u**	**p.(Ala804*); p.(Glu872*)**	***Novel***	**Pat**	**PVS1, PM2, PP1, PP4, PP5**
P12	Male	E 21	c.2540T>C	r.2540u>c	p.(Leu847Pro)	Yes - [50]	Pat	PM1, PP2, PM2
P13	Female	E 5	c.496_497del	r.496_497del	p.(Val166Leufs*7)	Yes - [56]	Pat	PVS1, PM2, PP5
P14	**Female**	**E 30-41**	**c.(3974+10_3982)_(6250_6387)del**	**r.3975_6364del**	**p.(Leu1326Argfs*25)**	***Novel***	**Pat**	**PVS1, PM2, PP4**
P15	Male	E 53	c.7846C>T	r.7846c>u	p.(Arg2616*)	Yes - [57]	Pat	PVS1, PM2, PP5
P16	Female	E 37	c.5242C>T	r.5242c>u	p.(Arg1748*)	Yes - [58]	Pat	PVS1, PM2, PP5
P17	Male	E 12	c.1381C>T	r.1381c>u	p.(Arg461*)	Yes - [50]	Pat	PVS1, PM2, PP5
P18	**Male**	**IVS 12**	**c.1392+751T>G**	**r.1392_1393ins1392+752_1392+822**	**p.(Pro464_Ser465insTyrTrpLeuLeuSerLeuThr*)**	***Novel***	**Pat**	**PVS1, PM2,** ~~PP4~~
P19	Male	E 21	c.2619dup	NA	p.(Lys874*)	Yes - [59]	Pat	PVS1, PM2, PP5
P20	Female	E 21	c.2540T>C	r.2540u>c	p.(Leu847Pro)	Yes - [50]	Pat	PM1, PP2, PM2

E – Exon; F – Familiar; IVS – InterVening Sequence; LPat – Likely Pathogenic; NA - Not Available; P – Proband; Pat - Pathogenic; The bold lines show the *novel* variants.

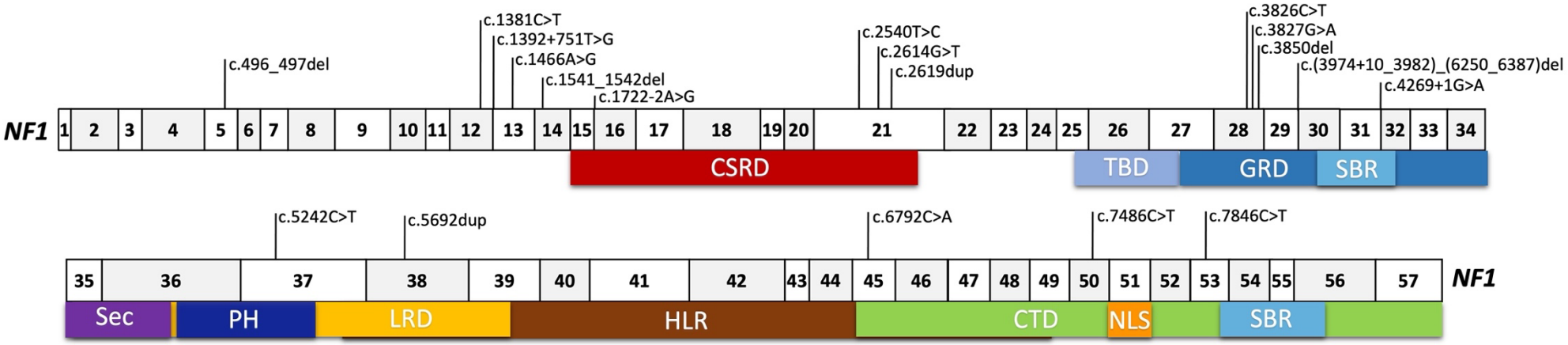

Figure 11: Schematic overview of the 19 *NF1* variants location identified in this study. Protein functional domains are indicated under the corresponding exons. CSRD - cysteine-serine-rich domain; TBD - tubulin-binding domain; GRD - central GAP-related domain; SBR - syndecan-binding region; Sec - Sec14 homologous domain; PH - pleckstrin homologous domain; LRD - leucine-rich domain; HLR - heat-like repeats; CTD - C-terminal; NLS - nuclear localization signal. Variant nomenclature is according to reference transcript NM_000267.3.

5.1.1. Sequencing of cDNA and detection of splicing variants

The P1, with a splicing variant c.4269+1G>A was included as a control for the cDNA analysis. Sanger sequencing of the cDNA revealed skipping of exon 31 (Figure 12). The fragment 4 used for the sequencing of cDNA of *NF1* (exon 21 to 33) was amplified and sequenced since as it was expected that the result, in cDNA, would be found in this region (Figure 12 - A). Three different sequences were found: the normal transcript (Wt); the normal transcript with an alternatively spliced exon, 63 bp inserted between exons 30 and 31 (30alt31, Wt*); and the mutated transcript (Mt). To obtain a cleaner sequence and to confirm the deletion of exon 31, two more primers were designed: an alternative exon-specific primer, NF1-F4-E31-202, and a deletion-specific primer, NF1-F4-E30.32-201 (Table S1). Thus, with the first primer (Figure 12 - B), two sequences were found, the normal one with exon 30alt31 and the mutated one. With the second primer, only the transcript with the deletion was amplified, as shown in Figure 12-C. With these results, it has been possible to validate our approach of designing primers to obtain the complete *NF1* cDNA and detect all possible deletions/insertions.

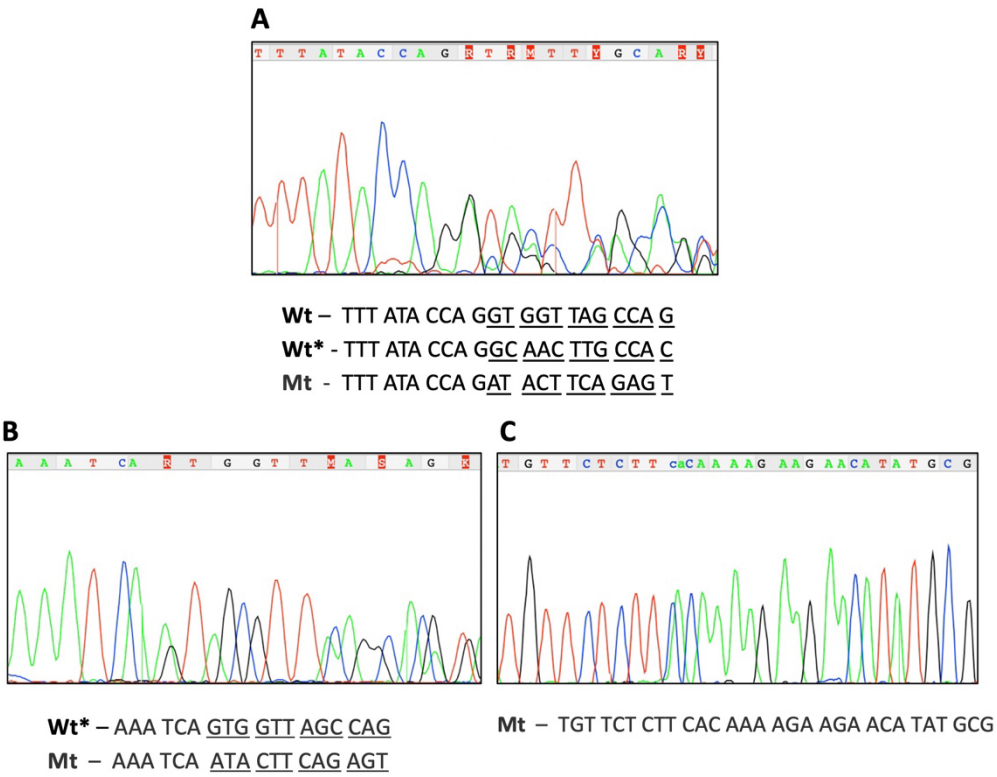

Figure 12: Effect on mRNA of *NF1* c.4269+1G>A pathogenic variant in patient P1. (A) Detection of wild type (Wt), wild type with exon 30alt31 (Wt*) and mutant type (Mt), deletion of exon 31. (B) Detection of wild type with exon 30alt31 (Wt*) and mutant type (Mt), deletion of exon 31. (C) Detection of mutant type.

5.1.2. Profiling of novel mutations and potential impact on cDNA

Four *novel* variants were detected in four patients (P5, P11, P14 and P18): two by NGS (c.5692dup and c.2614G>T), one by MLPA (large deletion), and one by cDNA sequencing (deep intronic).

Two *novel* exonic variants, c.5692dup and c.2614G>T, were found in P5 and P11, respectively. In both cases, the effect on mRNA transcript was also evaluated with cDNA sequencing. In c.5692dup apart from the expected duplication, no other effects have been detected (Figure 13). On the other hand, in c.2614G>T, cDNA sequencing revealed the deletion of the first 229 bp of exon 21 (Figure 14).

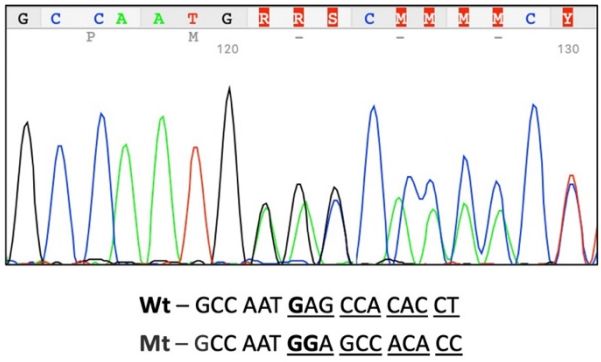

Figure 13: Effect on mRNA of *NF1* c.5692dup likely pathogenic variant in patient P5. Detection of wild type (Wt) and mutant type (Mt).

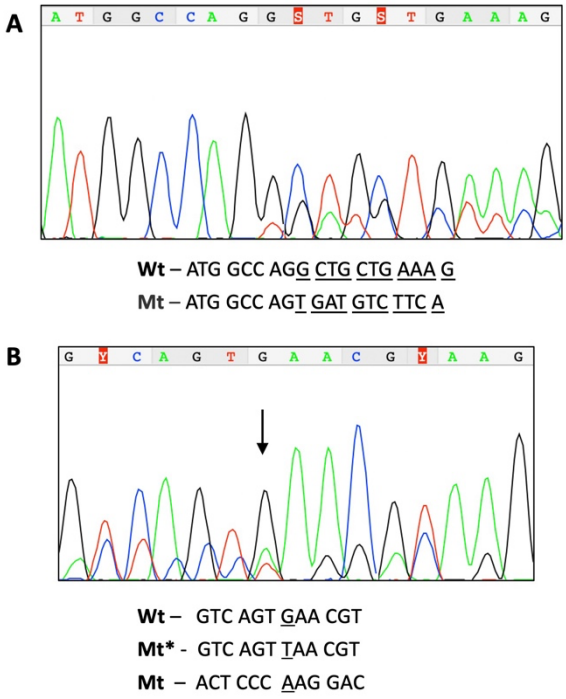

Figure 14: Effect on mRNA of *NF1* c.2614G>T pathogenic variant in patient P11. (A) Detection of wild type (Wt) and mutant type with aberrant splicing (Mt). (B) Detection of wild type (Wt), mutant type with aberrant splicing (Mt) and mutant type with point mutation (Mt*).

In P14, MLPA allowed the identification of a *novel NF1* large deletion, c.(3974+10_3982)_(6250_6387)del (Figure 15 – A). The cDNA confirmed complete deletion of exons 30 to 41 deletion resulting in a premature stop codon 25 amino acids upstream (Figure 15 – B).

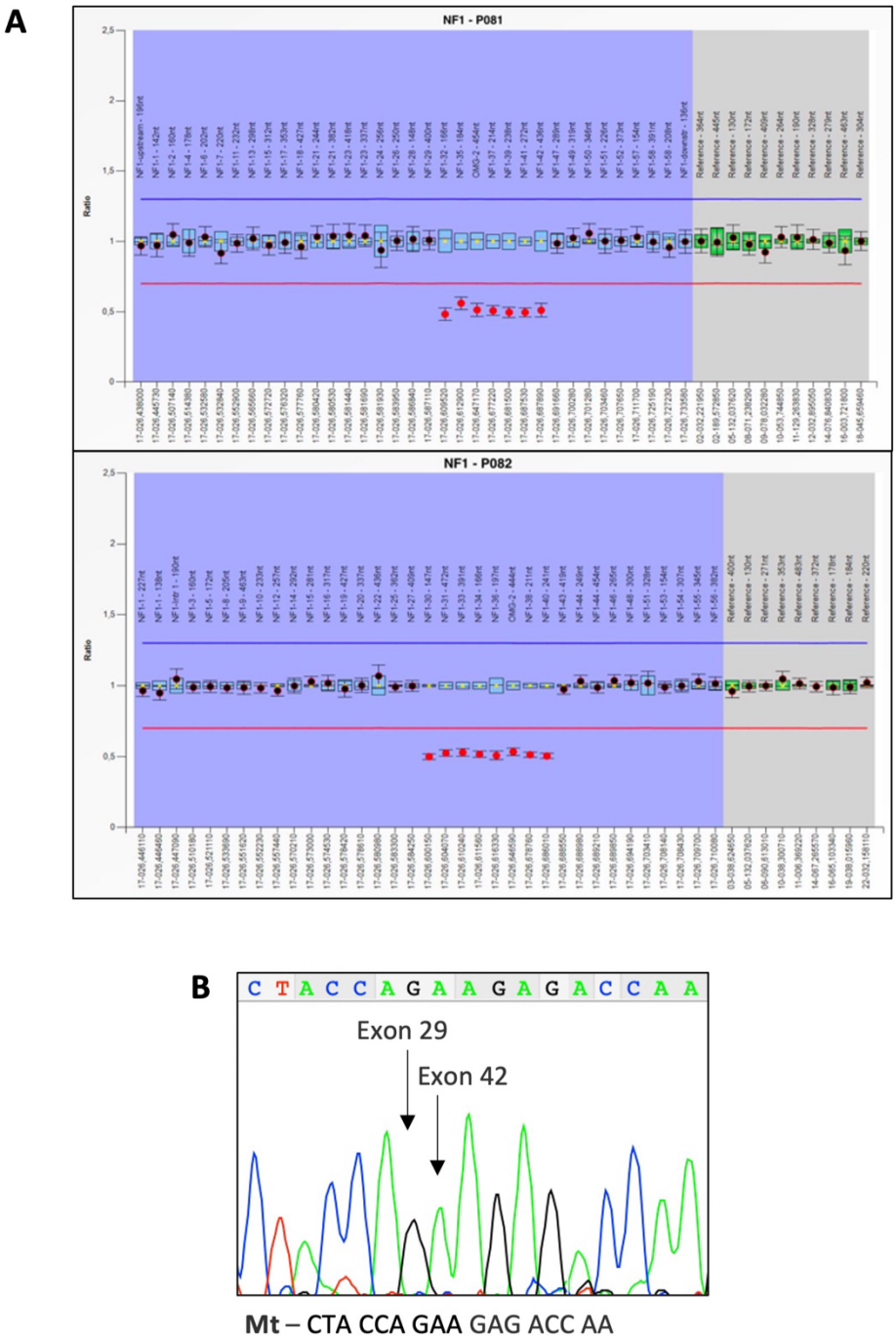

Figure 15: *NF1* pathogenic variant in patient P14. (A) Identification of a large deletion, c.(3974+10_3982)_(6250_6387)del (exon 30 to 41) by MLPA. (B) Exon 30 to 41 deletion in the mRNA, identified by Sanger sequencing with specific primers F4F and F6R (Table S1).
Mt – Mutant type.

In P18, with previous gDNA negative, we directly performed the cDNA study that revealed an insertion of 71 bp from intron 12 sequence between exons 12 and 13 (cryptic exon inclusion) (Figure 16 – A). In view of this result, the most plausible hypothesis would be a deep intronic variant in intron 12. Intron 12 was partially sequenced and the c.1392+751T>G variant was identified (Figure 16 – B). This variant leads to a premature stop codon, 8 amino acids ahead (p.Pro464_Ser465insTyrTrpLeuLeuSerLeuThr*).

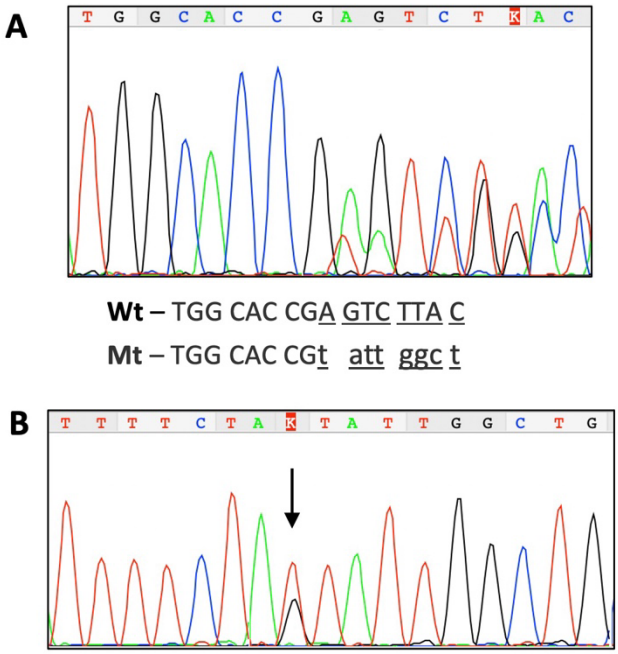

Figure 16: *NF1* pathogenic variant in patient P18. (A) mRNA insertion of 71 bp from intron 12, between exons 12 and 13. (B) Deep intronic variant in intron 12, c.1392+751T>G.
Wt – Wild type; Mt – Mutant type.

5.1.3. Classification of splicing variants

According to the classification of *NF1* splicing variants system proposed by Wimmer et al., the variants identified in this study were classified into five categories as follows:

Type I: splice site variants causing an exon skipping. In P1, it was observed the skipping of exon 31, as the result of the c.4269+1 variant (Figure 12 and Table 5).

Type II: deep intronic variants, which leads to a cryptic exon inclusion. This effect was identified as a result of the c.1392+751T>G variant, found in P18 (Figure 16 and Table 5).

Type III: exonic variants, that result in a skipping of part of the exon, due to a new splice site creation. The c.1466A>G, detected in P7, and c.2614G>T, detected in P11, leads to this aberrant splicing event, skipping the last 62 bp of the exon 13 and the first 229 bp of exon 21 (Figure 14), respectively (Table 5).

Type IV: variants that disrupt splice site, causing the use of cryptic exonic or intronic splice sites, leading to partial loss of an exon or partial inclusion of an intron, respectively. The splice variant c.1722-2A>G, found in P8, was classified as type IV (Table 5). It causes an insertion of 44 bp, from intron 15, between exons 15 and 16.

Type V: exonic variants leading to exon skipping. The variant c.6792C>A, identified in P4, gives rise to exon 45 skipping (Table 5).

Table 5: Classification of *NF1* variants according to *NF1* splicing variants system classification.

Location	Variant	mRNA Effect	Amino Acid Change	Classification of Splicing Variants
IVS12	c.1392+751T>G	r.1392_1393ins1392+752_1392+822	p.(Pro464_Ser465insTyrTrpLeuLeuSerLeuThr*)	Type II
E 13	c.1466A>G	r.1466_1527del; r.1466a>g	p.(Tyr489*); p.(Tyr489Cys)	Type III
IVS 15	c.1722-2A>G	r.1721_1722ins1722-43_1722-1	p.(Ser574fs)	Type IV
E 21	c.2614G>T	r.2410_2638del; r.2614g>u	p.(Ala804*); p.(Glu872*)	Type III
IVS 31	c.4269+1G>A	r.4111_4269del	p.(Val1371_Lys1423)	Type I
E 45	c.6792C>A	r.6757_6858del102; r.6792c>a	p.(Ala2253_Lys2286del); p.(Tyr2264*)	Type V

5.2. Genotype–Phenotype Correlations

In 22 patients (20 probands and 2 family members) positive for *NF1*, 91% (20) fulfilled NIH criteria at the time of the molecular diagnosis. In 9% (2/22) only one NIH criterion was present, probably due to their young age (Figure 17). These results were to be expected, since patients younger than 8 years of age usually do not show complete NF1 manifestations. In the remaining 14 probands who were negative for *NF1*, 79% (11/14) had 2 or more NIH criteria, while 21% (3/14) patients did not fulfil the NIH criteria (Figure 17).

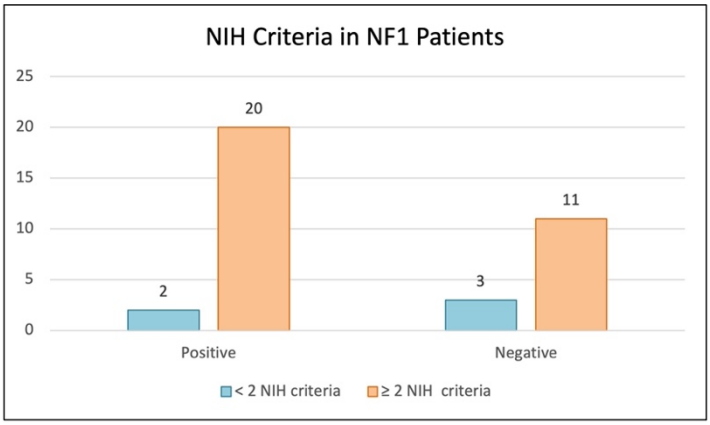

Figure 17: Patients' distribution, according to NIH minimal criteria for NF1 diagnosis.

P14, in whom the large NF1 deletion was identified, has a severe phenotype. Patients P12 and P20, with the p.(Leu847Pro) variant described in association with a severe phenotype, had only a mild phenotype, probably because they were only six and one years old, respectively.

P3 had the p.(Arg1276Gln) variant and did not have a cardiovascular malformation, the phenotype usually associated with this variant. The results of this study showed that the phenotypes of the patients were very heterogeneous and that in patients with similar phenotypes (e.g. cutaneous and subcutaneous neurofibromas) the variants found were different in type and/or localization. Furthermore, patients with variants of similar type or localization had different phenotypes. Family cases supported this, as several phenotypes can be observed for the same variant in patients from the same family.

5.3. Molecular Diagnosis of other RASopathies

Since P33 patient, who was negative result in the *NF1* study, had a phenotype that could be compatible with Legius syndrome, *SPRED1* sequencing was performed, and the result was also negative.

During the course of this study, six of the 14 patients (P21, P22, P23, P24, P25, P26) with a negative *NF1* result, could be investigated for differential diagnosis of RASopathies by WES applying a virtual RASopathies panel.

Among of the six patients studied by WES we found gene defects in four: in P25 and P26 two *novel* likely pathogenic variants were found in *SOS1*, c.1769A>G, and *LZTR1*, c.410C>A, respectively, genes associated with Noonan syndrome; in P23 and P24, were identified two gross deletions in *MAP2K2* (deletion of exons 9 to 11) and *HRAS* (complete gene deletion), respectively (Figure 10). These findings are still being characterized; in P21 and P22, no variants were found.

We can therefore conclude that, in this cohort of patients, the molecular diagnosis was carried out in 24/34 probands (70.6%) using our approach: 20/34 (58.8%) with NF1 and 4/34 (11.8%) with other RASopathies.

Chapter 6

Discussion

Chapter 6. Discussion

This study comprised a genotype–phenotype correlation in a cohort of 34 unrelated Portuguese families (40 individuals) with clinical suspicion of NF1. The NF1 is one of the most common Mendelian diseases and its molecular study is complex[7,24]. It shows great genetic and clinical variability, with a limited genotype-phenotype correlation. *NF1* is one of the largest genes in the human genome, with variants described throughout the gene (more than 4000 variants reported in HGMD® 2023.2, released June 2023), without hotspots, and with a high frequency of *de novo* variants. In addition, in *NF1*, the unusual splicing events are more frequent than in other genes[43], some of which, are not detected in the initial gDNA study. Not only the splice site variants are responsible for the splicing events, but also some exonic and deep intronic variants have been described as causative[24].

For this study, a protocol was followed that combined the gDNA sequencing by NGS, the search of CNVs by MLPA, and cDNA analysis by Sanger sequencing, to search for variants that disturb splicing and that are not detected with the previous methodologies.

Sixty-one percent of the patients were positive for *NF1* variants (22/36) and 39% (14/36) were negative. All patients with positive *NF1*, meet the minimum criteria for NF1 except P2 and P16. In both cases, multiple *café-au-lait* macules were present without any other criteria, but at the time of diagnosis, they were only two years and six months old, respectively. In NF1, the full clinical manifestation occurs usually after the age of eight years and is progressive over time[13]. These data are consistent with what was been reported previously, justifying the positive result in the absence of criteria at a young age[24,39].

When the clinical features were compared with the minimum NIH criteria for NF1 diagnosis[8,12], only 3/14 of the patients with negative results did not meet the minimum criteria (P25, P29, and P34). In P25, multiple *café-au-lait* macules were present without any other minimum criteria of NF1. In P29, with the suspicion of a segmental NF1, lacking the NF1 minimum criteria and the mosaic NF1 minimum criteria, the molecular study was performed in peripheral blood, which could justify the negative result, being

proposed to perform the study in one of the spots skin biopsies. The P34, with two CALMs and tibial dysplasia, was only 2 years old, which could justify the lack of the minimum criteria.

In the 20 probands with variants identified in *NF1* variants were detected by NGS in 17 (85%), by MLPA in one (5%), and by cDNA sequencing in the other two (10%). These results are in line with what was expected, with most of the *NF1* variants being detected by gDNA analysis, demonstrating the high sensitivity of the recently optimized methodology not only for molecular diagnosis but also for the identification of different types of variants[8]. Missense, nonsense, small deletions and duplications, splice site variants, intronic variants, and CNVs were identified. However, the diagnostic rate was only 61%, lower than the 95% expected with this workflow [8,36]. This can be explained not only by the small number of patients studied but also by the lack/non-exclusion of other differential diagnoses of NF1.

Four *novel* variants were detected, three by first-step approach (c.5692dup, c.2614G>T and a large deletion c.(3974+10_3982)_(6250_6387)del) and a deep intronic variant (c.1392+751T>G) by cDNA sequencing. This deep intronic variant is close to a reported variant[44], located in the same region (3 bp upstream). This was classified as a type II splice variant according to the *NF1* classification system of Wimmer et al[40].

Of the 15 already described variants, four have an effect on splicing. The c.4269+1G>A and c.1722-2A>G, are at the canonical sites and, as expected both disrupt the splicing event. The other two, were exonic variants c.1466A>G and c.6792C>A, located in the middle of exon 13 and 45, respectively, in addition to the expected nucleotide exchange in the mRNA, affect the mRNA splicing process in different ways. In this cohort, ~33% of variants affect the mRNA processing, two variants at splice site, three exonic variants, and one deep intronic variant. This rate is consistent with the 30% of previous reports[7,16,22]. The c.2619dup variant, identified in patient P19, was excluded from these statistics since it was not possible to evaluate the effect on mRNA due to the inability to obtain RNA from this sample. It would be of great interest to be able to obtain a viable sample to perform mRNA studies, as there is no publication reporting this.

There are exonic variants that act with a dual effect at the RNA and protein level. As reported previously more than 30% of missense variants disrupt pre-mRNA splicing,

showing a deleterious effect in a significant number of these variants[9,16]. In fact, in this study, 1/3 missense variants detected (c.1466A>G) causes an alteration of the splicing event. The cDNA sequencing also revealed that two nonsense variants out of seven (~29%) identified, disturb splicing (c.2614G>T and c.6792C>A), which is higher than the ~14% reported by Sabbagh et al.[16] According to the mRNA sequence, in c.6792C>A the transcript with abnormal splicing is transcribed in higher quantity compared to the transcript with the single-nucleotide variation (SNV).

This dual effect observed in these variants (c.1466A>G, c.2614G>T and c.6792C>A), which has also been reported by other groups[14,45,46], should be evaluated in order to try to understand its impact in phenotype. It would be interesting to recognize if it has a cumulative effect leading to a severe phenotype or if it can be a positive factor, causing a mild phenotype.

The genotype-phenotype correlation is very important since it allows one to understand how the disease might develop. Over the years, several groups have tried to evaluate this correlation in NF1, but with little success so far[13]. NF1 has a highly variable clinical expression even within members of the same family[31], and most of the time this correlation is scarce, although there are already some variants with an associated phenotype. Large deletions and missense variants affecting one of the five codons between 844 and 848 are related with a more severe phenotypes[8,28]. Indeed, P14 in whom the large deletion c.(3974+10_3982)_(6250_6387)del was identified has a more severe phenotype. However, P12 and P20, with the p.(Leu847Pro) variant, had only a mild phenotype, probably due to their young age.

Cardiovascular malformations and NF1-Noonan phenotype have been observed with higher frequency in patients with missense variants affecting p.Arg1276 and p.Lys1423[8,30]. In P3 a missense variant affecting p.Arg1276 (c.3827G>A) was identified, although the clinical features did not included cardiovascular malformations.

With the exception of one variant that was found in two different families, none of the other 18 variants were identified in more than one family. Therefore, it was only possible to attempt to establish some correlation based on variant type.

Several groups have reported a high prevalence of Lisch nodules in patients with frameshift[9] and nonsense variants[16]. In this cohort, Lisch nodules were observed in three

of the five patients with frameshift variants and in two of the eight patients with nonsense variants.

Neurofibromas (cNFs, sNFs, and pNFs), have been associated with different types of variants[16,47]. In fact, Kang et al. and Scala et al. described a negative association between the presence of cNFS, sNFs and pNFs and *NF1* missense variants[24,47]. In four patients with missense variants cNFs were not found and only P7, with a missense variant effecting splice, developed sNFs. Regarding pNFs, these were observed in two patients (P11 and P15) with nonsense variants, one of them causing an effect on splicing.

The cNFs, sNFs, and pNFs[16,47], Lisch nodules and scoliosis[47] have an increasing prevalence or number with age. In two patients (P9 and P11) one of these features was present in childhood, P9 had Lisch nodules at eight years of age and P11 had pNF at five years of age. All others with at least one of these features, were in or near adulthood, the youngest being 14 years of age.

In another report, a positive NF1 family history was negatively associated with freckles[47]. In this study, freckles were observed in eight patients with no family history and in four individuals with a family history, which is similar to those reported.

Although the results obtained in this study show some correlation between phenotype and variant type it does not allow us to confirm a clear relationship between phenotypes and specific types of *NF1* variants.

Fourteen of the variants identified, in this cohort, affect functional domains of neurofibromin. There are no clear differences in the phenotype of patients with variants in functional domains and variants that are located outside these domains. This again shows the difficulty of establishing correlations, as has been reported[24].

The RNA study is essential in the molecular diagnosis of *NF1* because it allows the detection of variants that were missed by the gDNA study alone. In addition, it allows the understanding and characterization of the effect of variants on mRNA, which is fundamental in the study of *NF1* due to the high frequency of aberrant splicing events. Classifying deep intronic variants can sometimes be challenging, and cDNA sequencing helps us understand their effect. In this cohort, without the RNA study, one variant would have been missed and three exonic variants would have been mispredicted at the RNA and protein level, even using SpliceAI as an *in silico* splicing prediction tool.

RASopathies are a large group of genetic syndromes, with an overlapping of clinical manifestations, that often make the clinical diagnosis complex[1].The features of these syndromes progress with age and findings in childhood are usually different from those observed in adolescents and adults, further complicating the clinical diagnosis[1]. Furthermore, the identification of NF1 at a young age is even more challenging, as it is a progressive disease, and often the minimum criteria may not be met. Experience and medical knowledge in NF1 are fundamental in the selection of patients for the genetic study, allowing not only the earliest possible diagnosis but also guiding patients in the best possible way enabling specific surveillance at a young age.

Six patients negative for *NF1* were re-evaluated using a virtual RASopathies panel after exome sequencing. We intended to assess the benefit of changing the study workflow in these diseases, starting with a panel more comprehensive including all genes associated with RASopathies, which includes *NF1*.

In two patients (P25 and P26) two *novel* likely pathogenic variants were found in *SOS1* and *LZTR1*, respectively, allowing the establishment of a diagnosis of Noonan syndrome. If the study of these patients had started with this approach, MLPA, and cDNA sequencing would have been avoided.

Starting the study with a targeted panel for RASopathies seems to be a better approach. This will provide the answer not only for NF1 but also for its differential diagnoses with more overlapping features. In this case, the first step should be a virtual RASopathies panel by NGS/WES sequencing, and if negative proceed with *NF1* complete study with MLPA and cDNA sequencing.

Chapter 7

Conclusion and Future Perspectives

Chapter 7. Conclusion and future perspectives

This study at the UFHM allowed the conclusion of a better approach to the diagnosis of NF1 patients. This approach combining three different methods, gDNA and cDNA sequencing, and MLPA, allowed the identification of 19 different variants, four of which were *novel*.

This is a highly efficient two-step strategy for *NF1* study, identifying missense, nonsense, splice variants, deep intronic variants, duplications and small and gross deletions. The cDNA sequencing, besides its importance in diagnosis, is also fundamental for understanding the impact of DNA variants on *NF1* mRNA.

With the implementation of a RASopathies panel as a first step, a more complete and broadening study of these patients is achieved.

Phenotype-genotype correlation in NF1 is usually difficult and it was no exception in this study, although some correlations could be established. With a larger number of patients, this point could be improved in the future.

In the future, cDNA analysis by Sanger sequencing could be replaced by transcriptome, which would provide the complete sequences of all RNA products.

This internship was a very enriching professional experience that allowed me to improve my skills in the different methods of molecular biology and to have greater contact with the work in a molecular biology laboratory.

Chapter 8

References

Chapter 8. References

1. Rauen KA. Defining RASopathy. *DMM Disease Models and Mechanisms* 2022; 15: 1–6.

2. Hebron KE, Hernandez ER, Yohe ME. The RASopathies: from pathogenetics to therapeutics. *DMM Disease Models and Mechanisms*; 15. Epub ahead of print 2022.

3. Tajan M, Paccoud R, Branka S, et al. The RASopathy family: Consequences of germline activation of the RAS/MAPK pathway. *Endocr Rev* 2018; 39: 676–700.

4. Rauen KA. The RASopathies. *Annu Rev Genomics Hum Genet* 2013; 14: 355–369.

5. Zhu L, Zhang Y, Tong H, et al. Clinical and molecular characterization of NF1 patients: single-center experience of 32 patients from China. *Medicine (United States)*; 95. Epub ahead of print 2016.

6. Vandenbroucke I, Vandesompele J, De Paepe A, et al. Quantification of NF1 transcripts reveals novel highly expressed splice variants. *FEBS Lett* 2002; 522: 71–76.

7. Koster R, Brandão RD, Tserpelis D, et al. Pathogenic neurofibromatosis type 1 (NF1) RNA splicing resolved by targeted RNAseq. *NPJ Genom Med*; 6. Epub ahead of print 2021.

8. Friedman JM. *Neurofibromatosis 1*. Seattle (WA): University of Washington, Seattle; 1993-2023., https://www.ncbi.nlm.nih.gov/books/ (1998).

9. Stella A, Lastella P, Loconte DC, et al. Accurate classification of NF1 gene variants in 84 Italian patients with neurofibromatosis type 1. *Genes (Basel)* 2018; 9: 1–18.

10. De Luca A, Bottillo I, Dasdia MC, et al. Deletions of NF1 gene and exons detected by multiplex ligation-dependent probe amplification. *J Med Genet* 2007; 44: 800–808.

11. Bianchessi D, Ibba MC, Saletti V, et al. Simultaneous detection of NF1, SPRED1, LZTR1, AND NF2 gene mutations by targeted NGS in an Italian cohort of suspected NF1 patients. *Genes (Basel)* 2020; 11: 1–12.

12. Legius E, Messiaen L, Wolkenstein P, et al. Revised diagnostic criteria for neurofibromatosis type 1 and Legius syndrome: an international consensus recommendation. *Genetics in Medicine* 2021; 23: 1506–1513.

13. Kırat E, Mutlu Albayrak H. The Spectrum of *NF1* Gene Variations in Southeastern Turkey. *The Journal of Pediatric Research* 2021; 8: 286–296.

14. Pros E, Gómez C, Martín T, et al. Nature and mRNA effect of 282 different NF1 point mutations: focus on splicing alterations. *Hum Mutat* 2008; 29: E173–E193.

15. Dragos VS, Blatnik A, Klancar G, et al. Two novel NF1 pathogenic variants causing the creation of a new splice site in patients with neurofibromatosis type I. *Front Genet* 2019; 10: 1–8.

16. Sabbagh A, Pasmant E, Imbard A, et al. NF1 molecular characterization and neurofibromatosis type i genotype-phenotype correlation: The french experience. *Hum Mutat* 2013; 34: 1510–1518.

17. Anastasaki C, Le LQ, Kesterson RA, et al. Updated nomenclature for human and mouse neurofibromatosis type 1 genes. *Neurol Genet* 2017; 3: 56–58.

18. Greenberg DS, Soreq H. *Alternative Splicing*. Elsevier Inc. Epub ahead of print 2013.

19. Bergoug M, Doudeau M, Godin F, et al. Neurofibromin Structure, Functions and Regulation. *Cells*; 9. Epub ahead of print 2020.

20. Barron, Victoria A. and Lou H. Alternative splicing the Neurofibromatosis type I pre-mRNA. *Biosci Rep* 2012; 32: 131–138.

21. Landrith T, Li B, Cass AA, et al. Splicing profile by capture RNA-seq identifies pathogenic germline variants in tumor suppressor genes. *NPJ Precis Oncol*; 4. Epub ahead of print 2020.

22. Tadini G, Legius E, Brems H. *Multidisciplinary Approach to Neurofibromatosis Type 1*. 2020. Epub ahead of print 2020.

23. Kiuru M, Busam KJ. The NF1 gene in tumor syndromes and melanoma. *Laboratory Investigation* 2017; 97: 146–157.

24. Kang E, Kim YM, Seo GH, et al. Phenotype categorization of neurofibromatosis type I and correlation to NF1 mutation types. *J Hum Genet* 2020; 65: 79–89.

25. Anastasaki C, Orozco P, Gutmann DH. RAS and beyond: the many faces of the neurofibromatosis type 1 protein. *DMM Disease Models and Mechanisms*; 15. Epub ahead of print 2022.

26. Leier A, Moore M, Liu H, et al. Targeted exon skipping of NF1 exon 17 as a therapeutic for neurofibromatosis type I. *Mol Ther Nucleic Acids* 2022; 28: 261–278.

27. Mo J, Moye SL, McKay RM, et al. Neurofibromin and suppression of tumorigenesis: beyond the GAP. *Oncogene* 2022; 41: 1235–1251.

28. Kehrer H, David S. Classification of NF1 microdeletions and its importance for establishing genotype / phenotype correlations in patients with NF1 microdeletions. *Hum Genet* 2021; 140: 1635–1649.

29. Riva M, Martorana D, Uliana V, et al. Recurrent NF1 gene variants and their genotype/phenotype correlations in patients with Neurofibromatosis type I. *Genes Chromosomes Cancer* 2022; 61: 10–21.

30. Koczkowska M, Callens T, Chen Y, et al. Clinical spectrum of individuals with pathogenic NF1 missense variants affecting p.Met1149, p.Arg1276, and p.Lys1423: genotype–phenotype study in neurofibromatosis type 1. *Hum Mutat* 2020; 41: 299–315.

31. Imbard A, Pasmant E, Sabbagh A, et al. NF1 single and multi-exons copy number variations in neurofibromatosis type 1. *J Hum Genet* 2015; 60: 221–224.

32. Behjati S, Tarpey PS. What is next generation sequencing? *Arch Dis Child Educ Pract Ed* 2013; 98: 236–238.

33. Yohe S, Thyagarajan B. Review of clinical next-generation sequencing. *Arch Pathol Lab Med* 2017; 141: 1544–1557.

34. Turnpenny, P & Ellard S. *Emery's Elements os Medical Genetics*. 15th ed. 2017.

35. Schouten JP, McElgunn CJ, Waaijer R, et al. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res* 2002; 30: e57.

36. Ha C, Kim JW, Jang JH. Performance evaluation of spliceai for the prediction of splicing of nf1 variants. *Genes (Basel)*; 12. Epub ahead of print 2021.

37. Stewart DR, Korf BR, Nathanson KL, et al. Care of adults with neurofibromatosis type 1: A clinical practice resource of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine* 2018; 20: 671–682.

38. Miller DT, Freedenberg D, Schorry E, et al. Health Supervision for Children With Neurofibromatosis Type 1. *American Academy of Pediactrics* 2021; 143: 1–16.

39. Gutmann DH, Ferner RE, Listernick RH, et al. Neurofibromatosis type 1. *Nat Rev Dis Primers* 2017; 3: 1–18.

40. Wimmer K, Roca X, Beiglbo¨ck H, et al. Extensive In Silico Analysis of NF1 Splicing Defects Uncovers Determinants for Splicing Outcome Upon 50 Splice-Site Disruption. *Hum Mutation* 2007; 28: 599–612.

41. Richards S, Aziz N, Bale S, et al. 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* 2015; 17: 405–424.

42. Tavtigian S V, Harrison SM, Boucher KM, et al. Fitting a naturally scaled point system to the ACMG/AMP variant classification guidelines. 2021; 41: 1734–1737.

43. Wimmer K, Schamschula E, Wernstedt A, et al. AG-exclusion zone revisited: Lessons to learn from 91 intronic NF1 3′ splice site mutations outside the canonical AG-dinucleotides. *Hum Mutat* 2020; 41: 1145–1156.

44. Jang MA, Kim YE, Kim SK, et al. Identification and characterization of NF1 splicing mutations in Korean patients with neurofibromatosis type 1. *J Hum Genet* 2016; 61: 705–709.

45. Morbidoni V, Baschiera E, Forzan M, et al. Hybrid minigene assay: An efficient tool to characterize mrna splicing profiles of nf1 variants. *Cancers (Basel)* 2021; 13: 1–26.

46. Wallis D, Li K, Lui H, et al. Neurofibromin (NF1) genetic variant structure–function analyses using a full-length mouse cDNA. *Hum Mutat* 2018; 39: 816–821.

47. Scala M, Schiavetti I, Madia F, et al. Genotype-phenotype correlations in neurofibromatosis type 1: A single-center cohort study. *Cancers (Basel)* 2021; 13: 1–27.

48. Aoki Y, Niihori T, Inoue SI, et al. Recent advances in RASopathies. *J Hum Genet* 2016; 61: 33–39.

49. De Luca A, Schirinzi A, Buccino A, et al. Novel and recurrent mutations in the NF1 gene in Italian patients with neurofibromatosis type 1. *Hum Mutat* 2004; 23: 629.

50. Fahsold R, Hoffmeyer S, Mischung C, et al. Minor lesion mutational spectrum of the entire NF1 gene does not explain its high mutability but points to a functional domain upstream of the GAP- related domain. *Am J Hum Genet* 2000; 66: 790–818.

51. Robinson PN, Böddrich A, Peters H, et al. Two recurrent nonsense mutations and a 4 bp deletion in a quasi-symmetric element in exon 37 of the NF1 gene. *Hum Genet* 1995; 96: 95–98.

52. Purandare SM, Lanyon WG, Connor JM. Characterisation of inherited and sporadic mutations in neurofibromatosis type-1. *Hum Mol Genet* 1994; 3: 1109–1115.

53. Laycock-Van Spyk S, Thomas N, Cooper DN, et al. Neurofibromatosis type 1-associated tumours: Their somatic mutational spectrum and pathogenesis. *Hum Genomics* 2011; 5: 623–690.

54. Heim RA, Kam-morgan LNW, Binnie CG, et al. Distribution of 13 truncating mutations in the neurofibromatosis 1 gene. *Hum Mol Genet* 1995; 4: 975–981.

55. Robinson PN, Buske A, Neumann R, et al. Recurrent 2-bp deletion in exon 10c of the NF1 gene in two cases of von Recklinghausen neurofibromatosis. *Hum Mutat* 1996; 7: 85–88.

56. Toliat MR, Erdogan F, Gewies A, et al. Analysis of the NF1 gene by temperature gradient gel electrophoresis reveals a high incidence of mutations in exon 4b. *Electrophoresis* 2000; 21: 541–544.

57. Upadhyaya M, Maynard J, Osbom M, et al. Characterisation of germline mutations in the neurofibromatosis type 1 (NF1) gene. *J Med Genet* 1995; 32: 706–710.

58. Valero MC, Velasco E, Moreno F, et al. Characterization of four mutations in the neurofibromatosis type 1 gene by denaturing gradient gel electrophoresis (DGGE). *Hum Mol Genet* 1994; 3: 639–641.

59. Nemethova M, Bolcekova A, Ilencikova D, et al. Thirty-Nine novel neurofibromatosis 1 (NF1) gene mutations identified in slovak patients. *Ann Hum Genet* 2013; 77: 364–379.

Chapter 9

Supplementary Material

Chapter 9. Supplementary Material

Table S1: *NF1* primers for cDNA amplification and Sanger sequencing.

Fragment	Target	Primer	Sequence 5' - 3'	Use For
F1	E 1	NF1-F1F	CTTGCCTCTTCCCTCACCTC	PCR
	E 9	NF1-F1R	TCAGCTGCCTACTTCCTCCA	PCR
	E 4	NF1-F1seqF	AGCTGCAACAACTTCAATG	Sequence
F2	E 8	NF1-F2F	AAAGACGTGGTTGATGAAAAC	PCR
	E 17	NF1-F2R	AAAAGGAGAAAGTGACAGGAAC	PCR
	E 15	NF1-F2seqR	TGATGAAGAACCAGCAGAG	Sequence
F3	E 12	NF1-F3F	GAGCACACCCAGCAATACGA	PCR
	E 12	NF1-F3F2	GATGCTGTGTATTGTCACTCGG	PCR
	E 22	NF1-F3R	CTAGCTTGCCCTAGATGTTCAGA	PCR
	E 18	NF1-F3seqF	GCCCAACTATAACACATTCA	Sequence
	E 19	NF1-F3seqR	CAGTGGGATGCTCAATGCG	Sequence
F4	E 21	NF1-F4F	AGTCCTGCTCTGTATCCAATG	PCR
	E 33	NF1-F4R	GTAGACGATGTAAAGCAAGCAC	PCR
	E 25	NF1-F4seqF1	GGAAGCCAAATCACAGTTAT	Sequence
	E 26	NF1-F4seqR	CGTTGGCATTGAGTAAGTTTGA	Sequence
	E 27	NF1-F4seqF2	AGGCACAGAATTTGACACAC	Sequence
	E 30-31	NF1-F4-E31-201	GTGGTTAGCCAGCGTTTC	Sequence
	E 30-30alt31	NF1-F4-E31-202	GCAACTTGCCACTCCCT	Sequence
	E 30-32	NF1-F4-E30.32-201	GTTTATACCAGATACTTCAGAG	Sequence
F5	E 32	NF1-F5F	CTTCAGAGTATTGCCAATCATGT	PCR
	E 37	NF1-F5R	TGATAGACTGGACAATGGCTTC	PCR
	E 34-35	NF1-F5seqF	ATGACTAGGCATCAGGTACA	Sequence
F6	E 36	NF1-F6F	CTGGCTGAGCACATAGAGC	PCR
	E 43	NF1-F6R	TTGCACGTTGGAATATCTCTCA	PCR
	E 40	NF1-F6seqR	TTCCAGAAGCCAAAGCTACA	Sequence
F7	E 42	NF1-F7F	TCATTGCCTTCCGTTCCAGT	PCR
	E 52	NF1-F7R	AAATGTGGGTGCTGTTGTGATG	PCR
	E 44	NF1-F7seqF	TGTATTAGCAAACGAGTGTCT	Sequence
	E 51	NF1-F7seqR	CAGTTTCTGCTACTCTCCTCAT	Sequence
F8	E 48	NF1-F8F	ACAGAGCGTGGCCTACTTAGCA	PCR
	E 48	NF1-F8F1	GAATACACAGAGCGTGGC	PCR
	E 57	NF1-F8R	AACAGGAAGTGCAGCATTACAACA	PCR
	E 57	NF1-F8R1	AACAGGAAGTGCAGCATTAC	PCR
	E 51	NF1-F8seqF	CCCCCCAAAATGAGGAGAGT	Sequence
	E 54	NF1-F8seqF2´	ATCCCCACCACAATACCAAA	Sequence
	E 57	NF1-F8seqR	GAAAGCAAGCAAGCTTCACA	Sequence
	E 54	NF1-F8seqR2	TTTGGTATTGTGGTGGGGAT	Sequence

E - Exon; F - Fragment

Table S2: *NF1* primers for DNA amplification and Sanger sequencing.

Target	Primer	Sequence 5' - 3'	Ta (ºC)
IVS 12	NF1-Int12R	AACTGGATAAGAGTGAACCACAA	59
	NF1-Int12seqF	ACTGATAGGAACTTTGATACTG	
Exon 21	NF1-21D2	GATCAGTGGCTCTTTAAAA	54
	NF1-21R2	TTAGGCACACATACACAC	
Exon 23	NF1-23R	TAACCGCATATCTACTCTT	54
Exon 45	NF1-45D	GTTCCTGAATTCATTCCGAGA	56
	NF1-45R	CTTCAAACAACTAAACTTTAAGAGG	
Exon 50	NF1-50D	TTGGAAGGAGCAAACGATG	58
	NF1-50R	ACTTTGCTACACTGACATGG	

Table S3: *SPRED1* primers for DNA amplification and Sanger sequencing.

Target	Primer	Sequence 5' - 3'
Exon 1	SPRED1-1D	TACCGTTCTGGGTGAGGC
	SPRED1-1R	CCCAAGTTTCGGATGGGTC
Exon 2	SPRED1-2D	CAAGACTGATGGCTTGGCT
	SPRED1-2R	TAACACAGAAACAGCTCCAG
Exon 3	SPRED1-3D	TAGCGTTGTATCACCTCAG
	SPRED1-3R	AAAGCCTGGTCACATATCAC
Exon 4	SPRED1-4D	AGTGGCCAGTACCTTAATTG
	SPRED1-4R	GATGCTCAACCTGTATTGGT
Exon 5	SPRED1-5D	TGGGAATTGCTATTCATAGCG
	SPRED1-5R	TACTGTGTCTGGTAAAGGGC
Exon 6	SPRED1-6D	TAGGTGGGGGGAAATGATTC
	SPRED1-6R	GCTCTGGCAATCTTTTAGACT
Exon 7	SPRED1-7D	CCTCATAGTCCACCAACTGA
	SPRED1-7R	CGCCAAGAACTGATGAGTTC

Table S4: PCR reaction for *NF1* cDNA sequencing.

Mix 1	**V_{final} = 18 μL**	**Mix 2**	**V_{final} = 23,5 μL**
SequalPrep Buffer (10x)	2 μL	*Taq* Buffer (10x)	2,5 μL
Enhancer A (10x)	2 μL	dNTPs 2,5 mM	2 μL
DMSO 100%	0,4 μL	DMSO 100%	1,25 μL
Forward Primer (50 $\mathrm{ng}/\mu\mathrm{L}$)	1,25 μL	Kit Primers (1,5 mM)	3,5 μL
Reverse Primer (50 $\mathrm{ng}/\mu\mathrm{L}$)	1,25 μL	*SequalPrep* (5 $\mathrm{U}/\mu\mathrm{L}$)	0,2 μL
SequalPrep (5 $\mathrm{U}/\mu\mathrm{L}$)	0,3 μL	dH_2O	14,05 μL
dH_2O	10,8 μL		

Table S5: PCR conditions for *NF1* cDNA sequencing.

Fragment	**Mix**	**Ta (ºC)**	**Cycles**
1	1	63	35
2		57	35
3		57	35
5		57	35
7		62	35
8		56	37
4	2	57	35
6		57	37

Table S6: PCR condition for *NF1* DNA sequencing.

Mix	**V_{final} = 25 μL**
Glucose Buffer* (10x)	2,5 μL
dNTPs (10 mM)	0,5 μL
Forward Primer (10 $\mu\mathrm{M}$)	0,5 μL
Reverse Primer (10 $\mu\mathrm{M}$)	0,5 μL
Taq polymersase (5 $\mathrm{U}/\mu\mathrm{L}$)	0,2 μL
dH_2O	21 μL

45 seconds of extension.
*Glucose buffer [10x]: Tris pH8.8 6.7 mM, $(NH4)2SO4$ 1.66 mM, $Na2$ EDTA 6.7 µM, $MgCl2$ 2.5 mM, BSA 16 µg/mL, B-Me 10 mM.

Table S7: PCR condition for *SPRED1* DNA sequencing.

Mix	**V_{final} = 25 μL**
Glucose Buffer* (10x)	2,5 μL
dNTPs (10 mM)	0,5 μL
Forward Primer (10 μM)	0,75 μL
Reverse Primer (10 μM)	0,75 μL
Taq Taq polymersase (5U/μL)	0,2 μL
dH_2O	20,5 μL

Ta – 60 ºC; 45 seconds of extension.
*Glucose buffer [10x]: Tris pH8.8 6.7 mM, $(NH_4)_2SO_4$ 1.66 mM, Na_2 EDTA 6.7 µM, $MgCl_2$ 2.5 mM, BSA 16 µg/mL, B-Me 10 mM.