

Development, validation and implementation of an NGS gene panel approach for the diagnosis of primary immunodeficiencies

Raquel Alexandra Fonseca da Silva

Mestrado em Biologia Celular e Molecular

Departamento de Biologia, FCUP

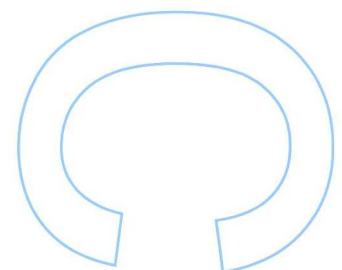
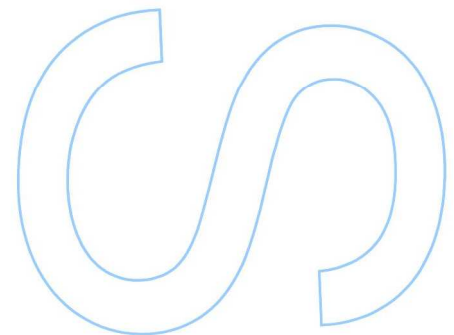
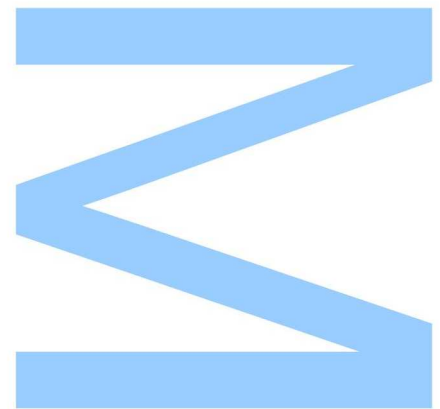
2020

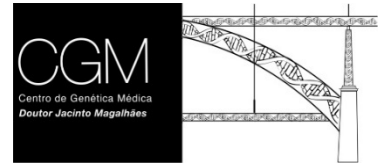
Orientador

Márcia Oliveira, Técnica Superior de Saúde, Centro de Genética Médica
Doutor Jacinto Magalhães, Centro Hospitalar Universitário do Porto

Co-orientador

Maria João Prata, Professora Associada, Faculdade de Ciências da
Universidade do Porto

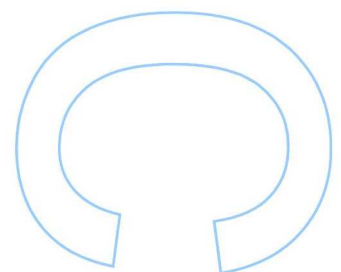
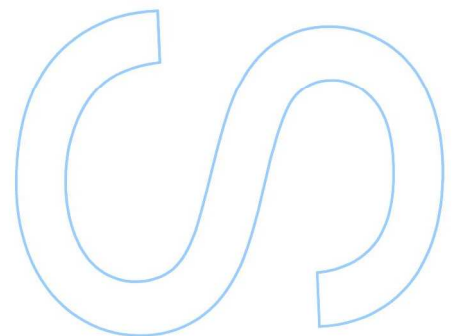
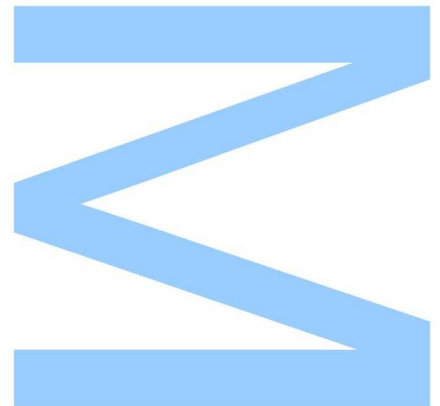




Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ____ / ____ / ____



Poster Presentation

Silva, R., Gonçalves, A., Oliveira, M. E., Neves, E., Marques, L., Fortuna, A. M., Santos, R. (2020) Design and Validation of an NGS approach for the diagnosis of Primary Immunodeficiencies. XLVIII Conferências de Genética Doutor Jacinto Magalhães '40 Anos Ao Serviço Da Comunidade', Centro de Genética Médica Doutor Jacinto Magalhães, Centro Hospitalar Universitário do Porto, 31st January and 01st February 2020, Porto.

Agradecimentos

Em primeiro lugar, gostaria de agradecer à Doutora Márcia Oliveira e à Ana Gonçalves pela orientação, motivação, paciência e por tudo o que ensinaram ao longo deste ano. Não poderia ter tido melhores pessoas a orientar-me neste último ano.

À Doutora Rosário Santos, por me ter acolhido e me ter dado a oportunidade de desenvolver o meu trabalho na Unidade de Genética Molecular do CGMJM.

À Professora Doutora Maria João Prata, pela disponibilidade, por me ter acompanhado neste projeto e por todos os conhecimentos que me transmitiu.

A todos os membros da Unidade de Genética Molecular, pela amizade, boa-disposição e toda a ajuda que me deram quando precisei. Sem dúvida que devo a vocês também a minha aprendizagem e evolução nesta maravilhosa área da genética. Um obrigada especial à Catarina e à Rita por terem acompanhado este trabalho de perto e por terem vivido comigo toda a experiência.

À minha mãe, por ser o maior apoio em tudo o que faço, por estar sempre disponível para me ouvir e para me dar bons conselhos. Sei que estarás ao meu lado aconteça o que acontecer.

Aos meus colegas do Mestrado em Biologia Celular e Molecular, por toda a experiência que partilhamos nestes dois últimos anos, pelos momentos de diversão e por me mostrarem que estamos todos juntos nisto.

Aos meus amigos, pelos bons momentos que me proporcionam, por me permitirem esquecer o trabalho por momentos quando estamos juntos e por poder partilhar convosco a conclusão de mais uma etapa.

Abstract

Primary immunodeficiencies (PIDs) are a heterogeneous group of inherited disorders that impair the development or function of the human immune system. PIDs cause increased mortality, so early diagnosis is crucial for initiation of the proper treatment and to improve the prognosis of patients. The diagnosis is complex due to the high genotypic and phenotypic heterogeneity associated to PIDs. Next-generation sequencing (NGS) is a comprehensive approach for diagnosis of PIDs that may provide a more efficient identification of the underlying genetic cause. The objectives of this study were to design, validate and implement a cost-effective NGS-based strategy for PIDs diagnosis. Three NGS multigene panels were designed targeting most of the currently known PID-related genes. One of these panels (namely Panel A), including a total of 117 genes, was tested and validated using 4 positive control DNA samples with previously identified pathogenic variants, and its clinical applicability was evaluated using genomic DNA samples from 4 patients without a genetic diagnosis. The experimental coverage obtained for Panel A was very satisfactory, originating an overall coverage of 315,65x with 99,04% of the target bases covered at least 20x. The assay was able to accurately detect the previously identified disease-causing variants in the 4 positive control DNA samples tested. However, none underlying genetic defect was identified when the 4 genetically uncharacterized patients were subjected to analysis with the panel, which demonstrates the heterogeneity and complexity associated to this type of diseases. Nevertheless, the present work showed that targeted NGS panel is an effective and rapid tool that can be used as the first-tier genetic assay in order to improve PIDs diagnosis.

Keywords

Primary immunodeficiency, genetic diagnosis, next-generation sequencing, targeted panel

Resumo

As imunodeficiências primárias (IDPs) são um grupo heterogêneo de doenças hereditárias que afetam o desenvolvimento ou a função do sistema imunológico humano. As IDPs diminuem a esperança e qualidade de vida dos doentes e, por isso, o diagnóstico precoce é fundamental para o início do tratamento adequado e a melhoria do prognóstico dos pacientes. O diagnóstico é complexo devido à alta heterogeneidade genotípica e fenotípica associada às IDPs. A sequenciação de nova geração (NGS, *next-generation sequencing*) é uma abordagem abrangente para o diagnóstico das IDPs que pode contribuir para uma identificação mais eficiente da causa genética subjacente. Os objetivos deste estudo foram desenhar, validar e implementar uma abordagem económica para o diagnóstico de IDPs, usando uma metodologia baseada em NGS. Foram desenhados três painéis de NGS multigénicos que incluem a maioria dos genes associados a IDPs atualmente conhecidos. Um desses painéis (nomeadamente o Painel A), que inclui um total de 117 genes, foi testado e validado usando 4 controlos positivos com variantes causais previamente identificadas, e sua aplicabilidade clínica foi avaliada usando amostras de DNA genómico de 4 pacientes sem diagnóstico genético. A fiabilidade experimental obtida para o Painel A foi bastante satisfatória, com uma cobertura geral de 315,65x com 99,04% das bases alvo cobertas pelo menos 20x. O ensaio foi capaz de detetar com precisão as variantes patogénicas identificadas anteriormente nas 4 amostras de DNA controlo testadas. No entanto, não foi identificado qualquer defeito genético subjacente quando os 4 pacientes geneticamente não caracterizados foram submetidos a análise com o painel, o que demonstra a heterogeneidade e a complexidade associada a este tipo de doenças. No entanto, o presente trabalho mostrou que este tipo de painel NGS direcionado é uma ferramenta rápida e eficaz que pode ser usada como estratégia de testagem genética de primeira linha para melhorar o diagnóstico das IDPs.

Palavras-chave

Imunodeficiência primária, diagnóstico genético, sequenciação de nova geração, painel direcionado

Table of Contents

Poster Presentation	III
Agradecimentos	IV
Abstract	V
Resumo	VI
List of Figures	IX
List of Tables	X
Abbreviations	XI
1. Introduction	1
1.1 Primary Immunodeficiencies.....	1
1.2 Classification of Primary Immunodeficiencies.....	2
1.3 Diagnosis and Treatment of Primary Immunodeficiencies	5
1.4 Genetic testing for PIDs.....	6
1.4.1 Next-generation sequencing technologies and their clinical use.....	6
1.4.2 Exome and genome sequencing	8
1.4.3 Functional studies for novel variants	9
1.5 Newborn Screening for Primary Immunodeficiencies.....	10
2. Aims	13
3. Materials and Methods.....	14
3.1 NGS multigene panels design	14
3.2 DNA samples studied	15
3.3 Panel A validation.....	22
3.3.1 DNA preparation and quantification	22
3.3.2 DNA library preparation and next-generation sequencing	23
3.3.3 Data analysis and variant interpretation	25
3.4 Panel A implementation.....	26
3.5 Development of a conventional molecular approach for the study of other PID-related genes	27
3.5.1 Primer design for PCR	27
3.5.2 Symmetric PCR	27
3.5.3 Enzymatic purification of PCR products	28
3.5.4 Sequencing PCR	28
4. Results.....	30
4.1 Panel A validation using positive control DNA samples	30
4.1.1 Gene coverage (predicted and experimental)	30
4.1.2 Variants detected in positive control patients	30

4.2 Panel A implementation.....	35
4.3 Development of a conventional molecular approach to complement the study of PID regions/genes.....	38
5. Discussion	40
6. Final Conclusions and Future Perspectives	44
7. Bibliography	46
8. Appendices	56
Appendix 1: Coverage metrics for Panel A genes	56
Appendix 2: Primers designed for the study of PID regions/genes.	59
Appendix 3: Sequencing metrics for control and patient samples.....	65
Appendix 4: Variants detected by Sanger sequencing in the 4 patients.....	66
Appendix 5: Poster presented at “XLVIII Conferências de Genética Doutor Jacinto Magalhães: 40 Anos Ao Serviço Da Comunidade”.....	67

List of Figures

Figure 1. Schematic representation of Ion AmpliSeq™ workflow	23
Figure 2. Conditions of the programs used for amplification of regions of interest.....	28
Figure 3. Conditions used in sequencing reactions	29
Figure 4. Binary alignment map file showing the variant identified in <i>BTK</i> gene in control patient 1.....	31
Figure 5. Binary alignment map file showing the variant identified in <i>ATM</i> gene in control patient 2.....	32
Figure 6. Binary alignment map file showing the variant identified in <i>LYST</i> gene in control patient 3.....	33
Figure 7. Binary alignment map file showing the variant identified in <i>IL7R</i> gene in control patient 4.....	34
Figure 8. Electrophoretic results obtained in the first amplification test of PID genes using a control DNA and standard conditions.....	38
Figure 9. Electrophoretic results obtained in the second amplification test of PID genes using a control DNA and high-GC conditions	39

List of Tables

Table 1. Classification of primary immunodeficiencies	2
Table 2. PID categories included in the three NGS multigene panels	14
Table 3. NGS multigene panels features.....	15
Table 4. Description of the genes included in Panel A and associated phenotypes	16
Table 5. Clinical data of the studied patients.....	22
Table 6. Reaction conditions for DNA quantification by qPCR	22
Table 7. Reaction conditions for targets amplification	23
Table 8. Reaction conditions for partial digestion of amplicons	24
Table 9. Conditions of the ligation reaction	24
Table 10. Conditions for library quantification by qPCR	25
Table 11. Description of the variants of uncertain significance identified in the 4 patients	35
Table 12. Bioinformatic analysis for the identified missense variants	37
Table A1. Coverage metrics for the genes included in Panel A	56
Table A2. Primers designed for low coverage regions	59
Table A3. Primers designed for the study of additional PID-related genes	61
Table A4. Sequencing metrics	65
Table A5. Description of the variants detected by Sanger sequencing in the 4 patients	66

Abbreviations

aCGH – Microarray-based Comparative Genomic Hybridization

ATM – Ataxia-telangiectasia mutated

BAM – Binary Alignment Map

bp – Base pairs

BTK – Bruton's tyrosine kinase

cDNA – Complementary deoxyribonucleic acid

CDS – Coding DNA sequence

CHS – Chediak-Higashi syndrome

CID – Combined immunodeficiency

CNV – Copy number variation

CVID – Common variable immunodeficiency

ddNTP – Dideoxynucleotide triphosphate

DNA – Deoxyribonucleic acid

dNTP – Deoxyribonucleotide triphosphate

gDNA – Genomic deoxyribonucleic acid

HSCT – Hematopoietic stem cell transplant

IL7R – Interleukin 7 receptor

KREC – Kappa-deleting recombination excision circles

LYST – Lysosomal trafficking regulator

MAF – Minor allele frequency

MHC – Major histocompatibility complex

MLPA – Multiplex Ligation-dependent Probe Amplification

mRNA – Messenger ribonucleic acid

NBS – Newborn screening

NGS – Next-generation sequencing

NTC – Negative Template Control

PCR – Polymerase chain reaction

PID – Primary immunodeficiency

qPCR – Quantitative real-time polymerase chain reaction

SCID – Severe combined immunodeficiency

SNV – Single-nucleotide variant

TREC – T cell receptor excision circles

VCF – Variant Caller Format

VUS – Variant of uncertain significance

WES – Whole exome sequencing

WGS – Whole genome sequencing

XLA – X-linked agammaglobulinemia

1. Introduction

1.1 Primary Immunodeficiencies

Primary immunodeficiencies (PIDs) are a genetically and phenotypically heterogeneous group of inherited diseases that result in impairment of the development and/or function of the human immune system. As a result, there is increased susceptibility to infections, autoimmunity, autoinflammation and malignancies (Bousfiha *et al.*, 2020). Most PIDs manifest early in infancy and childhood, but can also occur later in life, in the second or third decades of life.

The large majority of PIDs are considered monogenic disorders, following thus the classical patterns of Mendelian inheritance. However, with the advances in genomic technologies, multigenic defects have already been identified underlying PIDs (Germeshausen *et al.*, 2010; Zhang *et al.*, 2014). Moreover, although PIDs typically result from germline defects, they may also arise from somatic mutations (Dowdell *et al.*, 2010; Takagi *et al.*, 2011).

The mode of inheritance of these conditions can be dominant or recessive, autosomal or X-linked, showing in any case complete or incomplete penetrance. Their causative variants result in gain-of-function or loss-of-function of the encoded protein (Bousfiha *et al.*, 2020). The defective proteins may be involved in different biological processes, such as immune development, signaling cascades, effector-cell functions, and maintenance of immune homeostasis (Maródi and Notarangelo, 2007). Complete loss of protein function leads to severe forms of PIDs that normally manifest in early childhood, whereas hypomorphic variants that allow partial expression of the protein result in milder forms with later onset of the disease (Jacobs *et al.*, 2011; Babushok and Bessler, 2015).

PIDs are considered rare diseases, with an estimated global incidence of 1:10 000 to 1:100 000 (Joshi *et al.*, 2009). However, such incidence likely represents an underestimate as a result of the underdiagnosis of PIDs either due to limited access to diagnostic resources or due to difficulties in the diagnosis of cases with atypical presentations. Actually, recent studies indicate that PIDs may be more common than generally thought, reaching a prevalence between 1:1000 and 1:5000 (Zhang *et al.*, 2017).

In Portugal, despite the scarcity of data related to the epidemiology of PIDs, the number of cases in adults has been increasing, with 314 adult patients followed by immunoallergy services in 2015 (Duarte Ferreira *et al.*, 2018). Most of these patients

(92%) presented predominantly antibody deficiencies, particularly selective IgA deficiency and common variable immunodeficiency (CVID). However, only 6,7% of them had a molecular diagnosis (Duarte Ferreira *et al.*, 2018).

1.2 Classification of Primary Immunodeficiencies

The increasing use of next-generation sequencing (NGS) technologies has led to the rapid discovery of novel PID disorders, as well of previously unknown genetic causes for PIDs, in such an impressive way that currently 430 distinct genetic defects have been identified across more than 400 distinct disorders recognized so far (Bousfiha *et al.*, 2020). Since 1999, the International Union of Immunological Societies (IUIS) Expert Committee on Primary Immunodeficiencies has proposed a classification for PIDs (Picard *et al.*, 2018), which is updated every two years to include new disorders and disease-causing genes. To validate a new gene as a cause for PIDs and include it in the classification, its pathogenicity must be clearly supported. Thus, a new gene is included when multiple cases from unrelated kindreds with the same genetic cause are described, or, if few cases are described, when unequivocal pathogenic evidence is provided, generally by performing animal or cell culture studies (Tangye *et al.*, 2020).

This classification intends to ease the identification of the diseases and support immunologists and clinical researchers worldwide.

PIDs are classified in 9 categories according to common pathogenesis and components of the immune system involved (Table 1).

Table 1. Classification of primary immunodeficiencies.

PID categories
I. Immunodeficiencies affecting cellular and humoral immunity (a) Severe combined immunodeficiencies (SCID) (b) Combined immunodeficiencies (CID) generally less profound than SCID
II. Combined immunodeficiencies with associated or syndromic features
III. Predominantly antibody deficiencies
IV. Diseases of immune dysregulation
V. Congenital defects of phagocyte number or function
VI. Defects in intrinsic and innate immunity
VII. Autoinflammatory diseases
VIII. Complement deficiencies
IX. Phenocopies of primary immunodeficiencies

In the first category are immunodeficiencies affecting both cellular and humoral immunity, corresponding to the two types of adaptive immunity, i.e., the components of both cellular and humoral responses. Adaptive immunity is stimulated by the exposure to invading microorganisms and the magnitude of the response increases with repeated exposures to the infectious agent. Humoral immunity is mediated by B cells, which secrete antibodies that help in the recognition and elimination of the microbes. Cellular immunity is mediated by T cells, which are responsible for the defense against intracellular microbes and helping B cells in the production of antibodies (Abbas *et al.*, 2017).

Immunodeficiencies affecting both cellular and humoral immunity include both severe combined immunodeficiencies (SCID) and combined immunodeficiencies that are less profound than SCID. Severe combined immunodeficiencies are one of the most serious forms of PIDs and are characterized by deficiency of T cells. B cells may or may not be present, but when present do not function properly (Yu *et al.*, 2016). Combined immunodeficiencies less profound than SCID are caused by deficiencies in CD40 and CD40L proteins and in the major histocompatibility complex (MHC) (International Patient Organisation for Primary Immunodeficiencies - IPOPI, 2016). CD40 and CD40L are found on B and T cells, respectively, and help in B cell activation and in increasing T cell response. MHC molecules are present on lymphocyte's surface and present antigens to T cells, contributing to their activation.

The second category includes combined immunodeficiencies with associated or syndromic features. This group, like the first one, is associated with impairments in T and B cells. However, not only the immune system is affected, but also other organ systems. In these disorders, important cellular pathways are compromised, consequently affecting multiple cell types. Syndromic PIDs may be caused, for example, by DNA repair defects, whose effects have repercussions on all cell types, or by defects in signaling molecules that transmit signals from several cell types, thereby affecting different organs (Kersseboom *et al.*, 2011).

Predominantly antibody deficiencies constitute the third category and represent the most common PID worldwide (Rezaei *et al.*, 2017). In this form of PIDs, one or more immunoglobulin isotypes are decreased or non-functional and B cells may be present in reduced or normal numbers.

The fourth category includes diseases of immune dysregulation, a diverse group of diseases caused by abnormalities in the regulation of mechanisms of the immune system. These consist in disorders related to defects in regulatory T cells, whose function is to inhibit immune responses, including against self-antigens, syndromes with

autoimmunity, with or without uncontrolled proliferation of lymphocytes (lymphoproliferation), disorders associated with abnormal high levels of interferons that help in the initiation of immune responses, among others (International Patient Organisation for Primary Immunodeficiencies - IPOPI, 2016).

The fifth category corresponds to congenital defects of phagocyte. Phagocytes are cells involved in the first-line immune defense, that ingest invading microorganisms by phagocytosis and eliminate them. Phagocytes include neutrophils and macrophages. Diseases regarding phagocyte defects may be caused by a reduced number or an impaired function of these cells, like several congenital neutropenias, defects in phagocyte motility, inhibiting their migration toward the invading cells, and defects in phagocyte's respiratory burst, the process by which these cells release reactive oxygen species to kill the microbes (International Patient Organisation for Primary Immunodeficiencies - IPOPI, 2016).

Defects in intrinsic and innate immunity form the sixth category. The innate immune system is responsible for the early defense against infections, responding almost immediately to pathogens. These responses are stimulated by molecules shared by groups of related microorganisms, as innate immunity is not specific for particular antigens (Abbas *et al.*, 2017). The innate immune system is composed by various cell types, including phagocytes, mast cells, dendritic cells, natural killer (NK) cells, and also components of the complement system. Gene defects affecting these components result in several conditions that predispose patients to severe infections caused by certain microorganisms, like mycobacteria, viruses, and fungi (Parvaneh *et al.*, 2017).

In the seventh category are autoinflammatory disorders, which are caused by a dysregulation of the innate immune system, but rather than increasing the risk of infection, they result in excessive inflammation. Thus, autoinflammatory disorders are caused by defects in molecules involved in the regulation of innate immune responses. Inflammasomes are multiprotein complexes responsible for generating active forms of cytokines that promote inflammatory responses. Defects affecting these complexes, either their components or signaling molecules involved in their expression, that result in increased inflammasome activity, are one cause of autoinflammatory disease (Ciccarelli *et al.*, 2013).

The eighth category includes complement deficiencies. The complement system is a group of proteins that aid phagocytes identifying and ingesting pathogens, promote inflammatory responses and attack microorganism's cell membrane. Complement deficiencies may result from defects in components of the complement system or from defects in regulatory proteins of this system (Audemard-Verger *et al.*, 2016).

The ninth and final category includes phenocopies of PIDs. It is usually assumed that these conditions are not caused by inherited genetic variants, but instead either by somatic mutations acquired at some point during life or by autoantibodies that target self-proteins, such as components of the immune system (International Patient Organisation for Primary Immunodeficiencies - IPOPI, 2016).

1.3 Diagnosis and Treatment of Primary Immunodeficiencies

Early diagnosis of these disorders is critical, since the manifestations associated to PIDs cause an increased mortality in patients (Rae *et al.*, 2018), especially in those affected with severe forms. Identification of the underlying genetic cause allows prenatal and preimplantation diagnosis in families with a history of primary immunodeficiency, as well as carrier identification, this way supporting genetic counselling (Lee *et al.*, 2016). This can help families evaluating the risk of having affected children and pondering on future pregnancies. Importantly, it also allows identification of patients before the onset of the disease, decreasing the severity of the disease since the proper treatment is initiated in advance (Cifaldi *et al.*, 2019).

An accurate diagnosis greatly assists in the selection of treatment. Hematopoietic stem cell transplant (HSCT) is the standard treatment for patients with SCID and yields better results in younger patients, being the survival rate of infants transplanted before 3.5 months of age higher than those transplanted after the same age (Pai *et al.*, 2014). Besides HSCT, other treatments can be used for PIDs, especially if the patient has no matched HSC donor, if there is a high mortality rate related to HSCT for that specific PID form or if the patient has no access to a medical facility with experience in HSCT for PID cases (Seleman *et al.*, 2017).

Identification of the genetic alterations and their functional impact on proteins allows the use of precision therapies that target specific defects, such as gene therapy, targeted immunosuppression or use of biologics directed specifically to the affected pathway (Lenardo *et al.*, 2016). Thus, a prompt diagnosis and treatment of PIDs are essential to reduce the disease-associated mortality and improve the prognosis of PID patients.

The diagnostic approach for PIDs entails several steps: the first one consists in obtaining a detailed personal and family history, and the second step consists in performing laboratory evaluations (Al-Mousa *et al.*, 2016). Current laboratory assays involve a phenotypic and functional characterization, including complete blood count, to

assess the presence of lymphopenia and abnormal lymphocytes or phagocytic cells, lymphocyte proliferation assays and flow cytometry, to quantify both T and B cells and NK cells, measurement of serum immunoglobulin levels, neutrophil function assays and complement analyses, to evaluate the number and function of complement proteins (McCusker and Warrington, 2011). To confirm and establish a definitive diagnosis, identification of the disease's genetic cause through molecular assays is crucial.

1.4 Genetic testing for PIDs

Since the identification of the first primary immunodeficiencies in the early 1950s and the first application of genetic testing for PIDs in 1993 (Ochs and Hitzig, 2012), new methods have been developed in order to expedite identification of the cellular and molecular basis of these conditions.

Selection of candidate genes to be tested is guided by the clinical and immunological characteristics of the affected patients. However, gene selection is complex, not only because of the large number of PID-related genes, but also because of the high genotypic and phenotypic heterogeneity associated to PIDs. Similar genetic defects may result in distinct phenotypes and similar phenotypes may be caused by different genetic variations (Nijman *et al.*, 2014). The complexity of PIDs diagnosis is reflected in the significant number of cases that remain without a definitive genetic diagnosis (Gallo *et al.*, 2016). Thus, the use of a comprehensive and effective approach for the molecular testing of these disorders is of great importance.

Considering the rapid evolution in the field of PIDs, with novel disorders and disease-causing genes being continuously discovered, sequencing multiple candidate genes by the conventional method of Sanger sequencing is a laborious and time-consuming approach (Gallo *et al.*, 2016). Sanger sequencing may be a reliable method when is possible to come up to one gene as the most likely disease cause, otherwise, it can be an inefficient strategy to find “the needle in the haystack”.

1.4.1 Next-generation sequencing technologies and their clinical use

Application of NGS-based technologies has revolutionized the field of rare genetic diseases, including PIDs, expanding knowledge and providing a better understanding of the genetic basis of these diseases (Bisgin *et al.*, 2018).

Targeted NGS sequencing allows simultaneous sequencing of multiple genes, which is a great advantage considering the heterogeneity of PID disorders. Using this approach, a genetic diagnosis may be obtained in cases for which it would be difficult to select candidate genes to test, considering only the phenotype (Gallo *et al.*, 2016). This technique has the ability to generate sequence reads of billions of DNA fragments, with an increased sequencing coverage. This makes NGS an accurate and high-throughput diagnostic technology, with a rapid turn-around time (Raje *et al.*, 2014). Other important advantage of this method is its cost-efficiency, that is mainly due to the increased number of DNA samples tested per sequencing run, contributing to economize reagents (Stoddard *et al.*, 2014).

Several previous studies assessed the efficacy of targeted NGS gene panels for primary immunodeficiencies diagnosis. Nijman *et al.* (2014) developed an NGS-based approach using a panel containing 170 PID-related genes. For validation, they tested 33 PID patients with known variants, being able to detect 21 of 22 single-nucleotide variants (SNVs) and 11 of 13 indels. This resulted in a success rate of 95% for detection of point mutations and 85% for detection of deletions/insertions. In order to implement this approach, 26 patients without a genetic diagnosis, despite routine functional and extensive testing, were evaluated, with a diagnosis achieved for 4 of them. Al-Mousa *et al.* (2016) tested the efficiency of an NGS panel containing 162 PID genes. For validation of the assay, they selected 122 PID patients with known mutations, achieving an overall sensitivity of 96% for detection of SNVs. One hundred and thirty-nine patients with suspected PIDs but genetically undiagnosed were also included in the study. The authors were able to detect the genetic defects in 35 of the 139 patients, yielding a clinical sensitivity of 25%. Considering the total of 261 patients, the overall diagnosis efficiency was 58%. Rae *et al.* (2018) performed an assay including 27 participants with a PID compatible phenotype but without a genetic diagnosis, in which they used an NGS panel with 242 PID-related genes. A monogenic cause was identified in 46% of the patients, and, importantly, the NGS results had treatment implications in 37% of the cohort. Similarly, numerous other studies also demonstrated the effectiveness of NGS technology for diagnosis of PID disorders (Moens *et al.*, 2014; Gallo *et al.*, 2016; Cifaldi *et al.*, 2019). In the last few years, these and other works have revealed the unprecedented potential of NGS to improve the effective identification of the underlying genetic cause of immune diseases, with reduced cost and time as compared to more conventional genetic testing.

1.4.2 Exome and genome sequencing

Next-generation sequencing has been implemented in many laboratories worldwide for PID diagnosis and NGS-based methods used as a first-tier molecular approach are emerging as an effective method for the diagnosis of PIDs.

Besides targeted NGS gene panels, current methodologies used for PID investigation include whole genome sequencing (WGS) and whole exome sequencing (WES). WGS sequences the entire genome, being the most comprehensive method. It allows identification of both exonic and non-coding variants, such as intronic, splice site and regulatory variants (Royer-Bertrand and Rivolta, 2015). WES sequences only the protein-coding regions of the genome, i.e., the exons. Since it is an unbiased genetic test, WES may not only identify previously described pathogenic variants, but also novel genes and/or variants implicated in PID forms, just like WGS (Heimall *et al.*, 2018). Therefore, WES and WGS may be important tools in identification of known or novel genetic variants, which is especially edifying when the variants are connected with non-traditional phenotypes.

However, the cost of these two methodologies is high and the amount of information yielded implies an increased complexity of the analysis, as well as time required for interpretation of variants (Al-Mousa *et al.*, 2016). Since targeted gene panels are focused on a smaller number of genes, their cost is much lower, with an increased read depth of the sequenced genes when compared to WES or WGS (Sims *et al.*, 2014). Thus, they generate less data, which facilitates the analysis and provides faster results.

Other disadvantage of WGS and WES resides on the possibility of identifying incidental genetic findings. Since these methodologies evaluate many other genes that are not related to the patient's clinical phenotype, the possibility of finding variants associated with increased risk of malignancy, cardiovascular and other diseases is increased (Heimall *et al.*, 2018). This results in several ethical and medical issues regarding the decision to report the findings or not. The use of gene panels minimizes this risk, as it is a more focused approach that assesses only genes possibly related to the patient's phenotype (Evans, 2013).

Taken together the aspects considered above, the use of targeted gene panels can be envisaged as a valuable first-line option for clinical investigation of PIDs, since it offers a good compromise between cost, complexity of data analysis and diagnostic yield.

1.4.3 Functional studies for novel variants

Given the rarity of PIDs, novel variants are frequently detected in PID patients, and the advent of NGS has enabled a great increment of the rate of their discovery. Regardless of the method employed to identify novel variants, it is fundamental to clarify their effect on the function of the encoded final protein and establish an association with the specific phenotypes. This requires functional studies and model systems to prove the pathogenicity of the variants of unknown significance (VUS) (Bisgin *et al.*, 2018). Many assays can be performed to establish the biological effect of variants detected. Flow cytometry or Western blot may be used to assess gene expression, if the variant is predicted to impair it (Sobh *et al.*, 2016). Evaluation of receptor activation may be done through phosphorylation of downstream proteins or upregulation of target gene expression (Chang *et al.*, 2012). When the gene plays an important role in lymphocyte activation, the variant effect can be assessed through evaluation of the expression of activation markers, lymphocyte proliferation assays after antigen stimulation and measurement of secretion of immunoglobulins, after B cell activation, and cytokines, after T cell activation (Tan *et al.*, 2015).

In addition to *in vitro* assays, animal models of PIDs may afford important insights to understand the effect of a genetic defect within the context of an *in vivo* immune system.

The most commonly used animal model are mice, given their rapid reproductive rate and the high homology between their immune system and that of humans (Masopust *et al.*, 2017). Mice were used to model combined immunodeficiency caused by a missense variant in transferrin receptor 1 (TfR1), a cell surface receptor important for erythrocyte development and function, that is encoded by *TFRC* (Jabara *et al.*, 2015). This variant disrupts TfR1 intracellular internalization, which results in defective T and B cell proliferation and impaired class-switching. Mutant mice recapitulated the immune defects observed in patients, confirming the pathogenicity of the variant (Jabara *et al.*, 2015).

Zebrafish are another model system frequently used for PID understanding, due to their rapid reproductive rate and development, and the existence of orthologs of the majority of human genes in zebrafish genome (Iwanami, 2014). Zebrafish have been used to model severe combined immunodeficiencies, caused by genes like *ZAP70* or *BCL11B* (Moore *et al.*, 2016; Punwani *et al.*, 2016). A loss-of-function variant in *ZAP70* was proven to be pathogenic by generating mutant zebrafish that showed a complete loss of mature T cells (Moore *et al.*, 2016). In another study, functional analysis of a missense variant in *BCL11B* using zebrafish embryos revealed blocked development of

T cell progenitors and developmental abnormalities present in the patient that were reproduced in the zebrafish (Punwani *et al.*, 2016).

However, a natural limitation inherent to these animal models are the genetic differences between these animals and humans. In order to overcome this problem, humanized mice, generated by transplanting human hematopoietic stem cells into immunodeficient mice, which then produce human immune cells *de novo*, have already been used to investigate PIDs (Martinez-Torres *et al.*, 2014; Jangalwe *et al.*, 2016).

1.5 Newborn Screening for Primary Immunodeficiencies

In order to include diseases in population-based screening programs, a series of criteria must be taken in consideration (Wilson and Jungner, 1968). Severe PID forms, such as SCIDs, are life-threatening and effective treatments are capable of improving patient outcomes. This is only possible if the diagnosis is achieved before the onset of symptoms and the presence of severe infections, with possible organ damage. Additionally, the natural history of SCIDs is well known and there is a cost-effective laboratory method for detection of these diseases in a pre-symptomatic phase (Dorsey and Puck, 2017), which will be discussed later. Therefore, PIDs are suitable candidates for inclusion in newborn screening (NBS) programs.

Diagnosing primary immunodeficiencies in neonates is complex. The immune system is not fully developed at birth, which complicates testing and interpretation of results (Walkovich and Connelly, 2016). Moreover, most PID affected neonates are asymptomatic at birth because maternal antibodies prevent infections in the first months of life. Nevertheless, rapid diagnosis of these disorders is essential, as previously mentioned. Thus, understanding the development of the neonatal immune system and its limitations, as well as knowing the PIDs that more frequently occur early in life and their features, is crucial to identify PID forms faster. Furthermore, there must be a suspicion of PID when there is a history of early or unexplained deaths in the family, multiple miscarriages, atypical infections, or immune dysregulation (Walkovich and Connelly, 2016).

Currently, newborn screening methodologies are used to detect PID forms that are characterized by T cell and B cell deficiencies. T cell receptor excision circles (TREC) are small circular fragments of DNA that are formed during T cell development and maturation in the thymus, more precisely during T cell receptor gene arrangement.

TRECs can be assessed by quantitative reverse transcription polymerase chain reaction (qRT-PCR) from dried blood spots on Guthrie cards (Kwan and Puck, 2015). Absence or marked reduction of TRECs is indicative of T cell maturation defects. Likewise, quantification of kappa-deleting recombination excision circles (KREC) is used to measure B cell production. KRECs are produced during arrangement of the variable, diversity and joining domains (V(D)J recombination) of the B cell immunoglobulin kappa gene, during B cell maturation (Nakagawa *et al.*, 2011). Inclusion of these assays in newborn screening has brought great advances in early PID diagnosis. Simultaneous screening of TREC and KREC is more advantageous than TREC quantification alone, since SCIDs may present with different levels of T and B cell deficiency and some disorders are characterized by decreased B cells only, as is the case of some antibody deficiency disorders (Borte *et al.*, 2012).

Despite the contribution of these tests to prompt neonatal PID diagnosis, they are limited to disorders affecting T and/or B cells and do not provide a definitive molecular diagnosis. Moreover, this approach is unable to identify cases in which T cells are dysfunctional but present in normal numbers (Kuo *et al.*, 2013). Thus, in the future it is likely that screening for PIDs will extend to PID diseases other than T and B cell disorders, like phagocyte defects and complement deficiencies (King *et al.*, 2017). Phagocyte disorders may be identified through analysis of proteins specific of these cells and complement proteins can be eluted from dried blood spot samples and identified in order to diagnose complement disorders (Hamsten *et al.*, 2015).

To confirm the results obtained with phenotypic tests, like TREC and KREC assays, and establish a definitive diagnosis, genetic testing is necessary. Also, genetic tests allow differentiation of the underlying cause of T and B cell deficiency, which has great importance in determining the appropriate therapeutic approach (Heimall *et al.*, 2018).

Since PIDs include a long list of described disorders, the most efficient approach for their detection is the use of next-generation sequencing technology (King *et al.*, 2017). Using a multigene panel to identify the genetic defect in a newborn early diagnosed with SCID, for example, concerning that there are multiple candidate genes for the disease cause, seems to be a feasible strategy. In the process of establishing an NGS approach for newborn screening of PIDs, selection of candidate genes to include in screening panels is of great importance. A panel including all the PID-related genes identified so far is a suitable option, however, this panels need to be regularly updated in order to include new genes that are continuously being identified (King *et al.*, 2017).

Several studies have demonstrated the applicability of NGS panels for newborn screening of PIDs. Muramatsu *et al.* (2018) performed a combination of TREC assay

and next-generation sequencing for newborn screening of SCID. The authors identified 48 newborns with TREC negative results. These infants were subjected to further examination, including NGS using a panel containing 349 genes related to PIDs. Four patients with PIDs were identified, showing that combination of TREC and NGS assays may help in achieving a rapid and an accurate PID diagnosis.

NGS will have an important role in NBS of PIDs, allowing for its expansion. Therefore, combined TREC/KREC assays and NGS testing is emerging as the most efficient and high-throughput way for a timely molecular diagnosis for affected infants (Yu *et al.*, 2016).

Extensive efforts are needed for developing a national newborn screening program that includes screening for PIDs (Abolhassani *et al.*, 2018). In order to implement this approach, a robust laboratory-clinical pipeline must be established, ensuring that the tests are sensitive and specific. Also, other challenges are the incorporation of screening technologies on newborn screening laboratories and follow-up of positive cases.

Screening for PIDs is now included in the newborn screening programs of some countries, such as United States of America, where all the states perform screening for SCID, The Netherlands, the first European country to include SCID in the NBS program, Norway, among others (Kwan *et al.*, 2014; Blom *et al.*, 2017; International Patient Organisation for Primary Immunodeficiencies - IPOPI, 2020). Other countries are still assessing pilot TREC screening programs, as is the case of the United Kingdom, France, Spain, and several others (Adams *et al.*, 2014; Audrain *et al.*, 2014; de Felipe *et al.*, 2016). In Portugal, recent developments in SCID neonatal screening have been followed up. Application of NBS for PIDs is in progress, with discussions being outgoing to include SCID in national NBS panel at early stage (International Patient Organisation for Primary Immunodeficiencies - IPOPI, 2020).

2. Aims

Considering the advantages of next-generation DNA sequencing, the main goal of this project was to validate and implement a cost-efficient approach for the diagnosis of primary immunodeficiencies, designing 3 NGS multigene panels that target most of the genes known to be associated with the different categories of PIDs, in an attempt to improve PID diagnostic tools and develop an efficient and rapid strategy for the identification of causative mutations in PID affected patients, that ultimately will provide the definitive genetic cause for their disease phenotypes.

Among these panels, the first aim of the present work was the validation of the one containing not only the highest number of genes, but also the genes known to be implied in PIDs with the most elevated severity and prevalence (Panel A).

The second aim was related to the evaluation of the clinical efficacy of the validated multigene panel by testing molecularly undiagnosed patients.

The third aim consisted in the development of a molecular approach for the study of additional PID-related genes that were not included in Panel A, in order to continue the study of the cases that remained uncharacterized by the application of the validated NGS approach.

This work was conducted at Molecular Genetics Unit of Centro de Genética Médica Doutor Jacinto Magalhães (CGMJM) of Centro Hospitalar Universitário do Porto (CHUP). PID diagnosis was not previously performed at this unit and, since CHUP is a National Reference Center for PIDs, the final result of this work will be the inclusion of this molecular approach in the routine diagnosis with direct implications in follow-up and treatment of these patients at CHUP.

3. Materials and Methods

3.1 NGS multigene panels design

PID phenotypic categories were divided in 3 different panels, according to the classification proposed by Bousfiha *et al.* (2018), in order to develop a cost-effective NGS approach. Three panels targeting a total of 284 genes involved in different PIDs were designed, since patients can be distinguishable by their immunophenotypic profile and therefore be referred for analysis with a specific panel. The categories studied in each panel are presented in Table 2. One of the PID categories, phenocopies of PIDs, was not included in this study, since they are mostly caused by somatic mutations and affected individuals are mosaics; hence, the disease-causing variants may not be detected using the NGS conditions used in this work.

Table 2. PID categories included in the three NGS multigene panels.

Panel A	Severe combined immunodeficiencies
	Combined immunodeficiencies less profound than SCID
	Combined immunodeficiencies with syndromic features
	Antibody deficiencies
Panel B	Congenital defects of phagocyte number or function
	Defects in intrinsic and innate immunity
	Complement deficiencies
Panel C	Diseases of immune dysregulation
	Autoinflammatory diseases

Panels were designed using Ion AmpliSeq™ Designer software (Thermo Fisher Scientific, Waltham, Massachusetts, USA). All genes previously associated within the same PID category and with available on-demand primers in the software, were grouped in the same panel. Characterization of each panel, including number of the genes studied, panel size, number of amplicons (total and in each pool of primers), amplicon range and *in silico* coverage is presented in Table 3. These metrics were determined and provided by the manufacturer for the Ion AmpliSeq™ panels (Thermo Fisher Scientific). Coverage metrics for every individual gene is presented in Table A1 of Appendix 1. The genes comprising Panel A and the associated phenotypes are presented in Table 4. The panels were designed in order to target the coding DNA sequence (CDS) and splice site regions of these genes and for detection of different types of variants (such as single nucleotide variants and small insertions/deletions).

Table 3. NGS multigene panels features.

	Panel A	Panel B	Panel C
Number of genes	117	99	68
Panel Size	519.571 kb	336.607 kb	289.148 kb
Number of Amplicons	2734 Pool 1: 1372 amplicons Pool 2: 1362 amplicons	1784 Pool 1: 891 amplicons Pool 2: 893 amplicons	1504 Pool 1: 749 amplicons Pool 2: 755 amplicons
Amplicon Range	125-275 bp	125-275 bp	125-275 bp
<i>In silico</i> Overall Coverage	99 %	99 %	99 %

3.2 DNA samples studied

The present work focused primarily on the validation and implementation of Panel A, considering disease severity, prevalence, and sample availability. Hence, in this study were only recruited DNA samples from patients clinically diagnosed with any of the PID forms for which the panel was designed (see Table 2).

Validation experiments were done through genetically double-blind assays, using 4 anonymized genomic DNA control samples from patients previously characterized at the molecular level (positive controls). These patients were selected by physicians from the cohort followed at CHUP and all present PID disorders that belong to the categories included in Panel A.

In order to assess the clinical utility of Panel A, for future implementation as a routine diagnostic test for PIDs at Molecular Genetics Unit of CGMJM of CHUP, 4 anonymized DNA samples from patients without a genetic diagnosis, followed in the pediatric-immunology consultation of CHUP, were tested using the validated panel. Clinical and laboratory criteria were considered by the physicians in order to select these patients for evaluation with Panel A (Table 5).

The present work was approved by the Ethics Committee of CHUP [internal reference 2019.256 (209-DEFI/219-CE)].

Table 4. Description of the genes included in Panel A and associated phenotypes.

PID category	Gene Symbol	Location	Phenotype¹	Inheritance²	Transcript
Immunodeficiencies affecting cellular and humoral immunity (a) Severe combined immunodeficiencies	<i>ADA</i>	20q13.12	Severe combined immunodeficiency due to adenosine deaminase deficiency	AR, SMO	NM_000022.2
	<i>AK2</i>	1p35.1	Reticular dysgenesis	AR	NM_001625.3
	<i>CD247</i>	1q24.2	Immunodeficiency 25	AR	NM_198053.2
	<i>CD3D</i>	11q23.3	Immunodeficiency 19	AR	NM_000732.4
	<i>CD3E</i>	11q23.3	Immunodeficiency 18	AR	NM_000733.3
	<i>CORO1A</i>	16p11.2	Immunodeficiency 8	AR	NM_007074.3
	<i>DCLRE1C</i>	10p13	Omenn syndrome; Severe combined immunodeficiency with sensitivity to ionizing radiation	AR	NM_022487.2
	<i>FOXN1</i>	17q11.2	T-cell immunodeficiency, congenital alopecia, and nail dystrophy	AR	NM_003593.2
	<i>IL2RG</i>	Xq13.1	Severe combined immunodeficiency	XLR	NM_000206.2
	<i>IL7R</i>	5p13.2	Severe combined immunodeficiency, T-cell negative, B-cell positive, natural killer cell positive	AR	NM_002185.3
	<i>JAK3</i>	19p13.11	Severe combined immunodeficiency, T-cell negative, B-cell positive, natural killer cell negative	AR	NM_000215.3
	<i>LIG4</i>	13q33.3	Severe combined immunodeficiency with sensitivity to ionizing radiation; LIG4 syndrome	AR	NM_002312.3
	<i>NHEJ1</i>	2q35	Severe combined immunodeficiency with microcephaly, growth retardation, and sensitivity to ionizing radiation	-	NM_024782.2
	<i>PRKDC</i>	8q11.21	Immunodeficiency 26, with or without neurologic abnormalities	AR	NM_001081640.1
	<i>PTPRC</i>	1q31.3-q32.1	Severe combined immunodeficiency, T-cell negative, B-cell positive, natural killer cell positive	AR	NM_002838.4
<i>RAG1</i>	11p12	Omenn syndrome; Alpha/beta T-cell lymphopenia with gamma/delta T-cell expansion, severe cytomegalovirus infection, and autoimmunity; Severe combined immunodeficiency, T-cell negative, B-cell negative, natural killer cell positive; Combined cellular and humoral immune defects with granulomas	AR	NM_000448.2	

	<i>RAG2</i>	11p12	Omenn syndrome; Combined cellular and humoral immune defects with granulomas; Severe combined immunodeficiency, T-cell negative, B-cell negative, natural killer cell positive	AR	NM_000536.2
Immunodeficiencies affecting cellular and humoral immunity	<i>B2M</i>	15q21.1	Immunodeficiency 43; Amyloidosis, familial visceral	AD, AR	NM_004048.2
	<i>BCL10</i>	1p22.3	Immunodeficiency 37	AR	NM_003921.4
(b) Combined immunodeficiencies generally less profound than SCID	<i>CARD11</i>	7p22.2	B-cell expansion with NFkB and T-cell anergy; Immunodeficiency 11A; Immunodeficiency 11B with atopic dermatitis	AD, AR	NM_032415.4
	<i>CD3G</i>	11q23.3	Immunodeficiency 17, CD3 gamma deficient	AR	NM_000073.2
	<i>CD40</i>	20q13.12	Immunodeficiency with hyper-IgM	AR	NM_001250.4
	<i>CD40LG</i>	Xq26.3	Immunodeficiency with hyper-IgM	XLR	NM_000074.2
	<i>CIITA</i>	16p13.13	Bare lymphocyte syndrome	AR	NM_000246.3
	<i>DOCK2</i>	5q35.1	Immunodeficiency 40	AR	NM_004946.2
	<i>DOCK8</i>	9p24.3	Hyper-IgE recurrent infection syndrome	AR	NM_001190458.1
	<i>ICOS</i>	2q33.2	Immunodeficiency, common variable, 1	AR	NM_012092.3
	<i>IKBKB</i>	8p11.21	Immunodeficiency 15A/15B	AD, AR	NM_001556.2
	<i>IL21</i>	4q27	Immunodeficiency, common variable, 11	AR	NM_021803.3
	<i>IL21R</i>	16p12.1	Immunodeficiency 56; Elevated level of IgE	AD, AR	NM_181078.2
	<i>LAT</i>	16p11.2	Immunodeficiency 52	AR	NM_001014987.1
	<i>LCK</i>	1p35.2	Immunodeficiency 22	AR	NM_005356.3
	<i>MAGT1</i>	Xq21.1	Immunodeficiency, with magnesium defect, Epstein-Barr virus infection and neoplasia	XLR	NM_032121.5
	<i>MALT1</i>	18q21.32	Immunodeficiency 12	AR	NM_006785.2
	<i>MAP3K14</i>	17q21.31	NIK deficiency, Primary immunodeficiency with multifaceted aberrant lymphoid immunity	-	NM_003954.3
	<i>MSN</i>	Xq12	Immunodeficiency 50	XLR	NM_002444.2
	<i>RELB</i>	19q13.32	Immunodeficiency 53	AR	NM_006509.3
	<i>RFX5</i>	1q21.3	Bare lymphocyte syndrome	AR	NM_000449.3
	<i>RFXANK</i>	19p13.11	MHC class II deficiency	AR	NM_003721.2
<i>RFXAP</i>	13q13.3	Bare lymphocyte syndrome	AR	NM_000538.3	
<i>RHOH</i>	4p14	T-cell immunodeficiency with epidermodysplasia verruciformis	AR	NM_004310.3	

	<i>STK4</i>	20q13.12	T-cell immunodeficiency syndrome, recurrent infections, autoimmunity, and cardiac malformations	-	NM_006282.2
	<i>TAP1</i>	6p21.32	Bare lymphocyte syndrome	AR	NM_000593.5
	<i>TFRC</i>	3q29	Immunodeficiency 46	AR	NM_001128148.1
	<i>TNFRSF4</i>	1p36.33	Immunodeficiency 16	AR	NM_003327.3
	<i>UNC119</i>	17q11.2	Immunodeficiency 13; Cone-rod dystrophy 2	AD	NM_005148.3
	<i>ZAP70</i>	2q11.2	Immunodeficiency 48; Autoimmune disease, multisystem, infantile-onset, 2	AR	NM_001079.3
Combined immunodeficiencies with associated or syndromic features	<i>ARPC1B</i>	7q22.1	Immunodeficiency 71 with inflammatory disease and congenital thrombocytopenia	AR	NM_005720.3
	<i>ATM</i>	11q22.3	Ataxia-Telangiectasia; Breast cancer	AD, AR	NM_000051.3
	<i>BLM</i>	15q26.1	Bloom syndrome	AR	NM_000057.2
	<i>CCBE1</i>	18q21.32	Hennekam lymphangiectasia-lymphedema syndrome 1	AR	NM_133459.3
	<i>CDCA7</i>	2q31.1	Immunodeficiency-centromeric instability-facial anomalies syndrome 3	AR	NM_031942.4
	<i>DKC1</i>	Xq28	Hoyeraal-Hreidarsson syndrome; Dyskeratosis congenita	XLR	NM_001363.3
	<i>DNMT3B</i>	20q11.21	Immunodeficiency-centromeric instability-facial anomalies syndrome 1	AR	NM_006892.3
	<i>EPG5</i>	18q12.3-q21.1	Vici syndrome	AR	NM_020964.2
	<i>ERCC6L2</i>	9q22.32	Bone marrow failure syndrome 2	AR	NM_001010895.2
	<i>EXTL3</i>	8p21.1	Immunoskeletal dysplasia with neurodevelopmental abnormalities	AR	NM_001440.2
	<i>GINS1</i>	20p11.21	Immunodeficiency 55	AR	NM_021067.3
	<i>HELLS</i>	10q23.33	Immunodeficiency-centromeric instability-facial anomalies syndrome 4	AR	NM_018063.3
	<i>LIG1</i>	19q13.33	Ligase I deficiency; Immunodeficiency and predisposition to malignancies; Bloom syndrome-like	-	NM_000234.1
	<i>LYST</i>	1q42.3	Chediak-Higashi syndrome	AR	NM_000081.3
	<i>MCM4</i>	8q11.21	Immunodeficiency 54	AR	NM_005914.3
	<i>MTHFD1</i>	14q23.3	Combined immunodeficiency and megaloblastic anemia with or without hyperhomocysteinemia	AR	NM_005956.3
	<i>MYSM1</i>	1p32.1	Bone marrow failure syndrome 4	AR	NM_001085487.2
<i>NBN</i>	8q21.3	Nijmegen breakage syndrome; Leukemia, acute lymphoblastic; Aplastic anemia	AR	NM_002485.4	

<i>NFKBIA</i>	14q13.2	Ectodermal dysplasia and immunodeficiency 2	AD	NM_020529.2
<i>NOP10</i>	15q14	Dyskeratosis congenita	AR	NM_018648.3
<i>PARN</i>	16p13.12	Dyskeratosis congenita; Pulmonary fibrosis and/or bone marrow failure, telomere-related, 4	AD, AR	NM_002582.3
<i>PGM3</i>	6q14.1	Immunodeficiency 23	AR	NM_015599.2
<i>PMS2</i>	7p22.1	Mismatch repair cancer syndrome; Colorectal cancer, hereditary nonpolyposis, type 4	AR	NM_000535.5
<i>PNP</i>	14q11.2	Immunodeficiency due to purine nucleoside phosphorylase deficiency	AR	NM_000270.3
<i>POLE</i>	12q24.33	Colorectal cancer; Facial dysmorphism, immunodeficiency, livedo, and short stature syndrome (FILS syndrome)	AD, AR	NM_006231.2
<i>POLE2</i>	14q21.3	Polymerase ε subunit 2 deficiency; Colorectal cancer; Holoprosencephaly 2; Bardet-Biedl syndrome 1	-	NM_002692.3
<i>RBCK1</i>	20p13	Polyglucosan body myopathy 1 with or without immunodeficiency	AR	NM_031229.2
<i>RNF168</i>	3q29	RIDDLE syndrome	AR	NM_152617.3
<i>RNF31</i>	14q12	HOIP and LUBAC deficiency	-	NM_017999.4
<i>RTEL1</i>	20q13.33	Pulmonary fibrosis and/or bone marrow failure, telomere-related, 3; Dyskeratosis congenita	AD, AR	NM_016434.3
<i>SAMD9</i>	7q21.2	MIRAGE syndrome; Tumoral calcinosis, familial, normophosphatemic	AD, AR	NM_017654.3
<i>SAMD9L</i>	7q21.2	Ataxia-pancytopenia syndrome	AD	NM_152703.2
<i>SLC46A1</i>	17q11.2	Folate malabsorption, hereditary	AR	NM_080669.4
<i>SMARCAL1</i>	2q35	Schimke immunoosseous dysplasia	AR	NM_014140.3
<i>SP110</i>	2q37.1	Hepatic venoocclusive disease with immunodeficiency	AR	NM_080424.2
<i>SPINK5</i>	5q32	Netherton syndrome	AR	NM_001127698.1
<i>STAT3</i>	17q21.2	Hyper-IgE recurrent infection syndrome; Autoimmune disease, multisystem, infantile onset	AD	NM_139276.2
<i>STAT5B</i>	17q21.2	Growth hormone insensitivity with immunodeficiency	AD, AR	NM_012448.3
<i>STIM1</i>	11p15.4	Immunodeficiency 10; Stormorken syndrome; Myopathy, tubular aggregate, 1	AD, AR	NM_001277961.1
<i>TCN2</i>	22q12.2	Transcobalamin II deficiency	AR	NM_000355.3
<i>TERT</i>	5p15.33	Dyskeratosis congenita; Pulmonary fibrosis and/or bone marrow failure, telomere-related, 1	AD, AR	NM_198253.2

	<i>TINF2</i>	14q12	Revesz syndrome; Dyskeratosis congenita	AD	NM_001099274.1
	<i>TTC7A</i>	2p21	Gastrointestinal defects and immunodeficiency syndrome	AR	NM_020458.2
	<i>WAS</i>	Xp11.23	Neutropenia, severe congenital; Thrombocytopenia; Wiskott-Aldrich syndrome	XLR	NM_000377.2
	<i>WIPF1</i>	2q31.1	Wiskott-Aldrich syndrome 2	AR	NM_003387.4
	<i>ZBTB24</i>	6q21	Immunodeficiency-centromeric instability-facial anomalies 2	AR	NM_014797.2
Predominantly Antibody Deficiencies	<i>AICDA</i>	12p13.31	Immunodeficiency with hyper-IgM, type 2	AR	NM_020661.2
	<i>BLNK</i>	10q24.1	Agammaglobulinemia 4	AR	NM_013314.3
	<i>BTK</i>	Xq22.1	Agammaglobulinemia; Isolated growth hormone deficiency, type III, with agammaglobulinemia	XLR	NM_000061.2
	<i>CD19</i>	16p11.2	Immunodeficiency, common variable, 3	AR	NM_001178098.1
	<i>CD79A</i>	19q13.2	Agammaglobulinemia 3	AR	NM_001783.3
	<i>CD79B</i>	17q23.3	Agammaglobulinemia 6	AR	NM_000626.2
	<i>CD81</i>	11p15.5	Immunodeficiency, common variable, 6	AR	NM_004356.3
	<i>CR2</i>	1q32.2	Immunodeficiency, common variable, 7	AR	NM_001877.4
	<i>IGLL1</i>	22q11.23	Agammaglobulinemia 2	AR	NM_020070.2
	<i>IKZF1</i>	7p12.2	Immunodeficiency, common variable, 13	AD	NM_006060.5
	<i>INO80</i>	15q15.1	INO80 deficiency; Hepatosplenic t-cell lymphoma; Primary microcephaly; Fibromuscular dysplasia	-	NM_017553.1
	<i>MOGS</i>	2p13.1	Congenital disorder of glycosylation, type IIb	AR	NM_006302.2
	<i>MS4A1</i>	11q12.2	Immunodeficiency, common variable, 5	AR	NM_021950.3
	<i>MSH6</i>	2p16.3	Colorectal cancer, hereditary nonpolyposis, type 5; Mismatch repair cancer syndrome	AD, AR	NM_000179.2
	<i>NFKB1</i>	4q24	Immunodeficiency, common variable, 12	AD	NM_001165412.1
	<i>NFKB2</i>	10q24.32	Immunodeficiency, common variable, 10	AD	NM_001077494.2
	<i>PIK3CD</i>	1p36.22	Immunodeficiency 14	AD	NM_005026.3
	<i>PIK3R1</i>	5q13.1	Agammaglobulinemia 7; Immunodeficiency 36; SHORT syndrome	AD, AR	NM_181523.2
	<i>SKIV2L</i>	6p21.33	Trichohepatoenteric syndrome 2	AR	NM_006929.4
<i>TCF3</i>	19p13.3	Agammaglobulinemia 8	AD	NM_003200.3	

	<i>TNFRSF13B</i>	17p11.2	Immunodeficiency, common variable, 2; Immunoglobulin A deficiency 2	AD, AR	NM_012452.2
	<i>TNFRSF13C</i>	22q13.2	Immunodeficiency, common variable, 4	AR	NM_052945.3
	<i>TRNT1</i>	3p26.2	Retinitis pigmentosa and erythrocytic microcytosis; Sideroblastic anemia with B-cell immunodeficiency, periodic fevers, and developmental delay	AR	NM_182916.2
	<i>TTC37</i>	5q15	Trichohepatoenteric syndrome 1	AR	NM_014639.3

^{1,2} Extracted from OMIM® (Online Mendelian Inheritance in Man®) database (<https://omim.org/>); AR: autosomal recessive; AD: autosomal dominant; XLR: X-linked recessive; SMO: somatic mosaicism.

Table 5. Clinical data of the studied patients.

Case	Clinical information
P1	Hypogammaglobulinemia and epileptic encephalopathy
P2	Hypogammaglobulinemia and trichothiodystrophy
P3	Humoral immunodeficiency with recurrent pneumonia; peculiar facies; bronchiectasis
P4	Common variable immunodeficiency (CVID)

3.3 Panel A validation

3.3.1 DNA preparation and quantification

DNA samples integrity was verified by electrophoresis on 1% agarose gel (w/v) in 1x TAE (Tris 0.04 M, acetate 0.02 M, EDTA 1 M pH 8.0) for 30 minutes at 140 V. DNA quantification was made using NanoDrop® ND-1000 spectrophotometer (Thermo Fisher Scientific). Aliquots of DNA were prepared at a concentration of approximately 10 ng/μl.

Amplifiable genomic DNA was quantified by quantitative real-time PCR (qPCR) on a 7500 Fast Real-Time PCR System (Life Technologies, Carlsbad, California, USA) using TaqMan® RNase P Detection Reagents Kit (Applied Biosystems, Foster City, California, USA). A 1:20 dilution of each DNA sample (10 ng/μl) was prepared. Seven serial dilutions of a genomic DNA standard of known concentration were prepared, in order to generate a standard curve, as follows: 5 ng/μl, 2.5 ng/μl, 1.25 ng/μl, 0.625 ng/μl, 0.3125 ng/μl, 0.15625 ng/μl and 0.078125 ng/μl. The assays also included a Negative Template Control (NTC). For each sample/standard, duplicate qPCR reactions were done and consisted of 10 μl of 2x TaqMan® Universal PCR Master Mix, 1 μl of 20x RNase P Primer-Probe mix, 6.5 μl of nuclease-free water and 2.5 μl of DNA (10 ng/μl), making a total volume of 20 μl and then divided into two reactions. Reaction conditions used were those described by the manufacturer's protocol (Table 6). A standard curve plotting C_T vs. concentration (ng/μl) of input standard DNA was generated and the DNA amount for each sample was calculated using the analysis software provided with the equipment.

Table 6. Reaction conditions for DNA quantification by qPCR.

Stage	Temperature	Time
Hold	50°C	2 min
Hold	95°C	10 min
40 cycles	95°C	15 sec
	60°C	1 min

3.3.2 DNA library preparation and next-generation sequencing

DNA libraries were prepared using Ion AmpliSeq™ Library Kit 2.0 (Thermo Fisher Scientific) according to the manufacturer’s instructions (Figure 1).

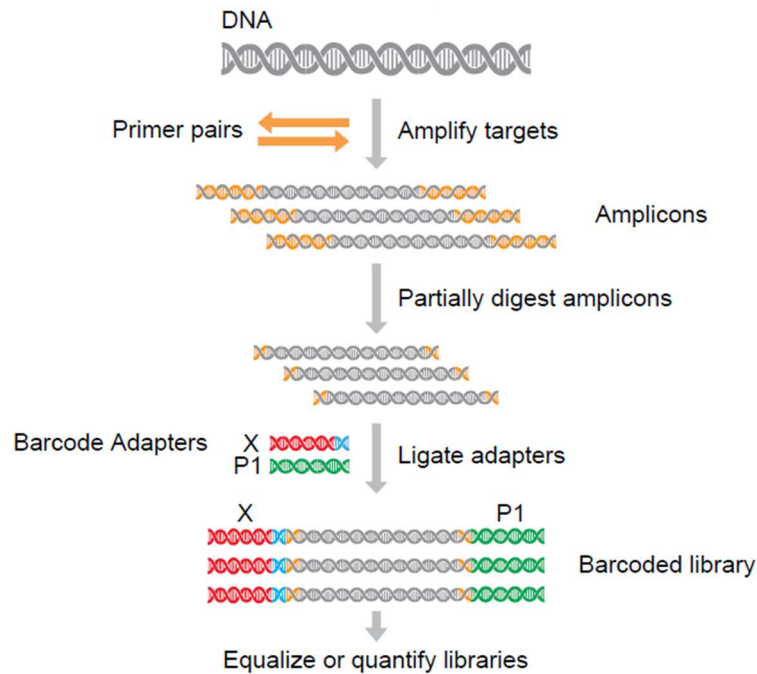


Figure 1. Schematic representation of Ion AmpliSeq™ workflow (Thermo Fisher Scientific).

Targets were amplified by PCR using the two independent primer pools designed for Panel A. First, DNA was diluted to a concentration of 3.5 ng/μl and, for each sample, a mixture was prepared containing 4.5 μl of 5x Ion AmpliSeq™ HiFi Mix, 5.4 μl of DNA (3.5 ng/μl) and 1.35 μl of nuclease-free water. Five microliters of this mixture were added to 5 μl of 2x primer pool 1 and the same was done for primer pool 2. Amplification conditions are presented in Table 7.

Table 7. Reaction conditions for targets amplification.

Stage	Temperature	Time
Hold	99°C	2 minutes
14 cycles	99°C	15 seconds
	60°C	8 minutes
Hold	10°C	Hold

After amplification, the two reaction's products were combined, resulting in a total volume of 20 µl, which was then partially digested for barcode ligation. For that, 2 µl of FuPa Reagent were added to each amplified sample and the mixture was loaded in the thermal cycler and the conditions used are described in Table 8.

Table 8. Reaction conditions for partial digestion of amplicons.

Temperature	Time
50°C	10 minutes
55°C	10 minutes
60°C	20 minutes
10°C	Hold (for up to 1 hour)

In order to identify each sample during the sequencing run, a different barcode adapter was ligated to each DNA library. A barcode adapter mix was prepared using 0.5 µl of Ion P1 Adapter, 0.5 µl of Ion Xpress™ Barcode X (X – number of the barcode chosen) and 1 µl of nuclease-free water. To perform the ligation reaction, 2 µl of the previous mix, 4 µl of Switch Solution and 2 µl of DNA Ligase were added to each DNA library. Reaction conditions are presented in Table 9.

Table 9. Conditions of the ligation reaction.

Temperature	Time
22°C	30 minutes
68°C	5 minutes
72°C	5 minutes
10°C	Hold (for up to 24 hours)

Subsequently, the libraries were purified. For that, 45 µl of AMPure™ XP Reagent (Beckman Coulter Life Sciences, Indianapolis, USA) were added to each library and the mixture was incubated for 5 minutes at room temperature. The tubes were placed in DynaMag™-96 Side Magnet and incubated for 2 minutes. The supernatant was discarded and 150 µl of freshly prepared 70% (v/v) ethanol were added to the pellet of beads. The tubes were moved side-to-side in the two positions of the magnet to wash the beads. The supernatant was discarded and the previous step was repeated for a second wash. With the tubes in the magnet, the beads were air-dried at room temperature for 5 minutes, to ensure that all ethanol droplets were removed.

The unique barcode libraries were then eluted in 40 µl of Low TE to disperse the beads, vortexed, incubated at room temperature for 2 minutes and placed on the magnet

for another 2 minutes to pellet the beads. The supernatant was removed and a 1:200 dilution was prepared for each library.

The purified libraries were quantified by qPCR with the Ion Library TaqMan[®] Quantitation Kit (Life Technologies). Three 10-fold serial dilutions of *E. coli* DH10B Control Library were prepared as follows: 6.8 pM, 0.68 pM and 0.068 pM. Each reaction was done in duplicate and, for each sample, a mix was prepared containing 20 µl of 2x Ion Library qPCR Master Mix, 2 µl of 20x Ion Library TaqMan[®] Quantitation Assay and 18 µl of the diluted Ion AmpliSeq[™] library or control dilution. This reaction mix was then divided in two wells of a PCR plate, for a total reaction volume of 20 µl. The assay included a negative control. qPCR was performed on a 7500 Fast Real-Time PCR System (Life Technologies), using the conditions described by the manufacturer's protocol (Table 10). Diluted library quantification was calculated using the software provided with the equipment.

Table 10. Conditions for library quantification by qPCR.

Stage	Temperature	Time
Hold	50°C	2 min
Hold	95°C	2 min
40 cycles	95°C	15 sec
	60°C	1 min

These unique barcode libraries were then diluted to 80 pM and combined to be sequenced on a single chip at a similar depth, as recommended by the manufacturer. Ion Chef[™] Instrument (Thermo Fisher Scientific) was used for template preparation and chip loading using Ion 510[™] & Ion 520[™] & Ion 530[™] Kit – Chef (Thermo Fisher Scientific). Template preparation included the clonal amplification of final library pools (by emulsion PCR), the recovery and the enrichment of template-positive Ion Sphere[™] Particles (ISPs). The loaded Ion 520 chips were then subjected to sequencing on an Ion GeneStudio[™] S5 System (Thermo Fisher Scientific).

3.3.3 Data analysis and variant interpretation

Sequence reads were aligned in the Ion Torrent Server 5.12 (Thermo Fisher Scientific) against the human genome assembly GRCh37/hg19. Torrent Suite Software 5.12.1 was used to obtain the basic run metrics (coverage of target regions and quality parameters) and to download the resulting Binary Alignment Map (BAM) file. This

software also generated the Variant Caller Format (VCF) file using Variant Caller plugin 5.12 and the predefined parameter “germline-low stringency”.

VCF file from each sample was uploaded to Ion Reporter 5.14 Software and variants were annotated using the “annotate variants single sample” workflow. Filter chains were applied to remove variants with a minor allele frequency (MAF) above 1% and present on the University of California Santa Cruz (UCSC) Genome Browser (<http://genome.ucsc.edu/>) common single-nucleotide polymorphism (SNP) list.

Alamut Visual™ software program version 2.11 (Interactive Biosoftware, Rouen, France) was used for analysis and interpretation of filtered-in variants. This software includes algorithms to predict functional impact of missense and splicing variants (such as PolyPhen-2, SIFT, MutationTaster, and Human Splicing Finder) and gives access to genetic variants databases (such as ClinVar, dbSNP, gnomAD, ESP, Uniprot). Other public databases consisting in collections of variants associated to human genetic diseases were also consulted, namely, Human Gene Mutation Database (HGMD®; <https://hgmd.org>) and Leiden Open Variation Database (LOVD®; <https://lovd.nl>).

Candidate variants were manually inspected on the BAM file using Alamut Visual™ Software. Prioritization was given to variants located in coding and splice-site regions, due to their likely higher impact on production of functional proteins. If these variants could not explain the phenotype, non-coding variants were subsequently analyzed. Detected variants were classified according to the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) guidelines (Richards *et al.*, 2015) as pathogenic, likely pathogenic, benign, likely benign or of uncertain significance. Variants were described according to the recommendations of Human Genome Variation Society (HGVS) (den Dunnen *et al.*, 2016).

The experimental coverage of the panel was ascertained using the data and metrics obtained with positive controls testing. Target regions with less than 20x read depth were considered not covered by the assay and selected to be screened by Sanger sequencing (as described in Section 3.5).

3.4 Panel A implementation

Once validated, the clinical efficacy of the NGS Panel A was evaluated by testing DNA samples of PID patients genetically uncharacterized. DNA and library preparation,

sequencing, data analysis and variant interpretation were performed as described above (Section 3.3).

3.5 Development of a conventional molecular approach for the study of other PID-related genes

To complement the study of PIDs, a molecular approach was developed for the analysis of additional PID-related genes that were not included in Panel A, because on-demand primers were unavailable in Ion AmpliSeq™ Designer software. By this way, primers were designed to study genes involved in severe forms of PID, such as CIDs less profound than SCID and CIDs with syndromic features (*BCL11B*, *CD8A*, *RMRP*, *TAP2*, *TAPBP*, *TERC*, and *TRAC*).

3.5.1 Primer design for PCR

Primers for conventional Sanger sequencing were designed using Primer Express 3.0 (Applied Biosystems). Prediction of the occurrence of possible dimers and/or hairpins, and the specificity of the primers in relation to the region of interest were evaluated using FastPCR version 3.7.7 (PrimerDigital, Helsinki, Finland) and Blast (NIH, Bethesda MD, USA), respectively. Alamut Visual™ software was used to ensure that the primers were not designed in regions with frequent variants.

M13 universal tails were incorporated in the sequence of most of the primers designed (forward - F and reverse - R) in order to facilitate the sequencing PCR reactions using the same unique M13 primers (F or R).

The sequence of each pair of primers and the expected size of the amplified product for each region of interest are presented in Tables A2 and A3 of Appendix 2.

3.5.2 Symmetric PCR

Amplification of the regions of interest was performed under different conditions according to the characteristics of the fragment to be amplified. The “standard” reaction consisted of a mixture of 10 µl of 2x PCR Master Mix (Promega, Madison, Wisconsin, USA), 1 µl of primer F (10 pmol/µl), 1 µl of primer R (10 pmol/µl), 7 µl of H₂O and 1 µl of DNA (100 ng/µl). For regions with “high G/C” content, the reaction consisted of 12.5 µl

of 2x Master Mix, 1.5 μl of DMSO, 5 μl of betaine 5 M, 1 μl of primer F (10 pmol/ μl), 1 μl of primer R (10 pmol/ μl), 3 μl of H₂O and 1 μl of DNA (250 ng/ μl). The conditions of the programs used to amplify the different regions of interest are presented in Figure 2. The programs for “standard” or “high G/C” amplifications differed in the number of cycles (38 and 40, respectively). To confirm the occurrence of amplification, PCR products were then visualized on a 1% (w/v) agarose gel (electrophoresis at 140 V for 30 minutes).

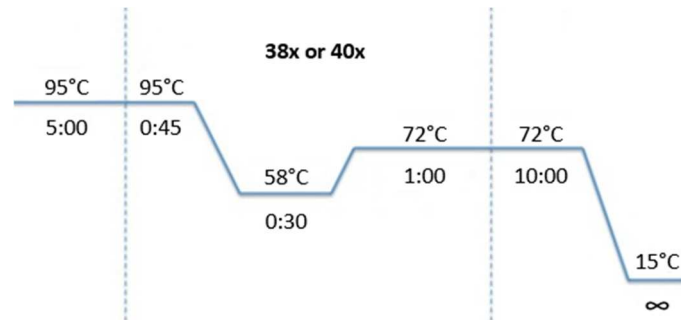


Figure 2. Conditions of the programs used for amplification of regions of interest.

3.5.3 Enzymatic purification of PCR products

After amplification, PCR products were purified by adding 1 μl of illustra™ ExoProStar™ 1-Step (GE Healthcare LifeSciences, Buckinghamshire, UK), a mixture of two enzymes that remove unused primers and nucleotides of the previous PCR reaction, to 5 μl of each product. The mixture was subsequently incubated in a thermal cycler for 30 minutes at 37 °C and 15 minutes at 80 °C.

3.5.4 Sequencing PCR

Sequencing PCR was performed using the commercial kit BigDye™ Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems). The reaction mixture consists of MgCl₂, fluorochrome-labeled ddNTP, dNTP, AmpliTaq Polymerase and buffer. In each reaction were used 6 μl of the enzymatic purified product, 2 μl of BigDye™ Terminator v.3.1 DNA Sequencing Mix and 2 μl of primer M13 – F or R (5 pmol/ μl) or 0.7 μl of primer F or R (10 pmol/ μl) used in symmetric PCR (see Tables A2 and A3 of Appendix 2). For “high G/C” fragments, 1.5 μl of betaine 5 M were also added to the sequencing mixture. Thermal cycler program used is presented in Figure 3.

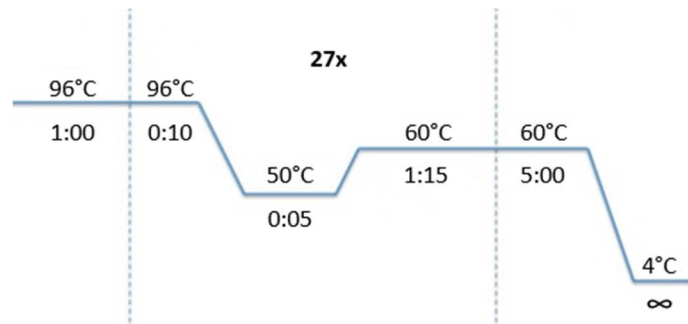


Figure 3. Conditions used in sequencing reactions.

Subsequently, purification of the products of the sequencing reaction was carried out, to eliminate dye interference, enzymes and dNTPs not used in the former reaction. For that, DyeEx 96 Plates with gel filtration system (QIAGEN, Hilden, Germany) were used. The samples were placed in the center of the resin columns in each well of the plate, and the plate was centrifuged for 3 minutes at 1000 g. Purified products were then dried in an appropriate thermal cyclers. To prepare the samples for subsequent sequencing run, 15 μ l of HiDi formamide (Applied Biosystems) were added to each dried sample. Sequencing runs were performed by capillary electrophoresis on an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems), the data obtained analyzed using SeqScape[®] V2.5 software (Applied Biosystems) and variants analyzed and interpreted using Alamut Visual[™] software (as mentioned before; Section 3.3.3).

4. Results

4.1 Panel A validation using positive control DNA samples

4.1.1 Gene coverage (predicted and experimental)

Overall *in silico* predicted coverage for Panel A was 99% of the coding regions of all genes included in the panel, according to the metrics data provided by the manufacturer.

After running the 4 control DNA samples, the observed average target base coverage at 20x was 99,04%, the average base coverage depth was 315,65x, and the average uniformity of base coverage was 97,60%. *In silico* and experimental coverage for every individual gene is presented in Table A1 of Appendix 1.

4.1.2 Variants detected in positive control patients

Targeted next-generation sequencing was performed using genomic DNA from 4 control patients with known causal variants for validation of Panel A. A bioinformatic pipeline was then created and tested for NGS data analysis (sequences alignment, variant calling and annotation, data visualization and variant interpretation).

After variant calling and annotation, an average of 436 variants per control was detected. In order to prioritize variants by frequency and thus reduce the number of variants, a filter was applied to consider only rare (minor allele frequency < 1%) exonic, intronic, splice site and UTR variants. After application of this filter, an average of 56 variants per control were filtered-in and subjected to further analysis with Alamut Visual software and other genetic databases (Section 3.3.3).

The NGS sequencing metrics obtained for the positive controls, including coverage data and number and type of variants, is presented in Table A4 of Appendix 3.

This approach was able to correctly detect the disease-causing variants previously identified in the 4 control samples.

Control 1 (C1) was a male hemizygous for a deletion of 4 bp (c.1581_1584del) in exon 16 of Bruton's tyrosine kinase (*BTK*) gene (Figure 4). This is a frameshift variant that results in substitution of a cysteine residue for a tryptophan in position 527, creating a premature stop codon (p.Cys527Trpfs*2). *BTK* gene is located at Xq22.1 and encodes cytoplasmic tyrosine kinase. BTK protein is expressed throughout differentiation of B cells and myeloid lineage cells, but is not present in T cells and plasma cells (Smith *et al.*, 1994). This protein is involved in signal transduction and has a great importance in regulation of B cell proliferation, differentiation, and activation (Fiorini *et al.*, 2004).

Defects in *BTK* gene are related to X-linked Agammaglobulinemia (XLA; MIM #300755), included in antibody deficiencies category. XLA is characterized by a markedly decreased number of B cells and profound hypogammaglobulinemia, i.e., very reduced serum levels of all immunoglobulin isotypes. Affected patients present recurrent bacterial infections caused mainly by encapsulated pathogens, like *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Pseudomonas*, and *Staphylococcus* (Conley and Rohrer, 1995). The most common symptoms include recurrent conjunctivitis, otitis, sinusitis and dermatitis. This variant detected in this control in *BTK* gene using Panel A has been previously reported in association with X-linked agammaglobulinemia (Conley *et al.*, 1994; Ozturk *et al.*, 2013; Singh *et al.*, 2016).

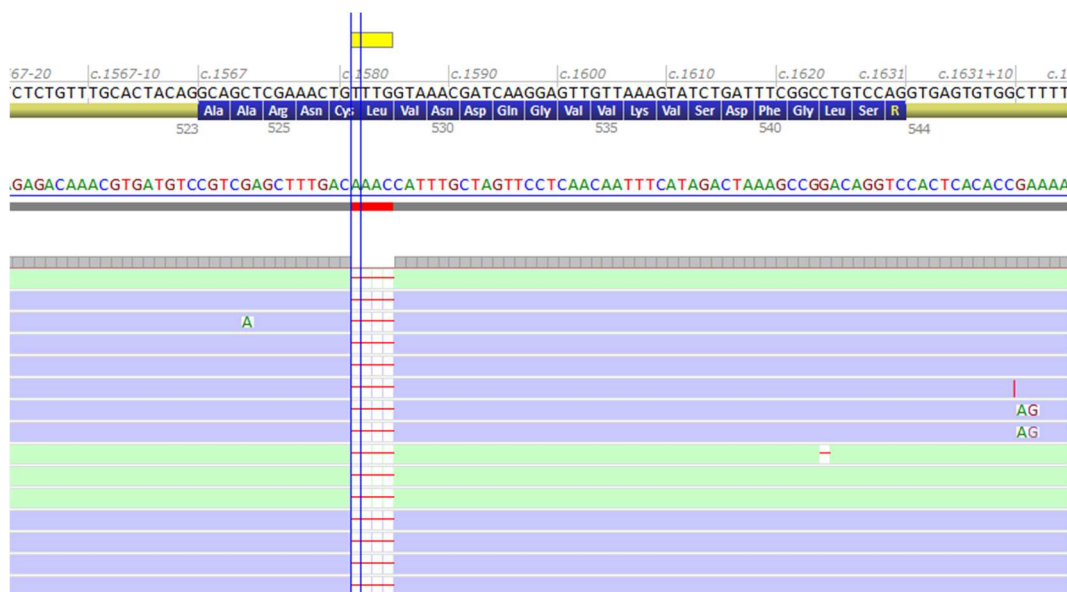


Figure 4. Binary alignment map (BAM) file showing the c.1581_1584del variant identified in hemizyosity in exon 16 of the *BTK* gene in control patient 1, using Alamut Visual software. Reference sequence (NM_00061.2) shown in the upper bar.

Control 2 (C2) had a homozygous duplication of 1 bp (c.4801dup) in exon 33 of ataxia-telangiectasia mutated (*ATM*) gene (Figure 5). This variant causes the substitution of a serine residue for a lysine in position 1601 and alters the reading frame, creating a premature stop codon (p.Ser1601Lysfs*4), which results in the production of a truncated protein. *ATM* gene, located at 11q22.3, encodes a serine/threonine kinase enzyme with the same name, which has a fundamental role in cell cycle checkpoint control and in coordination of cellular signaling pathways after DNA double strand breaks caused either by endogenous sources or oxidative stress (Ditch and Paull, 2012). Thus,

ATM is essential for preservation of genomic stability. Variants in *ATM* gene result in Ataxia-telangiectasia (MIM #208900), an autosomal recessive disease belonging to CID with syndromic features category. This disorder is characterized by progressive cerebellar ataxia, oculocutaneous telangiectasias, immunological defects, increased cancer predisposition, particularly of lymphoid origin, and increased sensitivity to radiation (Rothblum-Oviatt *et al.*, 2016). Many patients present premature aging of the skin and hair and growth factor deficiency (Schubert *et al.*, 2005). Regarding the immunological abnormalities, the most common are deficiencies in one or more isotypes of immunoglobulins, and lymphopenia, particularly affecting T cells. Low levels of new B and T cells leaving the bone marrow and the thymus, respectively, are also hallmarks of the disease (Kraus *et al.*, 2014). All these features can be explained by the absence of ATM protein. In order to generate clonal diversity, lymphocytes undergo gene rearrangements throughout their development. This is a process that causes DNA double strand breaks, and their repair is diffculted by the absence of a functional ATM protein (Bredemeyer *et al.*, 2006), resulting in a decreased number of lymphocytes and compromised function of these cells. The variant identified in this patient has not been previously reported in association with ataxia-telangiectasia.



Figure 5. Binary alignment map (BAM) file showing the c.4081dup variant identified in homozygosity in exon 33 of the *ATM* gene in control patient 2, using Alamut Visual software. Reference sequence (NM_001351834.1) shown in the upper bar.

In control 3 (C3), a homozygous variant (c.10747G>A) in exon 48 of lysosomal trafficking regulator (*LYST*) gene was identified (Figure 6). This is a missense variant

that results in substitution of a glycine residue for an arginine in position 3583 (p.Gly3583Arg). *LYST* gene, which is located at 1q42.3, encodes a protein with the same name that plays a role in lysosomal trafficking. It has been suggested that LYST protein is required for regulation of membrane fusion events and for sorting endosomal proteins into late multivesicular endosomes, although its precise function remains unknown (Faigle *et al.*, 1998; Möhlig *et al.*, 2007). This gene is implicated in Chediak-Higashi syndrome (CHS; MIM #214500), a rare autosomal recessive disease. CHS is characterized by hypopigmentation of the eyes, skin and hair, peripheral neuropathy, mild bleeding tendency, and immunodeficiency, such as abnormal NK cell function (Arulappan *et al.*, 2018). While this leads to recurrent bacterial infections mostly caused by *Staphylococcus* and *Streptococcus* species, viral and fungal infections are also common (Introne *et al.*, 1999). The classic clue to the diagnosis of CHS patients is the presence of enlarged lysosomes or lysosome-related organelles in many cell types. About 85% to 90% of the affected subjects develop hemophagocytic lymphohistiocytosis, which occurs in the so called accelerated phase, a fatal hyperinflammatory condition that affects multiple organs (Toro *et al.*, 2009). The *LYST* variant here identified was predicted to be damaging by SIFT, MutationTaster and Polyphen-2 algorithms. In addition, it has been previously reported in a homozygous patient with Chediak-Higashi syndrome (Antunes *et al.*, 2013).

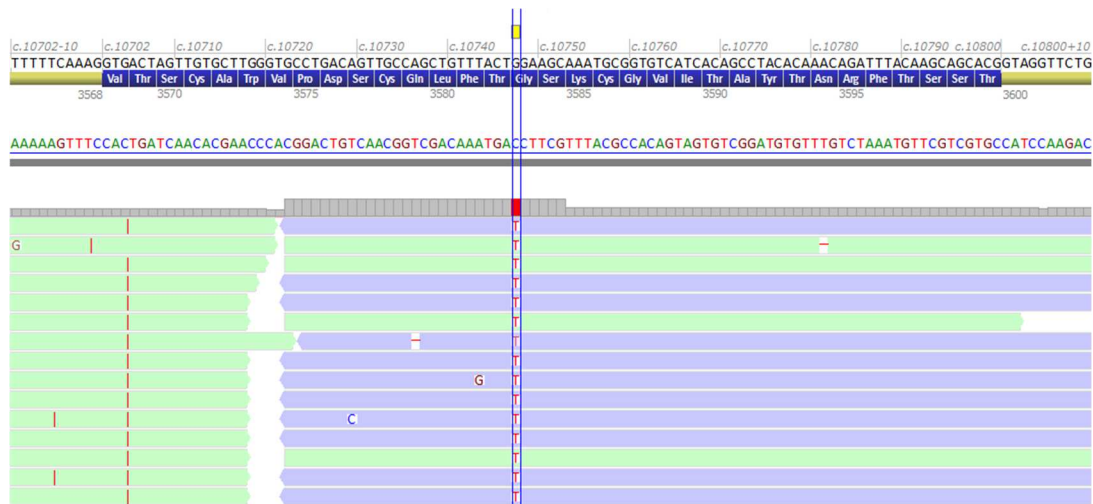


Figure 6. Binary alignment map (BAM) file showing the c.10747G>A variant identified in homozygosity in exon 48 of the *LYST* gene in control patient 3, using Alamut Visual software. Reference sequence (NM_000081.3) shown in the upper bar.

In control 4 (C4), a heterozygous variant (c.353G>A) was identified in exon 3 of interleukin 7 receptor (*IL7R*) gene (Figure 7). This missense variant causes the substitution of a cysteine residue for a tyrosine in position 118 (p.Cys118Tyr). *IL7R*, located at 5p13.2, encodes IL7R α , the α chain of the heterodimeric interleukin 7 receptor protein. This protein is essential for T cell development in the thymus and for proliferation and regulation of peripheral T cells (Huang and Luther, 2012). Variants in *IL7R* gene are associated with severe combined immunodeficiency (T cell negative, B cell positive, NK cell positive; MIM #608971). Patients with SCID present deficiencies in development or function of T and B cells, resulting in failure of the cellular and humoral immune system (Kalman *et al.*, 2004). Therefore, they usually present opportunistic or severe recurrent infections, as well as failure to thrive. Other symptoms include rashes, poor wound healing, and persistent diarrhea (IUIS Scientific Committee, 1999). B cell positive forms of SCID, as is the case of SCID caused by variants in *IL7R*, usually have normal numbers of B cells, but they are nonfunctional. This variant detected in this control in *IL7R* gene has been previously reported in association with SCID (Giliani *et al.*, 2005; Gallego-Bustos *et al.*, 2016). Since the identified variant in control 4 alone cannot explain the patient phenotype, very likely he must be a compound heterozygous with two variants in *IL7R*, one of which escaped detection with our approach. This is in line with previously performed studies, where some SCID patients were reported to be compound heterozygous for variants in *IL7R*, namely involving large deletions or variants affecting splicing in one of the mutated *IL7R* alleles (Bayer *et al.*, 2014; Gallego-Bustos *et al.*, 2016).

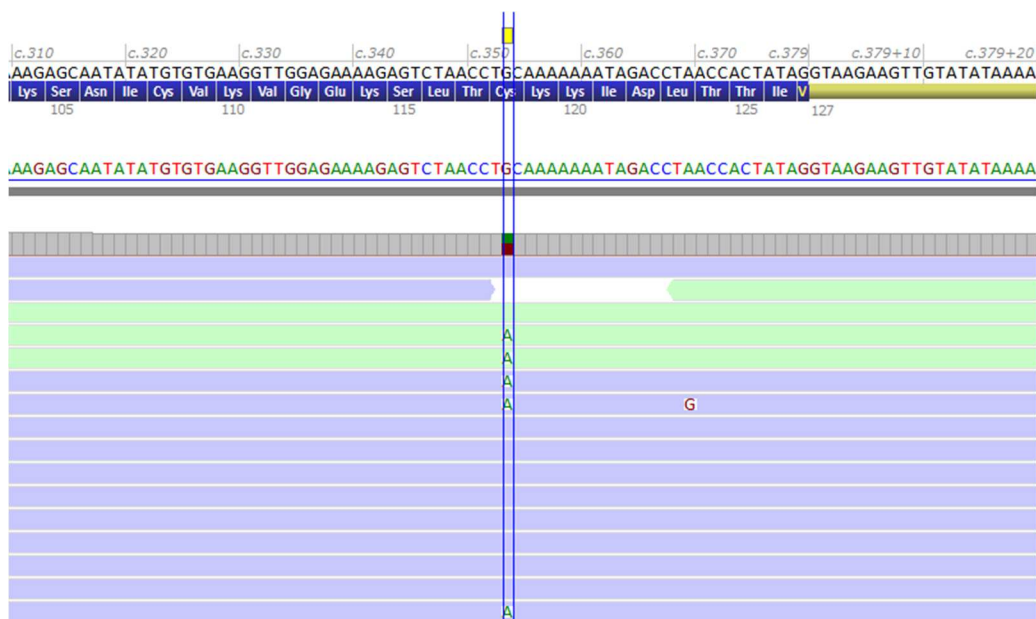


Figure 7. Binary alignment map (BAM) file showing the c.353G>A variant identified in heterozygosity in exon 3 of the *IL7R* gene in control patient 4, using Alamut Visual software. Reference sequence (NM_002185.4) shown in the upper bar.

4.2 Panel A implementation

After validation of Panel A analyzing 4 positive control samples, implementation of Panel A in PIDs diagnosis was performed through evaluation of 4 patient samples without a molecular diagnosis. As in the validation process and after conclusion of the NGS sequencing runs, variants were prioritized based on frequency, to identify among the non-polymorphic variants those with potential to be causal, resulting in an average of 52 variants per patient. The sequencing metrics obtained for each patient samples, including coverage data and number and type of variants, is presented in Table A4 of Appendix 3.

Analysis of the filtered-in variants did not include any variant previously reported or that could be classified as pathogenic or likely pathogenic. Variants of uncertain significance (VUS) were identified in all 4 patients, according to the ACMG/AMP guidelines (Richards *et al.*, 2015), and their description is presented in Table 11. Among these are 1 small in-frame deletion in *TTC37* gene (P2) and 2 missense variants, located in *JAK3* (P2) and *LYST* (P4) genes, which bioinformatic analysis is summarized in Table 12. The remaining correspond to variants located in intronic and regulatory regions for which splicing algorithms did not predict a significant impact.

Table 11. Description of the variants of uncertain significance identified in the 4 patients.

P1						
Gene	gDNA variant	Effect on protein	dbSNP reference	Zygoty	Allele Frequency (gnomAD)	ACMG criteria for classification
<i>PIK3CD</i>	c.1690-44C>A	p.?	-	Het.	-	PM2 and BP4
<i>RBCK1</i>	c.167+33G>C	p.?	rs145976650	Het.	0.12%	PM2 and BP4
<i>RTEL1</i>	c.477+13G>C	p.?	rs764646296	Het.	0.0016%	PM2 and BP4
<i>STIM1</i>	c.139+28G>A	p.?	rs199693618	Het.	0.23%	PM2 and BP4
<i>ZAP70</i>	c.1737-3C>T	p.?	rs56249179	Het.	0.086%	PM2 and BP4
P2						
Gene	gDNA variant	Effect on protein	dbSNP reference	Zygoty	Allele Frequency (gnomAD)	ACMG criteria for classification
<i>ARPC1B</i>	c.990-37A>C	p.?	rs201953746	Het.	0.14%	PM2 and BP4
<i>CR2</i>	c.58+53T>C	p.?	rs368839023	Het.	0.41%	PM2 and BP4
<i>HELLS</i>	c.-56C>T	p.?	rs191227493	Het.	0.42%	PM2 and BP4
<i>JAK3</i>	c.2152G>C	p.Val718Leu	rs146837396	Het.	0.065%	PM1, PM2, PP2 and BP4
<i>PTPRC</i>	c.1660-53_1660-50dup	p.?	rs889941644	Het.	0.17%	PM2 and BP4

<i>SPINK5</i>	c.1888-94C>A	p.?	rs531536632	Het.	0.33%	PM2 and BP4
<i>TRNT1</i>	c.1056+54A>G	p.?	rs750676939	Het.	0.0064%	PM2 and BP4
<i>TTC37</i>	c.2353_2361del	p.Asn785 Tyr787del	rs537964611	Het.	0.00040%	PM2, PM4 and PP3
P3						
Gene	gDNA variant	Effect on protein	dbSNP reference	Zygoty	Allele Frequency (gnomAD)	ACMG criteria for classification
<i>ARPC1B</i>	c.64+20C>T	p.?	rs116514404	Het.	0.11%	PM2 and BP4
<i>DKC1</i>	c.1155+162C>T	p.?	-	Het.	0.016%	BP4
<i>EXTL3</i>	c.-74C>T	p.?	rs542929533	Het.	0.067%	PM2 and BP4
<i>MSN</i>	c.192+99G>A	p.?	rs944426117	Het.	0.014%	PM2 and BP4
<i>PIK3CD</i>	c.1021-61G>T	p.?	rs566665752	Het.	-	PM2 and BP4
<i>PRKDC</i>	c.3464+4G>A	p.?	rs758498140	Het.	0.00099%	PM2 and BP4
P4						
Gene	gDNA variant	Effect on protein	dbSNP reference	Zygoty	Allele Frequency (gnomAD)	ACMG criteria for classification
<i>ATM</i>	c.7630-125_7630-122dup	p.?	rs376424710	Het.	0.74%	PM2 and BP4
<i>DCLRE1C</i>	c.678+69A>T	p.?	rs41297034	Het.	0.46%	PM2 and BP4
<i>INO80</i>	c.4238-24C>T	p.?	rs1313642023	Het.	0.0014%	PM2 and BP4
<i>JAK3</i>	c.2805+11del	p.?	-	Hom.	-	PM2 and BP4
<i>LYST</i>	c.4115C>T	p.Thr1372lle	rs1172596943	Het.	-	PM2
<i>PARN</i>	c.1006-78T>G	p.?	rs199652820	Het.	0.26%	PM2 and BP4
<i>PGM3</i>	c.1114-186G>A	p.?	rs1004172753	Het.	-	PM2 and BP4
<i>POLE2</i>	c.1498-53T>A	p.?	-	Het.	-	PM2 and BP4
<i>POLE2</i>	c.1498-56del	p.?	rs1387323059	Het.	-	PM2 and BP4
<i>NOP10</i>	c.-65C>T	p.?	rs550497412	Het.	0.032%	PM2 and BP4

Het.: Heterozygosity; Hom.: Homozygosity; PM1, PM2: moderate pathogenic criterion; PP2, PP3: supporting pathogenic criterion; BP4: supporting benign criterion.

The results obtained with NGS sequencing were complemented with conventional sequencing of the regions that were sub-optimally covered by Panel A and for which primers were designed. Only variants classified as benign or likely benign were identified. The description of the variants detected in the 4 patients is presented in Table A5 of Appendix 4.

Table 12. Bioinformatic analysis for the identified missense variants.

Patient	Gene	gDNA variant	Effect on Protein	Bioinformatic analysis						
				Nucleotide conservation	Amino acid conservation	Grantham dist.	SIFT	Mutation Taster	PolyPhen-2	CADD
P2	<i>JAK3</i>	c.2152G>C	p.Val718Leu	Weak	Weak	32 (Small)	Tolerated	Disease-causing	0.026 (Benign)	21.2
P4	<i>LYST</i>	c.4115C>T	p.Thr1372Ile	Weak	Moderate	89 (Moderate)	Tolerated	Disease-causing	0.835 (Possibly pathogenic)	25.2

4.3 Development of a conventional molecular approach to complement the study of PID regions/genes

The primers designed to sequence the regions presenting insufficient coverage with Panel A and to sequence other PID genes were tested with a control gDNA, using the commercial mix PCR Master Mix (Promega, Madison, USA) and standard conditions (Figure 2). Figure 8 shows the results obtained after separation by agarose gel electrophoresis for some of the tested primers, as an illustration of the obtained outcomes.

The regions of interest that did not amplify or originated unspecific products using the initial conditions were submitted to a new test using “high-GC” conditions (Figure 2). Figure 9 shows the obtained results after agarose gel electrophoresis for some of these regions.

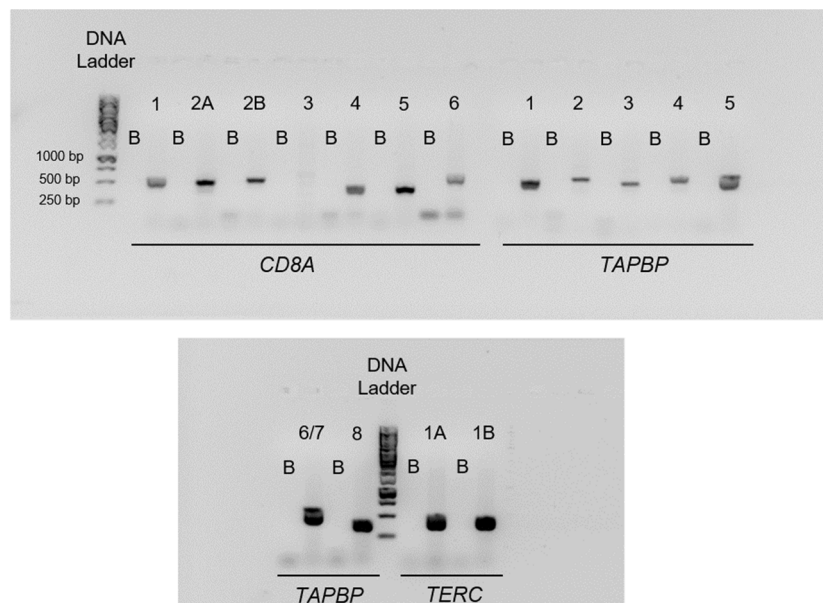


Figure 8. Electrophoretic results obtained in the first amplification test of PID genes using a control DNA and standard conditions. For each primer pair, a reaction in the absence of DNA was performed (B). The numbers represent different regions of interest amplified for each gene. DNA Ladder – GeneRuler™ 1 kb DNA Ladder (Thermo Fisher Scientific).

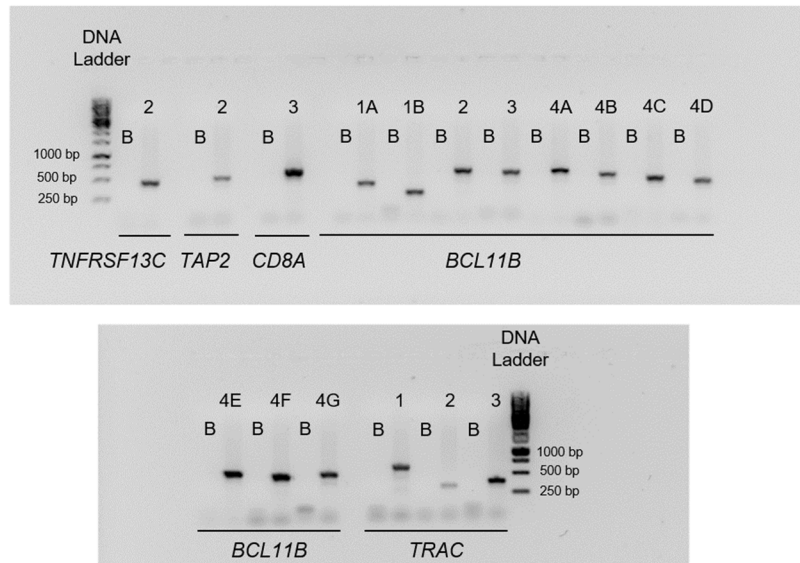


Figure 9. Electrophoretic results obtained in the second amplification test of PID genes using a control DNA and high-GC conditions. For each primer pair, a reaction in the absence of DNA was performed (B). The numbers represent different regions of interest amplified for each gene. DNA Ladder – GeneRuler™ 1 kb DNA Ladder (Thermo Fisher Scientific).

The sub-optimally covered regions targeted by Sanger sequencing for all patient DNA samples included *BTK* exon 14, *CD81* exons 1 and 3, *DKC1* exon 11, *DOCK2* exon 33, *DOCK8* exon 1, *EPG5* exons 25, 26, 27 and 39, *LYST* exon 42, *MALT1* exon 3, *MOGS* exon 1, *MS4A1* exon 8, *MSH6* exons 4 and 9, *NFKB1* exon 10, *NFKB2* exons 7 and 8, *PIK3CD* exons 12 and 13, *POLE* exons 24 and 26, *SLC46A1* exon 1, *TCF3* exons 10 and 19, *TERT* exon 1, *TFRC* exon 13, *TNFRSF13C* exon 2, *TNFRSF4* exons 2 and 4, *TTC37* exon 27, and *WAS* exon 10. As previously mentioned, no pathogenic variants were detected in the tested patients using this conventional approach.

The additional PID genes were not studied for the 4 patients since their phenotypes are not usually associated with those genes.

5. Discussion

Next-generation sequencing has revolutionized the field of PIDs, not only by allowing a more efficient genetic diagnosis of these disorders, but also by increasing the knowledge of PIDs. Over the years, novel genes related to immune disorders have been continuously discovered, thus revealing new and important roles for those genes, mechanisms of disease pathogenesis and new targets for therapies (Tangye *et al.*, 2020).

Since CHUP is a National Reference Center for PIDs, this project intended to develop an NGS-based approach for PID diagnosis, currently not offered in the routine molecular diagnostic setting.

For that purpose, 3 NGS multigene panels were designed, targeting genes described as associated with PIDs (Bousfiha *et al.*, 2018). This work focused on the validation and implementation of one panel containing 117 genes related to SCID, CID and antibody deficiencies.

In the panel validation stage, 4 control patients with known variants were tested. The generated data was initially used to evaluate the performance of the customized library, originating an overall coverage of 315,65x with 99,04% of the target bases covered at least 20x. These results demonstrate high quality of data since almost all target gene regions are covered at a high read depth. Even large genes, like *DOCK2*, *LYST*, *SPINK5* or *ATM*, showed an overall high coverage of the target coding regions.

Only 31 out of the 2734 amplicons (of 23 genes) consistently showed insufficient coverage, such as *CD81* (exons 1 and 3), *DOCK2* (exon 33) and *NFKB1* (exon 10), also reflected in the values calculated for experimental target base coverage for each gene (Table A1 of Appendix 1). In all of them, the sequence complexity, high GC content and the presence of repetitive sequences can explain the low coverage observed. Another reason for target regions not covered is high genomic homology, as observed in two regions absent in Panel A design: *STAT5B* exon 7, which has 99% homology with the same exon and adjacent introns of *STAT5A* gene, and *CORO1A* exon 12, which is affected by segmental duplication with > 98% homology. These results are in agreement with the performance of NGS panels reported by other studies (Valencia *et al.*, 2013; Moens *et al.*, 2014; Nijman *et al.*, 2014). The depth of sequence coverage is essential for a correct detection of genomic variants, so exons poorly covered, especially the ones containing mutational hotspots, must be complemented by conventional sequencing approaches. Viewing that, primers were designed for the exons showing insufficient or no coverage with Panel A, and the regions were then screened through Sanger

sequencing, allowing thus to fill the gaps experienced with NGS. For the reason mentioned above, this was only not possible for exon 7 of *STAT5B* gene and exon 12 of *CORO1A* gene, since the specificity of the primers could not be guaranteed due to genomic homology.

The study of positive control samples using the developed NGS approach resulted in the detection of all the pathogenic variants previously identified, demonstrating the high efficacy and sensitivity of Panel A.

Regarding the 4 patients without a molecular diagnosis, no variants classified as pathogenic/likely pathogenic were identified. Most of the identified VUS were heterozygous and were excluded as the disease cause because the associated disorders are autosomal recessive, with exception of 2 intronic variants in *PIK3CD* gene (P1, P3) which inheritance is autosomal dominant, and 2 intronic variants in *DKC1* gene (P3) and *MSN* gene (P3), which inheritance is X-linked recessive. However, the phenotypes associated with *DKC1* and *MSN* do not correlate to the phenotype observed in P3. The phenotype associated with *PIK3CD* is similar to those presented by P1 and P3, since variants in this gene are associated with B cell immunodeficiency, and hypogammaglobulinemia was described in some patients. Nevertheless, bioinformatic algorithms did not predict an impact on splicing for any of the 2 variants. Additionally, population data contributed to the exclusion of some variants, since they were found in homozygosity in healthy individuals, therefore proving their non-pathogenicity. These included the variants identified in *RBCK1* (P1), *ARPC1B* (P2, P3), *CR2* (P2), *DKC1* (P3), and *MSN* (P3) genes. Therefore, among the variants of uncertain significance, none was a strong candidate for the pathogenicity based on one or more of the following criteria: the phenotype described for the gene does not correlate to the patient's phenotype; the variant was predicted as non-damaging by bioinformatic tools such as SIFT, MutationTaster and/or PolyPhen-2; allele/population frequency data was not consistent with the classification of pathogenic variant; the zygosity of the variant(s) was not compatible with the gene/disease inheritance pattern. Therefore, the validated NGS approach did not succeed to reveal any genetic defect that could explain the phenotypes observed in these patients.

The low diagnostic yield obtained may be related to the low number of patients studied until now. The clinical utility of this panel will be better assessed as more patients are screened. Other studies showed satisfactory diagnostic yields in PID diagnosis using targeted NGS panels, with the percentage of patients for which a genetic diagnosis was achieved ranging from 25% to 48% (Moens *et al.*, 2014; Al-Mousa *et al.*, 2016; Bisgin *et al.*, 2018; Rae *et al.*, 2018).

The inability to detect the genetic defect in these 4 patients may be related to the absence of the causative gene in the NGS panel tested, either because the specific primers were not available for inclusion in the panel at the time of design (even having been already identified to cause PIDs), or because it is a “new” gene, not yet described as a PID cause. Despite the increased number of genes that have been associated with PIDs in the last years, especially since the introduction of new sequencing technologies, many forms of PIDs still do not have a defined underlying genetic defect. One example is Common variable immunodeficiency disorders (CVID), the most frequent symptomatic primary immune defect in adults, where the genetic basis of the majority of patients remains unknown (less than 20% genetically characterized) (de Valles-Ibáñez *et al.*, 2018).

One of the drawbacks inherent to targeted NGS gene panels is the relatively reduced number of genes included, especially considering the complexity and heterogeneity of PIDs. This requires the regular update of these panels to keep up with the rapid evolution of PID knowledge. Even so, out of a total of 146 genes associated to the PID categories of Panel A reported by Bousfiha *et al.* (2018), only 29 genes were not included in the panel. Other NGS-based approaches not restricted to a set of target genes, like whole exome sequencing or whole genome sequencing, may be useful in the identification of novel genes and more suitable to address a wider range of clinical phenotypes.

Additionally, this methodology may not accurately detect complex structural variants, other type of genetic defects associated to PIDs (Zhang *et al.*, 2009), such as large deletions/duplications, inversions, translocations or insertion of repetitive elements. A reliable detection of this type of variants using NGS panels is challenging, due to the generation of short reads (Clark *et al.*, 2011). Other techniques may be used to complement NGS applications, such as Microarray-based Comparative Genomic Hybridization (aCGH), MLPA (Multiplex Ligation-dependent Probe Amplification) or even whole genome sequencing.

Since these targeted-gene panels are mainly focused on coding sequences, the total of variants in regulatory, promotor and deep intronic regions cannot be analyzed. To assess the presence of this type of variants, a more comprehensive molecular approach using mRNA analysis and/or whole genome sequencing can be useful.

Generally, an efficient approach for PIDs diagnosis may be to first use targeted NGS gene panels based on the patient's phenotype to resolve most cases and then proceed to whole exome or genome sequencing if the cases remain unexplained (Moens *et al.*, 2014).

Despite the challenges in identification of disease-causing genetic defects using NGS gene panels, there are several advantages that make NGS a good option for application in the clinical setting, also demonstrated by the present work, like screening of several candidate genes in a single test with higher read depth, and multi-sample testing with reduced turnaround time. This study showed that NGS is an accurate and reliable screening tool for PIDs diagnosis which can lead to a prompt diagnosis and allow the initiation of the proper treatment on time. Therefore, targeted NGS may be a great option as a first-tier genetic approach for diagnosis of PID patients.

6. Final Conclusions and Future Perspectives

Genetic testing of PIDs has profound implications for patients and their families, with improved prognosis and decreased disease-associated mortality. Next-generation sequencing has been used for diagnosis of several Mendelian disorders with great success, and PID diagnosis is no exception.

However, in this study it was not possible to identify the genetic cause in the patients assessed with Panel A. A next step to achieve the molecular diagnosis in these patients could be the examination through Sanger sequencing of specific candidate genes that were not included in the panel, based on the patient's phenotype. For example, *IGHM* is a gene associated with agammaglobulinemia 1 and may be a candidate gene for patients P1 and P2 who present hypogammaglobulinemia. Similarly, *IRF2BP2*, a gene associated with common variable immunodeficiency, may also be studied for patient P4, since he presents this form of phenotype. Patient P3 has humoral immunodeficiency, so genes included in the antibody deficiencies category may be further analyzed, such as *ATP6AP1*, *IGKC*, or *UNG*, for example. Furthermore, additional testing tools can be used in an attempt to detect other type of genetic variants not detected by the NGS approach used in this study, like potential CNVs that may be detected by aCGH or MLPA.

Another alternative for further study the negative cases is the use of WES, especially considering the large number of novel genes included in the most recent PID classification (Bousfiha *et al.*, 2020), making the study of multiple candidate genes by Sanger sequencing time-consuming.

As previously mentioned, when candidate VUS are identified, a more detailed analysis should be performed in order to support or exclude pathogenicity, including study of familial co-segregation of the variant with disease, analysis of mRNA by cDNA sequencing in case of intronic variants, to assess changes in splicing mechanism, and functional studies using different methodologies (such as Western blot or flow cytometry), in an attempt to evaluate protein expression.

In the future, similarly to what was done for Panel A, Panels B and C, which target other PID categories like congenital defects of phagocyte, defects in intrinsic and innate immunity, diseases of immune dysregulation, among others, will also be tested with anonymous DNA samples with known variants and their clinical efficiency will then be evaluated with patients lacking a genetic diagnosis.

It is very important to update the designed panels regularly in order to include novel genes that are continuously being identified. In the near future, the panels designed for this study must be renewed with the genes that in the meantime were described, since

the design of the panels was based in a classification that reports 320 gene defects related to PIDs (Bousfiha *et al.*, 2018), and the most recent classification reports 430 different gene defects (Bousfiha *et al.*, 2020).

After validation and implementation of the 3 PID panels, there is the purpose to test the NGS panels sensitivity in other fluids than blood, namely in saliva. The clinical applicability of the developed approach will be assessed in PID patients that were submitted to hematopoietic stem cell transplant before having a molecular diagnosis.

In conclusion, this work allowed to successfully validate and implement a new molecular approach for diagnosis of PIDs, using one NGS multigene panel targeting several genes associated with SCID, CID and antibody deficiencies. With the future validation and implementation of Panels B and C, this may be a promising tool to improve PID patients lives.

7. Bibliography

- Abbas, A., Lichtman, A. H. and Pillai, S. (2017) *Cellular and Molecular Immunology 9th Ed., Cellular and Molecular Immunology*.
- Abolhassani, H., Chou, J., Bainter, W., *et al.* (2018) 'Clinical, immunologic, and genetic spectrum of 696 patients with combined immunodeficiency', *Journal of Allergy and Clinical Immunology*, 141(4), pp. 1450–1458.
- Adams, S. P., Rashid, S., Premachandra, T., *et al.* (2014) 'Screening of neonatal UK dried blood spots using a duplex TREC screening assay', *Journal of Clinical Immunology*, 34(3), pp. 323–330.
- Al-Mousa, H., Abouelhoda, M., Monies, D. M., *et al.* (2016) 'Unbiased targeted next-generation sequencing molecular approach for primary immunodeficiency diseases', *Journal of Allergy and Clinical Immunology*. Elsevier Ltd, 137(6), pp. 1780–1787.
- Antunes, H., Pereira, Â. and Cunha, I. (2013) 'Chediak-Higashi syndrome: Pathognomonic feature', *The Lancet*, 382(9903), p. 1514.
- Arulappan, J. Thomas, D. S., Wali, Y. A., Jayapal, S. K. and Venkatasalu, M. R. (2018) 'A child with Chediak-Higashi syndrome-a case study', *Current Pediatric Research*, 22(1), pp. 69–72.
- Audemard-Verger, A., Descloux, E., Ponard, D., *et al.* (2016) 'Infections revealing complement deficiency in adults: A French nationwide study enrolling 41 patients', *Medicine (United States)*, 95(19), pp. 1–4.
- Audrain, M., Thomas, C., Mirallie, S., *et al.* (2014) 'Evaluation of the T-cell receptor excision circle assay performances for severe combined immunodeficiency neonatal screening on Guthrie cards in a French single centre study', *Clinical Immunology*, 150(2), pp. 137–139.
- Babushok, D. V. and Bessler, M. (2015) 'Genetic predisposition syndromes: When should they be considered in the work-up of MDS?', *Best Practice and Research: Clinical Haematology*. Elsevier Ltd, 28(1), pp. 55–68.
- Bayer, D. K., Martinez, C. A., Sorte, H. S., *et al.* (2014) 'Vaccine-associated varicella and rubella infections in severe combined immunodeficiency with isolated CD4 lymphocytopenia and mutations in *IL7R* detected by tandem whole exome

- sequencing and chromosomal microarray', *Clinical and Experimental Immunology*, 178(3), pp. 459–469.
- Bisgin, A., Boga, I., Yilmaz, M., Bingol, G. and Altintas, D. (2018) 'The Utility of Next-Generation Sequencing for Primary Immunodeficiency Disorders: Experience from a Clinical Diagnostic Laboratory', *BioMed Research International*, 2018.
- Blom, M., Pico-Knijnenburg, I., Sijne-van Veen, M., *et al.* (2017) 'An evaluation of the TREC assay with regard to the integration of SCID screening into the Dutch newborn screening program', *Clinical Immunology*. Elsevier Inc, 180, pp. 106–110.
- Borte, S., Von Döbeln, U., Fasth, A., *et al.* (2012) 'Neonatal screening for severe primary immunodeficiency diseases using high-throughput triplex real-time PCR', *Blood*, 119(11), pp. 2552–2555.
- Bousfiha, A., Jeddane, L., Picard, C., *et al.* (2018) 'The 2017 IUIS Phenotypic Classification for Primary Immunodeficiencies', *Journal of Clinical Immunology*, 38(1), pp. 129–143.
- Bousfiha, A., Jeddane, L., Picard, C., *et al.* (2020) 'Human Inborn Errors of Immunity: 2019 Update of the IUIS Phenotypical Classification', *Journal of Clinical Immunology*.
- Bredemeyer, A. L., Sharma, G. G., Huang, C. Y., *et al.* (2006) 'ATM stabilizes DNA double-strand-break complexes during V(D)J recombination', *Nature*, 442(7101), pp. 466–470.
- Chang, J. J., Lacas, A., Lindsay, R. J., *et al.* (2012) 'Differential regulation of toll-like receptor pathways in acute and chronic HIV-1 infection', *Aids*, 26(5), pp. 533–541.
- Ciccarelli, F., Martinis, M. and Ginaldi, L. (2013) 'An Update on Autoinflammatory Diseases', *Current Medicinal Chemistry*, 21(3), pp. 261–269.
- Cifaldi, C., Brigida, I., Barzaghi, F., *et al.* (2019) 'Targeted NGS platforms for genetic screening and gene discovery in primary immunodeficiencies', *Frontiers in Immunology*, 10(APR).
- Clark, M. J., Chen, R., Lam, H. Y. K., *et al.* (2011) 'Performance comparison of exome DNA sequencing technologies', *Nature Biotechnology*. Nature Publishing Group, 29(10), pp. 908–916.
- Committee, I. S. (1999) 'Primary immunodeficiency diseases. Report of an IUIS Scientific Committee. International Union of Immunological Societies.', *Clinical and*

experimental immunology, 118 Suppl, pp. 1–28.

Conley, M. E., Fitch-hilgenberg, M. E., Cleveland, J. L., Parolini, O. and Rohrer, J. (1994) 'Screening of genomic DNA to identify mutations in the gene for bruton's tyrosine kinase', *Human Molecular Genetics*, 3(10), pp. 1751–1756.

Conley, M. E. and Rohrer, J. (1995) 'The spectrum of mutations in *Btk* that cause X-linked agammaglobulinemia', *Clinical Immunology and Immunopathology*, 76(3), pp. S192–S197.

Ditch, S. and Paull, T. T. (2012) 'The ATM protein kinase and cellular redox signaling: beyond the DNA damage response The protein kinase ataxia-telangiectasia mutated (ATM)', *Trends Biochem Sci*, 37(1), pp. 15–22.

Dorsey, M. and Puck, J. (2017) 'Newborn screening for severe combined immunodeficiency in the US: Current status and approach to management', *International Journal of Neonatal Screening*, 3(2).

Dowdell, K. C., Niemela, J. E., Price, S., *et al.* (2010) 'Somatic *FAS* mutations are common in patients with genetically undefined autoimmune lymphoproliferative syndrome', *Blood*, 115(25), pp. 5164–5169.

Duarte Ferreira, R., Silva, S., Carrapatoso, I., *et al.* (2018) 'Letter to the editor: Primary immunodeficiencies in adults - Multicentric cooperation to characterize the portuguese reality', *Acta Medica Portuguesa*, pp. 231–232.

den Dunnen, J. T., Dalgleish, R., Maglott, D. R., *et al.* (2016) 'HGVS Recommendations for the Description of Sequence Variants: 2016 Update', *Human Mutation*, 37(6), pp. 564–569.

Evans, B. J. (2013) 'Minimizing liability risks under the ACMG recommendations for reporting incidental findings in clinical exome and genome sequencing', *Genetics in Medicine*.

Faigle, W., Raposo, G., Tenza, D., *et al.* (1998) 'Deficient peptide loading and MHC class II endosomal sorting in a human genetic immunodeficiency disease: The Chediak-Higashi syndrome', *Journal of Cell Biology*, 141(5), pp. 1121–1134.

de Felipe, B., Olbrich, P., Lucenas, J. M., *et al.* (2016) 'Prospective neonatal screening for severe T- and B-lymphocyte deficiencies in Seville', *Pediatric Allergy and Immunology*, 27(1), pp. 70–77.

Fiorini, M., Franceschini, R., Soresina, A., *et al.* (2004) '*BTK*: 22 novel and 25 recurrent

- mutations in European patients with X-linked agammaglobulinemia.', *Human mutation*, 23(3), p. 286.
- Gallego-Bustos, F., Gotea, V., Ramos-Amador, J. T., *et al.* (2016) 'A case of IL-7R deficiency caused by a novel synonymous mutation and implications for mutation screening in SCID diagnosis', *Frontiers in Immunology*, 7(OCT), pp. 1–7.
- Gallo, V., Dotta, L., Giardino, G., *et al.* (2016) 'Diagnostics of primary immunodeficiencies through next-generation sequencing', *Frontiers in Immunology*, 7(NOV), pp. 1–10.
- Germeshausen, M., Zeidler, C., Stuhmann, M., Lanciotti, M., Ballmaier, M. and Welte, K. (2010) 'Digenic mutations in severe congenital neutropenia', *Haematologica*, 95(7), pp. 1207–1210.
- Giliani, S., Mori, L., De Saint Basile, G., *et al.* (2005) 'Interleukin-7 receptor α (IL-7R α) deficiency: Cellular and molecular bases. Analysis of clinical, immunological, and molecular features in 16 novel patients', *Immunological Reviews*, 203(7), pp. 110–126.
- Hamsten, C., Skattum, L., Truedsson, L., *et al.* (2015) 'Heat differentiated complement factor profiling', *Journal of Proteomics*, 126, pp. 155–162.
- Heimall, J. R., Hagin, D., Hajjar, J., *et al.* (2018) 'Use of Genetic Testing for Primary Immunodeficiency Patients', *Journal of Clinical Immunology*. *Journal of Clinical Immunology*, 38(3), pp. 320–329.
- Huang, H. Y. and Luther, S. A. (2012) 'Expression and function of interleukin-7 in secondary and tertiary lymphoid organs', *Seminars in Immunology*. Elsevier Ltd, 24(3), pp. 175–189.
- International Patient Organisation for Primary Immunodeficiencies (IPOPI) (2016) *How are primary immunodeficiency diseases classified?* 1st edition.
- International Patient Organisation for Primary Immunodeficiencies (IPOPI) (2020) Available at: <https://ipopi.org/> (Accessed: 27 April 2020).
- Introne, W., Boissy, R. E. and Gahl, W. A. (1999) 'Clinical, molecular, and cell biological aspects of Chediak-Higashi syndrome', *Molecular Genetics and Metabolism*, 68(2), pp. 283–303.
- Iwanami, N. (2014) 'Zebrafish as a model for understanding the evolution of the vertebrate immune system and human primary immunodeficiency', *Experimental Hematology*. Elsevier Ltd, 42(8), pp. 697–706.

- Jabara, H. H., Boyden, S. E., Chou, J., *et al.* (2015) 'A missense mutation in *TFRC*, encoding transferrin receptor 1, causes combined immunodeficiency', *Nature Genetics*, 48(1), pp. 74–78.
- Jacobs, C., Huang, Y., Masud, T., *et al.* (2011) 'A hypomorphic *Artemis* human disease allele causes aberrant chromosomal rearrangements and tumorigenesis', *Human Molecular Genetics*, 20(4), pp. 806–819.
- Jangalwe, S., Shultz, L. D., Mathew, A. and Brehm, M. A. (2016) 'Improved B cell development in humanized NOD-*scid* *IL2R γ* null mice transgenically expressing human stem cell factor, granulocyte-macrophage colony-stimulating factor and interleukin-3', *Immunity Inflammation and Disease*, 4(4), pp. 427–440.
- Joshi, A. Y., Iyer, V. N., Hagan, J. B., St. Sauver, J. L. and Boyce, T. G. (2009) 'Incidence and temporal trends of primary immunodeficiency: A population-based cohort study', *Mayo Clinic Proceedings*, 84(1), pp. 16–22.
- Kalman, L., Lindegren, M., Kobrynski, L., *et al.* (2004) 'Mutations in genes required for T-cell development: *IL7R*, *CD45*, *IL2RG*, *JAK3*, *IMG1*, *IMG2*, *ARTEMIS*, and *ADA* and severe combined immunodeficiency: HuGE review', *Genetics in Medicine*, 6(1), pp. 16–26.
- Kersseboom, R., Brooks, A. and Weemaes, C. (2011) 'Educational paper : Syndromic forms of primary immunodeficiency', *European Journal of Pediatrics*, 170(3), pp. 295–308.
- King, J., Ludvigsson, J. F. and Hammarström, L. (2017) 'Newborn screening for primary immunodeficiency diseases: The past, the present and the future', *International Journal of Neonatal Screening*, 3(3).
- Kraus, M., Lev, A., Simon, A. J., *et al.* (2014) 'Disturbed B and T cell homeostasis and neogenesis in patients with ataxia telangiectasia', *Journal of Clinical Immunology*, 34(5), pp. 561–572.
- Kuo, C. Y., Chase, J., Garcia Lloret, M., *et al.* (2013) 'Newborn screening for severe combined immunodeficiency does not identify bare lymphocyte syndrome', *Journal of Allergy and Clinical Immunology*, 131(6), pp. 1693–1695.
- Kwan, A., Abraham, R. S., Currier, R., *et al.* (2014) 'Newborn screening for severe combined immunodeficiency in 11 screening programs in the United States', *JAMA - Journal of the American Medical Association*, 312(7), pp. 729–738.

- Kwan, A. and Puck, J. M. (2015) 'History and current status of newborn screening for severe combined immunodeficiency', *Seminars in Perinatology*. Elsevier, 39(3), pp. 194–205.
- Lee, W. I., Huang, J. L., Yeh, K. W., *et al.* (2016) 'The effects of prenatal genetic analysis on fetuses born to carrier mothers with primary immunodeficiency diseases', *Annals of Medicine*, 48(1–2), pp. 103–110.
- Lenardo, M., Lo, B. and Lucas, C. L. (2016) 'Genomics of Immune Diseases and New Therapies', *Annual Review of Immunology*.
- Maródi, L. and Notarangelo, L. D. (2007) 'Immunological and genetic bases of new primary immunodeficiencies', *Nature Reviews Immunology*, 7(11), pp. 851–861.
- Martinez-Torres, F., Nochi, T., Wahl, A., Garcia, J. V. and Denton, P. W. (2014) 'Hypogammaglobulinemia in BLT humanized mice - An animal model of primary antibody deficiency', *PLoS ONE*, 9(10), pp. 1–15.
- Masopust, D., Sivula, C. P. and Jameson, S. C. (2017) 'Of Mice, Dirty Mice, and Men: Using Mice To Understand Human Immunology', *The Journal of Immunology*, 199(2), pp. 383–388.
- McCusker, C. and Warrington, R. (2011) 'Primary immunodeficiency', *Allergy, Asthma & Clinical Immunology*.
- Moens, L. N., Falk-Sörqvist, E., Asplund, A. C., Bernatowska, E., Smith, C. I. E. and Nilsson, M. (2014) 'Diagnostics of primary immunodeficiency diseases: A sequencing capture approach', *PLoS ONE*.
- Möhlig, H., Mathieu, S., Thon, L., *et al.* (2007) 'The WD repeat protein FAN regulates lysosome size independent from abnormal downregulation/membrane recruitment of protein kinase C', *Experimental Cell Research*, 313(12), pp. 2703–2718.
- Moore, J. C., Mulligan, T. S., Yordán, N. T., *et al.* (2016) 'T Cell Immune Deficiency in *zap70* Mutant Zebrafish', *Molecular and Cellular Biology*, 36(23), pp. 2868–2876.
- Nakagawa, N., Imai, K., Kanegane, H., *et al.* (2011) 'Quantification of κ -deleting recombination excision circles in Guthrie cards for the identification of early B-cell maturation defects', *Journal of Allergy and Clinical Immunology*, 128(1), pp. 223–225.e2.
- Nijman, I. J., Van Montfrans, J. M., Hoogstraal, M., *et al.* (2014) 'Targeted next-generation sequencing: A novel diagnostic tool for primary immunodeficiencies',

- Journal of Allergy and Clinical Immunology*. Elsevier Ltd, 133(2), pp. 529-534.e1.
- Ochs, H. D. and Hitzig, W. H. (2012) 'History of primary immunodeficiency diseases', *Current Opinion in Allergy and Clinical Immunology*, 12(6), pp. 577–587.
- Ozturk, C., Sutcuoglu, S., Atabay, B. and Berdeli, A. (2013) 'X-linked agammaglobulinemia presenting with secondary hemophagocytic syndrome: A case report', *Case Reports in Medicine*.
- Pai, S. Y., Logan, B. R., Griffith, L. M., *et al.* (2014) 'Transplantation outcomes for severe combined immunodeficiency, 2000-2009', *New England Journal of Medicine*, 371(5), pp. 434–446.
- Parvaneh, N., Lilic, D., Roesler, J., Niehues, T., Casanova, J. L. and Picard, C. (2017) 'Defects in Intrinsic and Innate Immunity: Receptors and Signaling Components', in *Primary Immunodeficiency Diseases*. Springer-Verlag Berlin Heidelberg, pp. 339–392.
- Picard, C., Bobby Gaspar, H., Al-Herz, W., *et al.* (2018) 'International Union of Immunological Societies: 2017 Primary Immunodeficiency Diseases Committee Report on Inborn Errors of Immunity', *Journal of Clinical Immunology*. *Journal of Clinical Immunology*, 38(1), pp. 96–128.
- Punwani, D., Zhang, Y., Yu, J., *et al.* (2016) 'Multisystem anomalies in severe combined immunodeficiency with mutant *BCL11B*', *New England Journal of Medicine*, 375(22), pp. 2165–2176.
- Rae, W., Ward, D., Mattocks, C., *et al.* (2018) 'Clinical efficacy of a next-generation sequencing gene panel for primary immunodeficiency diagnostics', *Clinical Genetics*, 93(3), pp. 647–655.
- Raje, N., Soden, S., Swanson, D., Ciaccio, C. E., Kingsmore, S. F. and Dinwiddie, D. L. (2014) 'Utility of Next Generation Sequencing in Clinical Primary Immunodeficiencies', *Current Allergy and Asthma Reports*, 14(10), pp. 1–13.
- Rezaei, N., Bonilla, F. A., Seppänen, M., *et al.* (2017) 'Introduction on Primary Immunodeficiency Diseases', in *Primary Immunodeficiency Diseases*. Springer-Verlag Berlin Heidelberg, pp. 1–81.
- Richards, S., Aziz, N., Bale, S., *et al.* (2015) 'Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for

- Molecular Pathology', *Genetics in Medicine*. IOP Publishing, 17(5), pp. 405–424.
- Rothblum-Oviatt, C., Wright, J., Lefton-Greif, M. A., McGrath-Morrow, S. A., Crawford, T. O. and Lederman, H. M. (2016) 'Ataxia telangiectasia: A review', *Orphanet Journal of Rare Diseases*. Orphanet Journal of Rare Diseases, 11(1), pp. 1–21.
- Royer-Bertrand, B. and Rivolta, C. (2015) 'Whole genome sequencing as a means to assess pathogenic mutations in medical genetics and cancer', *Cellular and Molecular Life Sciences*, 72(8), pp. 1463–1471.
- Schubert, R., Reichenbach, J. and Zielen, S. (2005) 'Growth factor deficiency in patients with ataxia telangiectasia', *Clinical and Experimental Immunology*, 140(3), pp. 517–519.
- Seleman, M., Hoyos-Bachiloglu, R., Geha, R. S. and Chou, J. (2017) 'Uses of next-generation sequencing technologies for the diagnosis of primary immunodeficiencies', *Frontiers in Immunology*, 8(JUL), pp. 1–8.
- Sims, D., Sudbery, I., Illott, N. E., Heger, A. and Ponting, C. P. (2014) 'Sequencing depth and coverage: Key considerations in genomic analyses', *Nature Reviews Genetics*. Nature Publishing Group, 15(2), pp. 121–132.
- Singh, S., Rawat, A., Suri, D., *et al.* (2016) 'X-linked agammaglobulinemia: Twenty years of single-center experience from North West India', *Annals of Allergy, Asthma and Immunology*. American College of Allergy, Asthma & Immunology, 117(4), pp. 405–411.
- Smith, C. I., Baskin, B., Humire-Greiff, P., *et al.* (1994) 'Expression of Bruton's agammaglobulinemia tyrosine kinase gene, *BTK*, is selectively down-regulated in T lymphocytes and plasma cells.', *Journal of immunology (Baltimore, Md. : 1950)*, 152(2), pp. 557–65.
- Sobh, A., Chou, J., Schneider, L., Geha, R. S. and Massaad, M. J. (2016) 'Chronic mucocutaneous candidiasis associated with an SH2 domain gain-of-function mutation that enhances STAT1 phosphorylation', *Journal of Allergy and Clinical Immunology*. Elsevier Ltd, 138(1), pp. 297–299.
- Stoddard, J. L., Niemela, J. E., Fleisher, T. A. and Rosenzweig, S. D. (2014) 'Targeted NGS: A cost-effective approach to molecular diagnosis of PIDs', *Frontiers in Immunology*, 5(NOV), pp. 1–7.
- Takagi, M., Shinoda, K., Piao, J., *et al.* (2011) 'Autoimmune lymphoproliferative

- syndrome-like disease with somatic *KRAS* mutation', *Blood*, 117(10), pp. 2887–2890.
- Tan, W., Yu, S., Lei, J., Wu, B. and Wu, C. (2015) 'A novel common gamma chain mutation in a Chinese family with X-linked severe combined immunodeficiency (X-SCID; T–NK–B+)', *Immunogenetics*, 67(11–12), pp. 629–639.
- Tangye, S. G., Al-Herz, W., Bousfiha, A., *et al.* (2020) 'Human Inborn Errors of Immunity: 2019 Update on the Classification from the International Union of Immunological Societies Expert Committee', *Journal of Clinical Immunology*. *Journal of Clinical Immunology*, 40(1), pp. 24–64.
- Toro, C., Nicoli, E., Malicdan, M. C., Adams, D. R. and Introne, W. J. (2009) 'Chediak-Higashi Syndrome Summary Suggestive Findings', *GeneReviews*, pp. 1–22.
- Valencia, C. A., Ankala, A., Rhodenizer, D., *et al.* (2013) 'Comprehensive Mutation Analysis for Congenital Muscular Dystrophy: A Clinical PCR-Based Enrichment and Next-Generation Sequencing Panel', *PLoS ONE*, 8(1), pp. 1–11.
- de Valles-Ibáñez, G., Esteve-Solé, A., Piquer, M., *et al.* (2018) 'Evaluating the genetics of common variable immunodeficiency: Monogenetic model and beyond', *Frontiers in Immunology*, 9, p. 636.
- Walkovich, K. and Connelly, J. A. (2016) 'Primary immunodeficiency in the neonate: Early diagnosis and management', *Seminars in Fetal and Neonatal Medicine*. Elsevier Ltd, 21(1), pp. 35–43.
- Wilson, J. M. G. and Jungner, G. (1968) 'Principles and Practice of Screening for Disease. WHO Public Paper 34', *Geneva: World Health Bibliography Organization*, p. 168.
- Yu, H., Zhang, V. W., Stray-Pedersen, A., *et al.* (2016) 'Rapid molecular diagnostics of severe primary immunodeficiency determined by using targeted next-generation sequencing', *Journal of Allergy and Clinical Immunology*. Elsevier Inc., 138(4), pp. 1142-1151.
- Zhang, K., Chandrakasan, S., Chapman, H., *et al.* (2014) 'Synergistic defects of different molecules in the cytotoxic pathway lead to clinical familial hemophagocytic lymphohistiocytosis', *Blood*, 124(8), pp. 1331–1334.
- Zhang, Q., Davis, J. C., Lamborn, I. T., *et al.* (2009) 'Combined immunodeficiency associated with *DOCK8* mutations', *New England Journal of Medicine*, 361(21), pp.

2046–2055.

Zhang, Q., Frange, P., Blanche, S. and Casanova, J. L. (2017) 'Pathogenesis of infections in HIV-infected individuals: insights from primary immunodeficiencies', *Current Opinion in Immunology*, 48, pp. 122–133.

8. Appendices

Appendix 1: Coverage metrics for Panel A genes

Table A1. Coverage metrics for the genes included in Panel A.

Gene	Number of exons	Total target bases (bp)	<i>In silico</i> missed bases (bp)	<i>In silico</i> target base coverage	<i>In silico</i> + Experimental target base coverage
ADA	12	1212	0	100%	100%
AICDA	5	847	0	100%	100%
AK2	7	795	0	100%	100%
ARPC1B	9	1569	0	100%	100%
ATM	62	9791	0	100%	100%
B2M	3	510	0	100%	100%
BCL10	3	852	0	100%	100%
BLM	21	4464	0	100%	100%
BLNK	17	1541	0	100%	100%
BTK	19	2272	0	100%	90,54%
CARD11	24	3705	0	100%	100%
CCBE1	11	1661	0	100%	100%
CD19	14	2374	0	100%	100%
CD247	8	575	0	100%	100%
CD3D	5	566	0	100%	100%
CD3E	8	704	0	100%	100%
CD3G	6	609	0	100%	100%
CD40	9	1324	0	100%	100%
CD40LG	5	836	0	100%	100%
CD79A	5	931	0	100%	100%
CD79B	6	753	0	100%	100%
CD81	8	1111	0	100%	61,93%
CDCA7	10	1853	0	100%	100%
CIITA	19	4392	0	100%	100%
CORO1A	10	1886	155	91,80%	91,80%
CR2	19	4229	0	100%	100%
DCLRE1C	14	2219	0	100%	100%
DKC1	15	1699	0	100%	86,82%
DNMT3B	23	2828	0	100%	100%
DOCK2	52	8093	0	100%	97,23%
DOCK8	48	6780	0	100%	96,59%
EPG5	44	9940	0	100%	91,52%
ERCC6L2	19	5642	0	100%	100%
EXTL3	5	3010	0	100%	100%
FOXN1	8	2027	0	100%	100%

<i>GINS1</i>	7	941	0	100%	100%
<i>HELLS</i>	23	3805	0	100%	100%
<i>ICOS</i>	5	850	0	100%	100%
<i>IIGLL1</i>	3	792	0	100%	100%
<i>IKBKB</i>	22	3470	0	100%	100%
<i>IKZF1</i>	9	2029	0	100%	100%
<i>IL21</i>	5	763	0	100%	100%
<i>IL21R</i>	9	2133	0	100%	100%
<i>IL2RG</i>	8	1190	0	100%	100%
<i>IL7R</i>	8	1460	0	100%	100%
<i>INO80</i>	35	6421	0	100%	100%
<i>JAK3</i>	23	3605	0	100%	100%
<i>LAT</i>	11	1447	0	100%	100%
<i>LCK</i>	12	2130	0	100%	100%
<i>LIG1</i>	27	4110	0	100%	100%
<i>LIG4</i>	1	2746	0	100%	100%
<i>LYST</i>	51	11916	3	99,97%	98,54%
<i>MAGT1</i>	10	1604	0	100%	100%
<i>MALT1</i>	17	3325	50	98,50%	96,33%
<i>MAP3K14</i>	15	3593	0	100%	100%
<i>MCM4</i>	16	3392	0	100%	100%
<i>MOGS</i>	4	2714	0	100%	91,38%
<i>MS4A1</i>	6	1194	0	100%	85,59%
<i>MSH6</i>	10	4183	0	100%	90,63%
<i>MSN</i>	13	2384	0	100%	100%
<i>MTHFD1</i>	27	3078	0	100%	100%
<i>MYSM1</i>	20	3487	0	100%	100%
<i>NBN</i>	16	2425	0	100%	100%
<i>NFKB1</i>	23	4060	0	100%	95,71%
<i>NFKB2</i>	22	3803	0	100%	94,74%
<i>NFKBIA</i>	6	1254	0	100%	100%
<i>NHEJ1</i>	7	970	0	100%	100%
<i>NOP10</i>	2	295	0	100%	100%
<i>PARN</i>	24	3120	0	100%	100%
<i>PGM3</i>	14	2497	0	100%	100%
<i>PIK3CD</i>	22	3355	0	100%	95,02%
<i>PIK3R1</i>	17	2467	0	100%	100%
<i>PMS2</i>	15	2739	0	100%	100%
<i>PNP</i>	6	930	0	100%	100%
<i>POLE</i>	49	7351	3	99,90%	94,04%
<i>POLE2</i>	19	2546	0	100%	100%
<i>PRKDC</i>	86	13246	0	100%	100%
<i>PTPRC</i>	33	4415	0	100%	100%
<i>RAG1</i>	1	3142	0	100%	100%

<i>RAG2</i>	1	1594	0	100%	100%
<i>RBCK1</i>	12	2152	0	100%	100%
<i>RELB</i>	11	2286	7	99,70%	99,70%
<i>RFX5</i>	9	2301	0	100%	100%
<i>RFXANK</i>	8	1183	0	100%	100%
<i>RFXAP</i>	3	969	0	100%	100%
<i>RHOH</i>	1	626	0	100%	100%
<i>RNF168</i>	6	2016	0	100%	100%
<i>RNF31</i>	22	4361	0	100%	100%
<i>RTEL1</i>	34	5675	0	100%	100%
<i>SAMD9</i>	1	4820	0	100%	100%
<i>SAMD9L</i>	1	4805	0	100%	100%
<i>SKIV2L</i>	28	5141	0	100%	100%
<i>SLC46A1</i>	6	1440	0	100%	84,86%
<i>SMARCAL1</i>	16	3665	0	100%	100%
<i>SP110</i>	20	3220	0	100%	100%
<i>SPINK5</i>	34	4997	0	100%	100%
<i>STAT3</i>	23	2543	0	100%	100%
<i>STAT5B</i>	18	2544	43	98,30%	98,30%
<i>STIM1</i>	13	3063	0	100%	100%
<i>STK4</i>	11	2014	0	100%	100%
<i>TAP1</i>	11	2977	0	100%	100%
<i>TCF3</i>	19	3142	0	100%	85,81%
<i>TCN2</i>	9	1734	0	100%	100%
<i>TERT</i>	16	3559	0	100%	93,59%
<i>TFRC</i>	18	3003	10	99,70%	91,81%
<i>TINF2</i>	9	1810	0	100%	100%
<i>TNFRSF13B</i>	5	1132	0	100%	100%
<i>TNFRSF13C</i>	3	705	13	98,20%	84,96%
<i>TNFRSF4</i>	7	1184	0	100%	65,03%
<i>TRNT1</i>	7	1655	0	100%	100%
<i>TTC37</i>	40	6695	0	100%	97,03%
<i>TTC7A</i>	22	3831	0	100%	100%
<i>UNC119</i>	5	1026	0	100%	100%
<i>WAS</i>	12	1629	0	100%	87,17%
<i>WIPF1</i>	7	1862	0	100%	100%
<i>ZAP70</i>	12	1980	0	100%	100%
<i>ZBTB24</i>	6	2444	0	100%	100%

Appendix 2: Primers designed for the study of PID regions/genes.

Table A2. Primers designed for low coverage regions.

Gene	Region	Primer	Nucleotide Sequence 5' - 3'	Amplified product size	Amplification conditions
<i>BTK</i>	14	g.BTK-14-F	TGTA AACGACGGCCAGTACCAATGAATCCCGTTTCTGAG	452 bp	Standard
		g.BTK-14-R	CAGGAAACAGCTATGACCTCCCTCGTCCCAAACCTCTC		
<i>CD81</i>	1	g.CD81-1-F	GGCGCCCTATAAGTACTGCGG	564 bp	High G/C
		g.CD81-1-R	GAGCACAGAGGCCCGAA		
	3	g.CD81-3-F	GAGGTCCCTTGCTGCTCATC	441 bp	Standard
		g.CD81-3-R	GTAGGA ACTGGCCACGATGG		
<i>DKC1</i>	11	g.DKC1-11-F	TGTA AACGACGGCCAGTACGTCTTGAGCTGCAAGCCTG	409 bp	Standard
		g.DKC1-11-R	CAGGAAACAGCTATGACCCGCAACCCAGTACCACTTACC		
<i>DOCK2</i>	33	g.DOCK2-33-F	TGTA AACGACGGCCAGTAAAACGGTGCCTGGCTCATG	395 bp	Standard
		g.DOCK2-33-R	CAGGAAACAGCTATGACCTGAAGCTCAGGGAGATGGAATG		
<i>DOCK8</i>	1	g.DOCK8-1-F	TTCGGCTTAGAAGGTGGAAATG	506 bp	Standard
		g.DOCK8-1-R	TGTAAGAACCTGGAGGGCGG		
<i>EPG5</i>	25	g.EPG5-25-F	TGTA AACGACGGCCAGTACTGCAGGGATTGTTAGAGTGG	435 bp	Standard
		g.EPG5-25-R	CAGGAAACAGCTATGACCGCACCAGAGGAAAAATGAACAC		
	26	g.EPG5-26-F	CCACATGCTTCTGTTCTCCTGC	317 bp	Standard
		g.EPG5-26-R	CCACTCCAGGTAGCAAATTATCCAG		
	27	g.EPG5-27-F	TGTA AACGACGGCCAGTGCTCACAACTTTGGTCCTCTG	452 bp	Standard
		g.EPG5-27-R	CAGGAAACAGCTATGACCTGAGTGTTTTTCATGCCCTGC		
	39	g.EPG5-39-F	TGTA AACGACGGCCAGTCTTCATGTGGCTTCCTGTTCTG	446 bp	Standard
		g.EPG5-39-R	CAGGAAACAGCTATGACCTCCTTCCCCACTGCTGAGTC		
<i>LYST</i>	42	g.LYST-42-F	TGTA AACGACGGCCAGTTCCTGGGTAAGAGTGGGAGGG	454 bp	Standard
		g.LYST-42-R	CAGGAAACAGCTATGACCAGGAGCACTTTGGGCAAGG		

<i>MALT1</i>	3	g.MALT1-3-F	TCAAAGTCTTCTGCAAGTAAGTTTC	599 bp	Standard
		g.MALT1-3-R	CTGCTGCCCTAAGTAATCCC		
<i>MOGS</i>	1	g.MOGS-1-F	GACCAACGCCCAGGACAGAC	551 bp	Standard
		g.MOGS-1-R	TTCCCCAGAAGAGGTCCG		
<i>MS4A1</i>	8	g.MS4A1-8-F	CAGGAGGTATTTGATAAATGTTTGTGG	458 bp	Standard
		g.MS4A1-8-R	AGAGAAGGAGCTATGGTCACTCC		
<i>MSH6</i>	4	g.MSH6-4-F	TGTA AACGACGGCCAGTACGAATGGTGCTAGATGCAGTG	521 bp	Standard
		g.MSH6-4-R	CAGGAAACAGCTATGACCAATCAGGAAAACGACCTTCAGG		
	9	g.MSH6-9-F	GGTACTGCAACATTTGATGGG	392 bp	Standard
		g.MSH6-9-R	TGTCCCTTTTGAATAACTTCCTCTGG		
<i>NFKB1</i>	10	g.NFKB1-10-F	GTTCGTGTTTCTCAGGATTGGC	532 bp	Standard
		g.NFKB1-10-R	GAGGATCCCTCAAGCCCAAC		
<i>NFKB2</i>	7/8	g.NFKB2-7-8-F	TGTA AACGACGGCCAGTCCAAACCCAAGGGCCTAAG	479 bp	Standard
		g.NFKB2-7-8-R	CAGGAAACAGCTATGACCCAGGGAGAAGGAGCCATCAC		
<i>PIK3CD</i>	12/13	g.PIK3CD-12-13-F	TTGGCATCTCGTGAGTGAGG	447 bp	Standard
		g.PIK3CD-12-13-R	CCTCATGCTTGTTCCACTTGG		
<i>POLE</i>	24	g.POLE-24-F	TGTA AACGACGGCCAGTTTGTCAATCCATCCACTCCTGG	344 bp	Standard
		g.POLE-24-R	CAGGAAACAGCTATGACCTCCTGGCTCCTGATCCAACC		
	26	g.POLE-26-F	TGTA AACGACGGCCAGTCAGGCTAGAGAAGGGCATGG	523 bp	Standard
		g.POLE-26-R	CAGGAAACAGCTATGACCTCCATTTCTGGTTGATGGACAC		
<i>SLC46A1</i>	1	g.SLC46A1-1-F	TGTA AACGACGGCCAGTTTACAGGTGAGGTCATCCCG	640 bp	Standard
		g.SLC46A1-1-R	CAGGAAACAGCTATGACCGCTGAGGTTCTCGGTCAGG		
<i>TCF3</i>	10	g.TCF3-10-F	TGTA AACGACGGCCAGTTCAGGTGAGGACTACGGCAG	487 bp	Standard
		g.TCF3-10-R	CAGGAAACAGCTATGACCGAAGGAGGATGCAGATGGGAG		
	19	g.TCF3-19-F	TGTA AACGACGGCCAGTTTTGTGGCTGGACTCAGACCC	571 bp	Standard
		g.TCF3-19-R	CAGGAAACAGCTATGACCTGGACCCAGTGTACCTTGG		
<i>TERT</i>	1	g.TERT-1-F	GGCTCCCAGTGGATTCCG	570 bp	High G/C
		g.TERT-1-R	ATGTCGCTGGTTCCCCC		

<i>TFRC</i>	13	g.TFRC-13-F	TGTA ^g AAACGACGGCCAGT ^g TTTTACCACTTTTGCCTTTGCG	315 bp	Standard
		g.TFRC-13-R	CAGGAAACAGCTATGACCTCTGGGAGACAGAGGTTGCAG		
<i>TNFRSF13C</i>	2	g.TNFRSF13C-2-F	TGTCCCCTCCCGAAGCAG	461 bp	High G/C
		g.TNFRSF13C-2-R	AGCTGTTCTGGCTCCGAC		
<i>TNFRSF4</i>	2	g.TNFRSF4-2-F	GGACACCTACCCAGCAACG	435 bp	Standard
		g.TNFRSF4-2-R	TGGGAATGTGGGTGCCAG		
	4	g.TNFRSF4-4-F	TGTA ^g AAACGACGGCCAGTAGGGCTTCCAGAGGCCAAAC	416 bp	Standard
		g.TNFRSF4-4-R	CAGGAAACAGCTATGACCGCTCCACCCACAGGGAAAG		
<i>TTC37</i>	27	g.TTC37-27-F	TTACTGTGCTAGCCTCTGTGTCAG	530 bp	Standard
		g.TTC37-27-R	TCTATTCATTCCCTTAACAAGGATAAC		
<i>WAS</i>	10	g.WAS-10-F	TGTA ^g AAACGACGGCCAGTTCAGTCAGGAGTTGGTCAGTGG	499 bp	Standard
		g.WAS-10-R	CAGGAAACAGCTATGACCTCCCTGCCGATTTGATC		

F – forward; R – reverse. Sequence of M13 universal primer tails highlighted in gray.

Table A3. Primers designed for the study of additional PID-related genes.

Gene	Region	Primer	Nucleotide Sequence 5' - 3'	Amplified product size	Amplification conditions
<i>BCL11B</i>	1A	g.BCL11B-1A-F	TGTA ^g AAACGACGGCCAGTAAGGGAAAAA ^g AAACACCAACCC	345 bp	High G/C
		g.BCL11B-1-R	CAGGAAACAGCTATGACCCGTAGACTCTGCCAGCCAGC		
	1B	g.BCL11B-1B-F	TGTA ^g AAACGACGGCCAGTGGCTCCCCCATCAGTGC	242 bp	High G/C
		g.BCL11B-1-R	CAGGAAACAGCTATGACCCGTAGACTCTGCCAGCCAGC		
	2	g.BCL11B-2-F	GGAGGACCTGGTGAGCTCTG	586 bp	High G/C
		g.BCL11B-2-R	AGCACACTCAGGCAGAGGG		
	3	g.BCL11B-3-F	TGTA ^g AAACGACGGCCAGTTGTCCCGATTGCCTTCCC	518 bp	High G/C
		g.BCL11B-3-R	CAGGAAACAGCTATGACCTTCTTACCACTTCCCCTGGC		
	4A	g.BCL11B-4A-F	AAGGCGGGTCTTACACAATTGAAG	576 bp	High G/C
		g.BCL11B-4A-R	TGAAGAGAGGCGGCGTGC		

	4B	g.BCL11B-4B-F	TGTA AACGACGGCCAGTATGAATTTCTGGGCGACAG	497 bp	High G/C	
		g.BCL11B-4B-R	CAGGAAACAGCTATGACCTGCACGATGAGATTGCTCTGG			
	4C	g.BCL11B-4C-F	GCCATGGACTTCTCGCGG	479 bp	High G/C	
		g.BCL11B-4C-R	TGGCGGAAGTCACCGTCG			
	4D	g.BCL11B-4D-F	TGTA AACGACGGCCAGTAAGCTCAAGCGCCACATG	410 bp	High G/C	
		g.BCL11B-4D-R	CAGGAAACAGCTATGACCCTAGGCCACGTTCTCC			
	4E	g.BCL11B-4E-F	CCCGAGTCGAGCTTCAGC	546 bp	High G/C	
		g.BCL11B-4E-R	TCGTGCGTCCGTGAAGCC			
	4F	g.BCL11B-4F-F	TGTA AACGACGGCCAGTGAAGGACCTGGAGCTGCCG	465 bp	High G/C	
		g.BCL11B-4F-R	CAGGAAACAGCTATGACCTGCGTCTTCATGTGGCGC			
	4G	g.BCL11B-4G-F	TGTA AACGACGGCCAGTAGAACTGCAGCAACTTGACGG	473 bp	High G/C	
		g.BCL11B-4G-R	CAGGAAACAGCTATGACCGGCTCCACAATTTGTA CTGCC			
	CD8A	1	g.CD8A-1-F	TGTA AACGACGGCCAGTCCCTTTCTCATCCCCAAAC	493 bp	Standard
			g.CD8A-1-R	CAGGAAACAGCTATGACCTTCCAGGTCCGATCCAGC		
2A		g.CD8A-2A-F	TGTA AACGACGGCCAGTTCCTCCAAATGCGAAATCAGG	462 bp	Standard	
		g.CD8A-2A-R	CAGGAAACAGCTATGACCTTGGGCTTGTTTTGGGAGAG			
2B		g.CD8A-2B-F	TGTA AACGACGGCCAGTGATCGGACCTGGAACCTGG	478 bp	Standard	
		g.CD8A-2B-R	CAGGAAACAGCTATGACCGCCTCGCGTTAGCCTCAG			
3		g.CD8A-3-F	TGTA AACGACGGCCAGTTGCTCGTGGGTAAAATGCG	561 bp	High G/C	
		g.CD8A-3-R	CAGGAAACAGCTATGACCGGAACGTCTCTCCGCCTC			
4		g.CD8A-4-F	TGTA AACGACGGCCAGTTGTCCAAGTGGTGGGAAACC	367 bp	Standard	
		g.CD8A-4-R	CAGGAAACAGCTATGACCAGAAGTGCCTGTACCTGCGG			
5		g.CD8A-5-F	TGTA AACGACGGCCAGTATTTGCCTCATCCAATTT CAGC	353 bp	Standard	
		g.CD8A-5-R	CAGGAAACAGCTATGACCGCCTTGGCCTCTCTTTGCTC			
6		g.CD8A-6-F	TGTA AACGACGGCCAGTTGACGTCATTCACCCAGTCCC	529 bp	Standard	
		g.CD8A-6-R	CAGGAAACAGCTATGACCAGAGTTGAGATGGCATGGG			
RMRP		g.RMRP-1-F	TTCCTGTATTTGTTCAATTCGTGTC			

	1	g.RMRP-1-R	CGAGGTGGACTGATCGCTTG	658 bp	Standard
TAP2	1	g.TAP2-1-F	TGTA AACGACGGCCAGTCCCAGAGTAACCGCCACTAAAG	385 bp	Standard
		g.TAP2-1-R	CAGGAAACAGCTATGACCGCAGAGGGACTGGGAAGCAG		
	2A1	g.TAP2-2A1-F	TGTA AACGACGGCCAGTCAGGGACAAGGATTGGGACTC	498 bp	High G/C
		g.TAP2-2A-R	CAGGAAACAGCTATGACCTCACCTGGTCCTGCTCCTTC		
	2A2	g.TAP2-2A2-F	TGTA AACGACGGCCAGTGAGGCTCCTGATGGTCATGC	623 bp	Standard
		g.TAP2-2A-R	CAGGAAACAGCTATGACCTCACCTGGTCCTGCTCCTTC		
	2B	g.TAP2-2B-F	ACTGCTGCTCCCGCTCTGTC	409 bp	Standard
		g.TAP2-2B-R	GAGGGATTAATGTCTCACCTGAAAG		
	3	g.TAP2-3-F	TGTA AACGACGGCCAGTGTTTTGGGTGAGTCAGGAGAGG	309 bp	Standard
		g.TAP2-3-R	CAGGAAACAGCTATGACCTGTTTTGCGCCTGAAAGGG		
	4	g.TAP2-4-F	CTTACCTCTTCTCTCTCTGTGTGTCTG	399 bp	Standard
		g.TAP2-4-R	CATACGGGTGAAGGCAGGAG		
	5	g.TAP2-5-F	TGTA AACGACGGCCAGTCCCTCTGCAAGTGAAGGAGG	416 bp	Standard
		g.TAP2-5-R	CAGGAAACAGCTATGACCGGGATCCTCTAGCCACAAATG		
	6	g.TAP2-6-F	ACTGGCTGGAGTAAGATGTGGG	343 bp	Standard
		g.TAP2-6-R	CAGCTGTAGTTTCTCTTCCCTTG		
	7	g.TAP2-7-F	TGTA AACGACGGCCAGTGATTGGCGGAGAGACCTGG	421 bp	Standard
		g.TAP2-7-R	CAGGAAACAGCTATGACCCAAATAAAGCCCAAGGCC		
	8	g.TAP2-8-F	TGTA AACGACGGCCAGTAGGAGGAGCAGTGCAGTTGTG	493 bp	Standard
		g.TAP2-8-R	CAGGAAACAGCTATGACCAGGACGTAGGGTAAACGTCAGC		
	9	g.TAP2-9-F	TGTA AACGACGGCCAGTAAACCTGGACCCTTGCTCTCTG	480 bp	Standard
g.TAP2-9-R		CAGGAAACAGCTATGACCTGTGTAAACCCCAAGGCAG			
10	g.TAP2-10-F	TGTA AACGACGGCCAGTACCAGGTGTTCAATTCTGAGGG	384 bp	Standard	
	g.TAP2-10-R	CAGGAAACAGCTATGACCGGAATGGAGGAAAGGGCAG			
11	g.TAP2-11-F	TGTA AACGACGGCCAGTACCTGCTGTGCACTTGTCCC	448 bp	Standard	
	g.TAP2-11-R	CAGGAAACAGCTATGACCCCTGGACTGCCCTTCTCTCC			

	12	g.TAP2-12-F	CTTTATCTACTGCCCTTTCCTACCTTC	540 bp	Standard
		g.TAP2-12-R	AGGCAAAGCTCTAGTCACCAGG		
<i>TAPBP</i>	1	g.TAPBP-1-F	GGGACGCGGCACAGATAG	499 bp	Standard
		g.TAPBP-1-R	GGAGCGGTTTGTATGTCTGGTG		
	2	g.TAPBP-2-F	TGTA ^g AAACGACGGCCAGTCTTCACTAACCAAGAGCGGG	533 bp	Standard
		g.TAPBP-2-R	CAGGAAACAGCTATGACCCTGCCGGATCTAAAGAGGAGG		
	3	g.TAPBP-3-F	TGTA ^g AAACGACGGCCAGTGCATTTGAAGTTCCCGAAC	443 bp	Standard
		g.TAPBP-3-R	CAGGAAACAGCTATGACCTTGGCGAAATGTCTTGCTC		
	4	g.TAPBP-4-F	TTTAGCCCTCCCAATAACTGC	570 bp	Standard
		g.TAPBP-4-R	CATCCCTCCCCCTATTACGG		
	5	g.TAPBP-5-F	TGTA ^g AAACGACGGCCAGTAGCCAGCAGGCAAAGTGGAG	567 bp	Standard
		g.TAPBP-5-R	CAGGAAACAGCTATGACCCCAAGCAGAAGAAAGGCAG		
	6/7	g.TAPBP-6-7-F	TGTA ^g AAACGACGGCCAGTCTCTCTGGGCACTTGCAG	546 bp	Standard
		g.TAPBP-6-7-R	CAGGAAACAGCTATGACCCTCCCTCCATGGGTTTTGAG		
8	g.TAPBP-8-F	TGTA ^g AAACGACGGCCAGTAATAGCCTAATGTGGGTGGTGG	375 bp	Standard	
	g.TAPBP-8-R	CAGGAAACAGCTATGACCTTGGAGGGATTAGGAGCAGATG			
<i>TERC</i>	1A	g.TERC-1A-F	GCAACGTCCTTCCTCATGGC	485 bp	Standard
		g.TERC-1A-R	CGGCTGACAGAGCCCAAC		
	1B	g.TERC-1B-F	CGCCTTCCACCGTTCATTC	455 bp	Standard
		g.TERC-1B-R	ACTTTGGAGGTGCCTTCACG		
<i>TRAC</i>	1	g.TRAC-1-F	TGTA ^g AAACGACGGCCAGTTTGTGCCTGTCCCTGAGTCC	600 bp	High G/C
		g.TRAC-1-R	CAGGAAACAGCTATGACCGGTGGCAATGGATAAGGCC		
	2	g.TRAC-2-F	TGTA ^g AAACGACGGCCAGTCCCTCCAGATCTCTCCACAAGG	295 bp	Standard
		g.TRAC-2-R	CAGGAAACAGCTATGACCCATTCTCATTCATTCCC		
	3	g.TRAC-3-F	TGTA ^g AAACGACGGCCAGTAAGGGACAGGAGGTGCAGGAG	350 bp	High G/C
		g.TRAC-3-R	CAGGAAACAGCTATGACCGGACTTAAAAGCAAGACCCTGC		

F – forward; R – reverse. Sequence of M13 universal primer tails highlighted in gray.

Appendix 3: Sequencing metrics for control and patient samples

Table A4. Sequencing metrics.

ID	On target (%)	Mean coverage of target regions (x)	Bases with coverage >20x (%)	Total number of variants	Filtered-in variants	Pathogenic/Likely pathogenic variants	Benign/Likely benign variants	Variants of uncertain significance
C1	98.77	383.4	99.08	459	61	1	40	14
C2	98.69	206.9	98.70	391	56	1	32	11
C3	98.60	355.9	99.24	449	55	1	40	9
C4	98.69	316.4	99.12	446	52	1	39	6
P1	98.30	271.9	99.11	437	49	0	38	5
P2	98.03	256.0	99.16	394	53	0	35	8
P3	98.23	306.5	99.36	400	47	0	34	7
P4	98.21	291.4	99.13	460	59	0	40	10


Appendix 4: Variants detected by Sanger sequencing in the 4 patients

Table A5. Description of the variants detected by Sanger sequencing in the 4 patients.

Gene	gDNA variant	Effect on protein	Cases		Allele Frequency
			Het.	Hom.	
EPG5	c.4532C>T	p.Ala1511Val	P2, P4	-	23.81%
	c.4646+35T>C	p.?	P2, P4	-	22.02%
DOCK8	c.53+28T>C	p.?	P3	-	55.36%
	c.53+240A>C	p.?	P3	-	23.84%
LYST	c.9784+67_9784+69dup	p.?	P3	-	9.93%
MALT1	c.377-64A>C	p.?	P2, P3	-	3.80%
	c.377-104A>T	p.?	P1	-	9.88%
MS4A1	c.676-40G>T	p.?	P2, P3, P4	-	7.35%
MSH6	c.3802-40C>G	p.?	P1, P2, P3	P4	73.81%
NFKB1	c.836-199G>C	p.?	-	P1, P3, P4	41.49%
POLE	c.2864+37A>G	p.?	P1, P2, P4	P3	49.78%
	c.3156G>A	p.Thr1052=	P1, P2, P4	P3	50.62%
	c.3275+92T>C	p.?	P1, P2, P4	P3	52.69%
TCF3	c.759C>T	p.Ser253=	P3	-	22.29%
	c.1823-8C>T	p.?	P2, P3, P4	-	16.15%
TFRC	c.1468+25G>T	p.?	P3	P1, P2, P4	81.63%
	c.1468+39A>G	p.?	P1, P4	-	50.07%
	c.1468+77G>A	p.?	P1, P2, P3	-	18.31%
TNFRSF13C	c.191G>T	p.Gly64Val	P2	-	0.87%
	c.192C>T	p.Gly64=	P2	-	0.87%
	c.367+89G>C	p.?	-	P4	6.57%
TTC37	c.2704-83G>A	p.?	P4	-	1.88%

Het.: Heterozygosity; Hom.: Homozygosity.


Appendix 5: Poster presented at “XLVIII Conferências de Genética Doutor Jacinto Magalhães: 40 Anos Ao Serviço Da Comunidade”.



DESIGN AND VALIDATION OF AN NGS APPROACH FOR THE DIAGNOSIS OF PRIMARY IMMUNODEFICIENCIES

Raquel Silva^{1,2}, Ana Gonçalves^{2,3}, Márcia E. Oliveira^{2,3}, Esmeralda Neves^{3,4}, Laura Marques^{3,5}, Ana Maria Fortuna^{3,6}, Rosário Santos^{2,3}

¹ Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Porto, Portugal; ² Unidade de Genética Molecular, Centro de Genética Médica Doutor Jacinto Magalhães, Centro Hospitalar Universitário do Porto, Porto, Portugal; ³ Unidade Multidisciplinar de Investigação Biomédica, Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto - UMIB/ICBAS/UP; ⁴ Serviço de Imunologia, Departamento de Patologia, Centro Hospitalar Universitário do Porto, Porto, Portugal; ⁵ Unidade de Infecologia Pediátrica e Imunodeficiências, Serviço de Pediatria, Centro Hospitalar Universitário do Porto, Porto, Portugal; ⁶ Unidade de Genética Médica, Centro de Genética Médica Doutor Jacinto Magalhães, Centro Hospitalar Universitário do Porto, Porto, Portugal



INTRODUCTION

Primary immunodeficiencies (PIDs) are a heterogeneous group of inherited diseases that affect the development or function of the human immune system. These diseases cause susceptibility to infections, autoimmunity/autoinflammation and malignancies^[1]. The field of PIDs has evolved rapidly in recent years, with over 300 different disorders and gene defects identified so far^[1]. Early and efficient diagnosis of these manifestations is essential to reduce the disease-associated mortality. Due to the great genotypic and phenotypic variability among PIDs, selection of candidate genes for testing becomes difficult^[1]. Next-generation sequencing (NGS) allows simultaneous screening of multiple candidate genes and is a comprehensive approach for PID diagnosis. The main goal of this work is to design and validate an efficient NGS approach for diagnosis of PIDs.

MATERIAL AND METHODS

Design | In order to develop a cost-effective NGS approach for PID diagnosis, 3 gene panels targeting different phenotypic categories were designed, as patients are distinguishable on an immunophenotypic basis. The panels target most of the genes known to be associated with PIDs.

Validation | Validation experiments are done through double-blind assays using control samples (previously molecularly characterized patients) for the three panels. Here are presented the preliminary results for validation of panel A, using 2 control samples.

Panel design

- PID categories divided in 3 panels (Table 1)
- Total of 283 genes targeted

Next-generation sequencing

- Library preparation → 2734 primers divided into 2 pools
- Template preparation (clonal amplification of the library)
- Sequencing on Ion GeneStudio S5 system (Thermo Fisher Scientific)

Variant analysis

- Sequence alignment with reference sequence (Ion Torrent Suite)
- Variant calling and annotation (Ion Torrent Suite, Ion Reporter)
- Variant filtering
- Prediction of variant effects
- Visual inspection of candidate variants using BAM File (Alamut Visual)

Table 1. PID categories included in the gene panels.

Panel	Categories	Number of genes
Panel A	Severe combined immunodeficiencies	117 genes
	Combined immunodeficiencies with syndromic features	
	Antibody deficiencies	
Panel B	Congenital defects of phagocyte	99 genes
	Defects in intrinsic and innate immunity	
Panel C	Complement deficiencies	68 genes
	Diseases of immune dysregulation	
	Autoinflammatory disorders	

RESULTS

Coverage analysis was performed for the 2 assays:

- **In-silico coverage:** 99%

Two previously identified variants were successfully detected in the two control patients (C1 and C2).

Control 1

A hemizygous variant **c.1581_1584del** was found in exon 16 of the gene **BTK** (Ref. seq. NM_00061.2). This is a frameshift variant that creates a premature stop codon → p.(Cys527Trpfs*2)




Fig. 1. Binary alignment map (BAM) file showing the c.1581_1584del variant in BTK gene in control patient 1.

X-linked Agammaglobulinaemia

Control 2

A homozygous variant **c.4081dup** was found in exon 33 of the gene **ATM** (Ref. seq. NM_001351834.1). This is a frameshift variant that creates a premature stop codon → p.(Ser1601Lysfs*4)


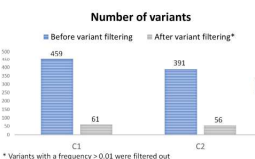


Fig. 2. Binary alignment map (BAM) file showing the c.4081dup variant in ATM gene in control patient 2.

Ataxia-telangiectasia

Number of variants



1 strong candidate variant for each control

* Variants with a frequency > 0.01 were filtered out

CONCLUSIONS

- Coverage results were very satisfactory, given the similarity with the *in-silico* coverage prediction.
- We successfully identified the disease-causing variants in the 2 patients used as controls.
- The variant found in **BTK** gene in control 1 is associated with Agammaglobulinaemia (X-linked recessive inheritance)^[1], which is part of **Antibody deficiencies** category.
- The variant found in **ATM** gene in control 2 is associated with Ataxia-telangiectasia (autosomal recessive inheritance), which is part of **CID with syndromic features** category.
- NGS showed to be an efficient approach for variant detection in these 2 cases.

FUTURE PERSPECTIVES

- More characterized patients will be assessed using this gene panel in order to determine the specificity of this methodology.
- Once this approach is validated, its clinical sensitivity will be tested by assessing undiagnosed patients.
- An efficient and rapid diagnosis of the underlying genetic defects is essential to establish the proper treatment and improve the prognosis of PID patients.
- In order to determine the sensitivity of the panel in post-hematopoietic stem cell transplant patients, screening of these patients will be performed on specific tissues.

REFERENCES

[1] Bouwffaa A, et al. The 2017 IUIS Phenotypic Classification for Primary Immunodeficiencies. *J Clin Immunol*. 2018. doi:10.1007/s10075-017-0465-8

[2] Bao M, et al. Clinical efficacy of a next-generation sequencing gene panel for primary immunodeficiency diagnosis. *Clin Genet*. 2019. 193(1):647-655. doi:10.1111/cgg.13163

[3] Moens LN, et al. Diagnosis of primary immunodeficiency disease: A sequencing capture approach. *PLoS One*. 2014. doi:10.1371/journal.pone.0114901

[4] Al-Akhouli H, et al. Unbiased targeted next-generation sequencing molecular approach for primary immunodeficiency disease. *J Allergy Clin Immunol*. 2016. 137(6):1780-1787. doi:10.1016/j.jaci.2015.12.1310

[5] Conley ME, et al. Screening of genomic DNA to identify mutations in the gene for Bruton's tyrosine kinase. *Hum Mol Genet*. 1994. 3(10):1751-6. doi:10.1093/hmg/3.10.1751

ACKNOWLEDGMENTS

UMIB is funded by FCT Strategic Project UID/Multi/0215/2019
Project approval: DEFI, CHUP, E.P.E 2019.256(209-DEFI/219-CE)