UNIVERSITÀ DEGLI STUDI
DI GENOVA

Dottorato di ricerca in Scienze Pediatriche - XXXI ciclo

Curriculum del corso: Genetica

Development of a diagnostic protocol, mutation search, and genotype-phenotype correlation in haematological and immunological diseases by targeted resequencing using three different gene panels

Student: Dr.ssa Alice Grossi

Advisor: Dr.ssa Isabella Ceccherini

Contents

1.Introduction

1.1 Primary Immunodeficiency Disorders

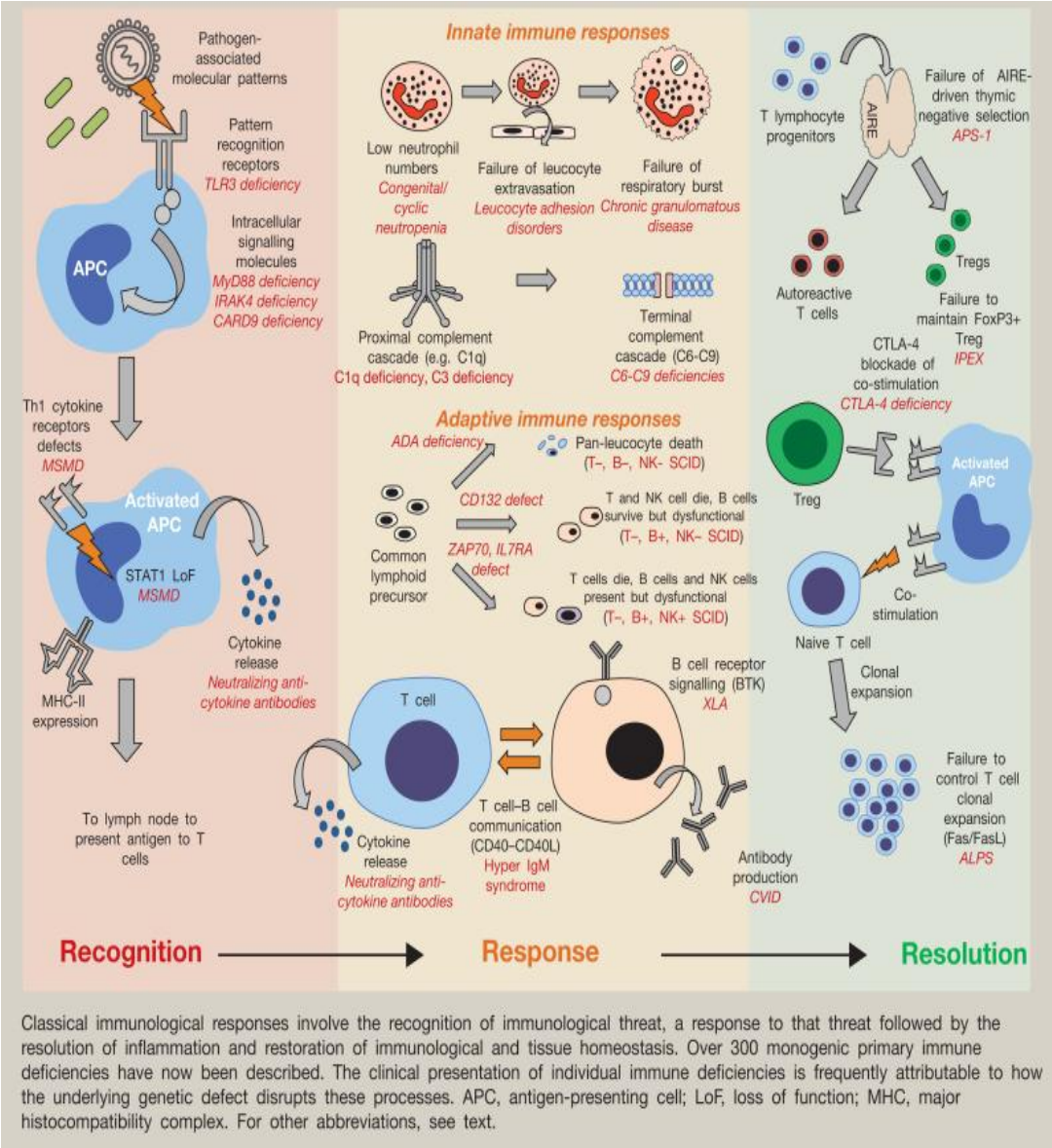
Primary immunodeficiency disorders (PID) include a large heterogeneous group of inherited diseases sharing either a poor or absent function in one or more components of the immune system. More than 350 different monogenic immune disorders and corresponding genes have been identified to date and many new others are continuously being recognized. Though rare, these disorders are chronic and severe and timely diagnosis can be lifesaving, at least for identifying the most suitable drug treatment(s) until bone marrow transplantation can take place. Moreover, even in single PID gene mutation, a genotype-phenotype correlation is often lacking, due to frequent cases of reduced penetrance and variable expressivity, as well as to a wide phenotypic heterogeneity due to allelic series.[1] Despite the variability of clinical presentations, most disorders involve autoimmunity and immune dysregulation, strongly associated with frequent infections.

With the exception of IgA deficiency (1/300-500), PIDs are more frequent than previously believed, with an estimated prevalence of 1 in 1200 live births. PIDs are essentially classified based on the component of the immune system that is involved, either the adaptive or the innate immunity (Figure 1). They are also distinct from secondary immune-deficiencies, resulting from other causes, such as viral or bacterial infections, malnutrition, treatments that induce immunosuppression or immunoglobulin loss.[2,3] In February 2017, the International Union of Immunological Societies (IUIS) established the classification of the inborn errors of immunity:[4]

- immune-deficiencies affecting cellular and humoral immunity;
- combined immune-deficiencies with associated or syndromic features;

- predominantly antibody deficiencies;
- diseases of immune dysregulation;
- congenital defects of phagocyte number or function;
- defects in intrinsic and innate immunity;
- autoinflammatory disorders;
- complement deficiencies;
- phenocopies of inborn errors of immunity.

Figure 1: Archetypal primary immunodeficiencies in the context of the classical immune response (from Shields and Patel 2017).



Classical immunological responses involve the recognition of immunological threat, a response to that threat followed by the resolution of inflammation and restoration of immunological and tissue homeostasis. Over 300 monogenic primary immune deficiencies have now been described. The clinical presentation of individual immune deficiencies is frequently attributable to how the underlying genetic defect disrupts these processes. APC, antigen-presenting cell; LoF, loss of function; MHC, major histocompatibility complex. For other abbreviations, see text.

1.2 Pathogenic mechanisms of PID

1.2.1 Disorders of innate immunity

The first line of defense against potential pathogens is represented by the innate immune system. Responses and reactions are not specific to each particular pathogen, reflecting a simple and unrefined mechanism that is, however, highly conserved in vertebrates, invertebrates and plants. The innate immune system recognizes microbes through a class of proteins found either inside or on the surface of the immune cells, termed pattern recognition receptors (PRRs), binding unique proteins of various microorganisms. A major class of PRRs are called toll-like receptors (TLRs), and they are responsible for triggering host cell expression in response to pathogens. Upon contact with these microbes, TLRs send internal messages to the nucleus of the cell to secrete cytokines, which stimulate the immune system to fight against the invading microorganisms. The innate immune system relies on the recognition of particular types of molecules that are common to many pathogens but are absent in the host.

These pathogen-associated molecules (called pathogen-associated immune-stimulants) may induce two different innate immune responses: the inflammatory responses and the phagocytosis. Both these responses can occur quickly, even if the host has not previously been exposed to a particular pathogen. Numerous cells and proteins are involved in this response, including phagocytes, dendritic cells and complement proteins. Phagocytes are primarily responsible for phagocytosis, the process by which pathogen material is engulfed and eliminate by a cell.[5]

Complement system consists of about 20 interacting soluble proteins that are synthetized mainly by the liver and circulate in the blood and extracellular fluid. Most

of them are inactive until they are triggered by an infection. Their function is to identify and opsonize foreign antigens, so that they can be eliminated. The early complement components are activated first. There are three sets of these, belonging to three distinct pathways of complement activation: the classical pathway, the lectin pathway, and the alternative pathway. The early components of all three pathways act locally to activate C3, which is the pivotal component of complement. The early components and C3 are all pro-enzymes, that are activated sequentially by proteolytic cleavage. The classical pathway is activated by IgG or IgM antibody molecules bound to the surface of a microbe. The lectine pathway is mediated by mannan-binding lectin, a serum protein that forms clusters of six carbohydrate-binding heads around a central collagen-like stalk, and binds specifically mannose and fucose residues in bacterial cell walls. In the alternative pathway, C3 is spontaneously activated at low levels and can attach to both host cells and pathogens. Host cells produce a series of proteins that prevent the complement reaction from proceeding on their cell surfaces. Because pathogens lack these proteins, they are singled out for destruction. Activation of the classical or lectin pathways also activates the alternative pathway through a positive feedback loop, amplifying their effects.

Another key-component of the innate immunity system is represented by the Natural Killer (NK) cells. Their function is to destroy virus-infected cells by inducing their apoptosis. NK cells do not express antigen-specific receptors, so they can monitor the level of class I Major Histocompatibility Complex (MHC) proteins, a sophisticated and highly polymorphic group of genes in vertebrates that code for a large family of cell-surface glycoproteins. The presence of high levels of these proteins inhibits the killing

activity of NK cells that, therefore, selectively kill cells expressing low levels class I MHC proteins, including both virally-infected cells and some cancer cells.

Defects in the development and function of anyone of these elements may lead to PIDs. Defects of phagocyte number or function involve chronic granulomatous disease (CGD), severe pyogenic bacterial infection of skin and mucosal, and leukocyte adhesion deficiency (e.g. some of the genes found with mutations in these disorders are ELANE, WAS, VPS13B, GATA2). Defects in complement system entail deficiency in early complement pathway (C1q, C1r, C2, C4), in late complement pathway (C5, C6, C7, C8, C9) and in C3 and regulatory components. Also the family of cytokines may be subject to defects, in particular the group of interleukins (ILs) and interferon (IFN-α, IFN-β and IFN-γ).[2,6]

1.2.2 Disorders of adaptive immunity

The adaptive immunity is a sophisticated defensive response, unique prerogative of vertebrates. Adaptive responses are highly specific to each particular pathogen and can also provide long-lasting protection. Being this type of defensive answer more precise but also more powerful, it is important that the system can clearly distinguish foreign antigens from the self. Any substance capable of eliciting an adaptive immune response is referred to as an antigen (antibody generator). The activity of the adaptive immune system is carried out by blood cells called lymphocytes, which include B cells producing antibody (humoral) responses and T cells involved in cell-mediated immune responses. In antibody responses, B cells are activated to secrete antibodies, which are proteins called immunoglobulins. The antibodies circulate in the bloodstream and pervade any other body fluid and district. They bind specifically the foreign antigen

thus stimulating their further production and inactivating viruses and microbial toxins. In the cell-mediated immune responses, activated T cells react directly against a foreign antigen that is presented on the surface of a host cell. The two different response systems are so intertwined in their activities so that the defect of one is often enough to entail a combined disorder: since B-cell-mediated antibody production requires intact T-cell function, most T-cell defects lead to combined (B- and T-cell) immunodeficiency disorders (CIDs).[2,6]

1.2.3 Autoimmunity

The balance between host defense and protection against self-directed immune attack is essential and depends on lymphocyte proliferation and immune tolerance. The immune system becomes tolerant to self-antigens through the process of central and peripheral tolerance. The main mechanism for the induction of central tolerance in bone marrow and thymus (for B- and T-cells, respectively) consists in the deletion of high-affinity auto-reactive lymphocytes. The main mechanisms for peripheral tolerance are anergy, antigene ignorance, deletion by apoptosis, effect of inhibitory receptors and inhibition of auto-reactive lymphocytes by T regulatory (Tregs) cells.[7] In PID patients, inflammation and persistent antigen presentation due to recurrent infections are essential mechanisms for autoimmunity.[7]

There are PIDs associated with autoimmune disease due to dysregulation of the whole immune system. Generally, lymphocytes may be present though dysfunctional, allowing for the development of excessive auto-reactivity and resulting in autoimmune disease and/or other symptoms of immune dysregulation. Autoimmune lymphoproliferative syndrome (ALPS), hemophagocytic lymphohistiocytosis (HLH) and

many other disorders belong to this category. In particular, ALPS is one of the first well-characterized human genetic disorders of the apoptosis and represents a very good example of how the improvement in genomic technologies in the latest years has led to the recognition of a large number of ALPS-like autoimmune and lymphoproliferative disorders.[2,8]

1.3 Classical diagnostics of PID

The current diagnostic approach to PIDs is dominated by time-consuming phenotypic and functional characterization. The diagnostic procedure for PIDs is a multistep process involving collection of a detailed personal and family history and data from several complex laboratory assays, thus allowing to define the immunologic defect.[9]

In the latest few years, molecular genetic testing has become an essential diagnostic tool for PIDs as it often provides a conclusive diagnosis, assists in genetic counseling, permits early prenatal diagnosis and carrier identification, determines the diagnosis in atypical cases, affords genotype-phenotype correlation, and allows pre-symptomatic identification of patients with PIDs.[10]

Sanger sequencing has played the crucial final step common to every genetic approach for many years and still represents the gold standard for DNA sequencing, the “first generation” process of reading the sequence of nucleotides present in a DNA molecule, thus confirming the presence of nucleotide variants in genes of interest.[11] Unfortunately, Sanger sequencing is not only laborious, expensive and time-consuming, but it is also not available in a diagnostic setting for many genes in different labs.

1.4 Next Generation Sequencing

The advent of the Next Generation Sequencing (NGS) has solved and overcome most problems in both diagnostic and genetic research.

In particular, in PIDs, NGS has driven the rapid increase in the number of recognizable disorders, often hampered by the wide heterogeneity of the many genetically diverse but phenotypically overlapping diseases, and has led to the discovery of new genes implicated in well-defined biological pathways, revisiting frequencies and broadening the phenotypic spectra.[4,12,13]

NGS is a revolutionary diagnostic tool for genetic investigations, allowing the simultaneous analysis of multiple genes and the effective detection of gene mosaicism. There are a variety of different NGS technological platforms making use of different sequencing chemistries.[14,15] Nevertheless, most of these share a common set of features concerning sequencing reactions such as: i) taking place in parallel, at the same time, ii) micro scaled so that a very high number of genes can be accommodated on the same chip, iii) requiring a very tiny amount of DNA per test, iv) cheaper than Sanger sequencing, v) producing shorter reads (typically 50-700 nt in length).

Till now, most clinical applications have been in diagnostic testing for hereditary disorders and, more recently, for risk screening for hereditary cancers and therapeutic decision-making for somatic cancers. The testing target has evolved from hotspot panels, actionable gene panels, and disease-focused panels to more comprehensive panels such as the targeted whole exome and the unbiased whole genome, sequencing approaches these latter that are beginning to emerge in specific cases also on a diagnostic setting. Panel-based testing is more practical at the present time, especially in small labs, and is still widely applied in clinical applications. The hotspot

panel is a collection of frequently mutated hotspots that are either clinically actionable or with diagnostic/prognostic significance. The actionable gene panels evolved from hotspot panels by including all exons of targeted genes (or all clinical relevant regions) so that other pathogenic mutations outside frequently mutated sites can be interrogated.

The disease-focused panels are comprised of the genes for a particular disease and are largely used to screen for the risk of inherited diseases, or to diagnose suspected genetic diseases. The common feature of these panels is to focus on genes known in the literature to be associated or related to the disease. Although disease-focused panels have gained popularity, clinical laboratories are facing serious financial and practical challenges associated with 1) the development and validation of different disease-focused panels according to the international guidelines; 2) the limited number of samples in need of molecular testing for any given disease at any given time; 3) the requirement to constantly update the content of existing panels. The challenges that clinicians are beginning to face today involve the choice between starting their analysis with a disease-targeted test versus jumping immediately to exome (WES) or genome (WGS) approaches. Besides cost issues, laboratories hesitate to switch to large unbiased approaches to avoid facing with the hundreds of variants with unknown clinical significance detected when WES and WGS are applied and that is why targeted testing will remain a cornerstone of the diagnostic evaluation for at least a few more years.[16,17]

Regardless of the target size, the NGS technology is based on the parallel sequencing of multiple small fragments of a given DNA target, which are ligated to proper adaptors and pooled in so-called “libraries” for the successive sequencing, rather than

on the sequencing of single fragments like in the Sanger sequencing technology. Next generation methods of DNA sequencing have therefore three main steps: (1) creation of DNA libraries including the whole target DNA, first captured in the form of DNA segments that are then ligated to custom linkers, (2) amplification of the libraries using clonal methods to separate each fragment, and (3) sequencing of each fragment of the library using one of several different chemistries.[11]

The library preparation can take place through different technologies generally based on probe hybridization to enrich sequencing libraries or based on highly multiplexed PCR reactions (Figure 2). Amplicon assays offer a slight advantage in being able to work with smaller quantities of input DNA, often down to 10ng. Hybridization assays generally require more input DNA, typically ~500 ng. Hybridization protocols start with random shearing of the DNA, followed by "capture" of the randomly sheared overlapping fragments with long oligonucleotide (oligo) baits. This allows independent sequencing of a large number of unique fragments. Any duplicates (assay artifacts) can be easily identified and removed, leaving high-quality data for analysis. Because the fragments are randomly sheared they should not align perfectly with one another and if they do, they are most certainly duplicates. Hybridization capture approach generally demonstrates better uniformity but can undergo off-target capture of sequences with high levels of repetition or low complexity (i.e., the Human Histocompatibility Locus region). Hybridization assay protocols are more time consuming and require large numbers of manual steps.

The PCR-based method is more efficient with lower amounts of DNA and has usually higher on-target rates. Tricky quality parameters of each runs, such as limited coverage, low variant frequency, and vicinity to read starts/ends, lead to a significant

number of potential false positive and false negative results. Primer competition and non-uniform amplification of target regions caused by varied GC content or amplicon length for example in the presence of an insertion (under-represented) or deletion (over-represented), contribute to variation in amplification efficiency. PCR-based approaches are usually faster with fewer steps. It is important to note, however, that PCR itself is the most common source of bias and error in any enrichment assay, so the faster protocol is ultimately balanced by the requirement for high-quality data, and the additional time required to validate potential false positive results.

Performance, namely the likelihood that the assay can detect all variants present in any region of interest, avoiding false negatives and false positives, should be a key requirement for all applications. The most common reason for false negatives in a targeted sequencing assay is poor coverage at the locus. The most common cause of false positives are artifacts introduced by PCR polymerases, even when using proofreading enzymes. Hybridization assays use very few PCR cycles, in comparison to amplicon assays, and therefore the data is less “noisy”.

Price is also a factor to be taken into consideration: for larger regions, hybrid-based panels are very convenient; for smaller regions, amplicon tests may be cheaper because the lower cost of a small number of the primers.

In any case, a limitation of targeted re-sequencing, whatever target capture procedure, is that probes and oligos are based on a reference sequence, and variations that significantly deviate from the reference, as well as large insertion/deletion mutations, are not always going to be determined.[18]

The following sequencing step can mainly take place through two technological platforms supplied by two distinct companies: the Ion Torrent Personal Genome Machine by ThermoFisher and the HiSeq 2000 by Illumina.

The Ion Torrent PGM "harnesses the power of semiconductor technology" detecting the protons released as nucleotides are incorporated during synthesis. DNA fragments with specific adapter sequences are linked to and then clonally amplified by emulsion PCR on the surface of 3-micron diameter beads, known as Ion Sphere Particles. The templated beads are loaded into proton-sensing wells that are fabricated on a silicon wafer and sequencing is primed from a specific location in the adapter sequence. As sequencing proceeds, each of the four bases is introduced sequentially. If bases of that type are incorporated, protons are released and a signal is detected proportional to the number of bases incorporated. Conversely, Illumina has adopted a sequencing-by-synthesis approach, utilizing fluorescently labeled reversible-terminator nucleotides, on clonally amplified DNA templates immobilized to an acrylamide coating on the surface of a glass flowcell.[19]

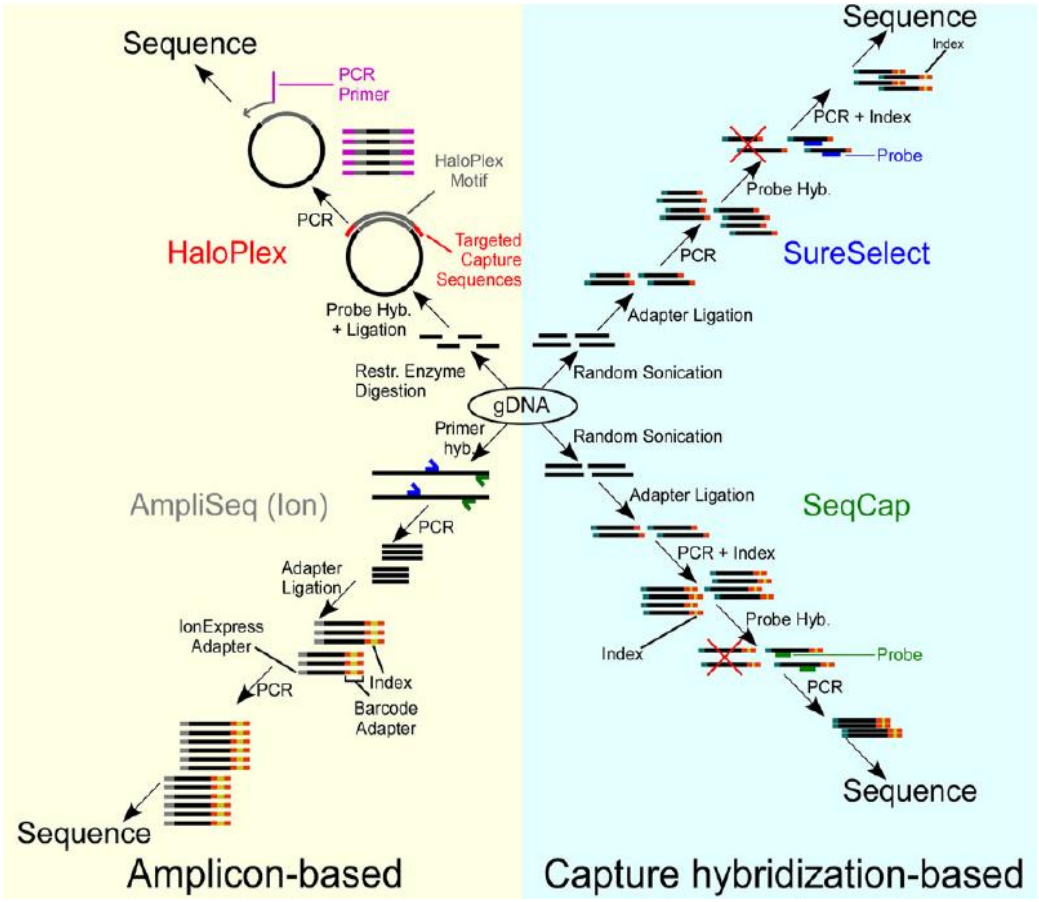


Figure 2: AmpliSeq and HaloPlex are the protocols used in this study. Noteworthy, though reported as two amplification procedures, HaloPlex uses also a hybridization step in its procedure.

2. Aim of the study

Since its development, Next-Generation Sequencing (NGS) has had outstanding implications on both clinical diagnosis and research applications to study simple Mendelian disorders, such as PIDs. In the last few years, our knowledge about PIDs etiology has largely increased resulting in the discovery of novel genes and the recognition of new diseases, a step forward the establishment of robust phenotype-genotype correlations.

The present project has focused on:

1) the identification of the underlying genetic causes of already known and unknown PIDs, particularly of bone marrow failure and autoimmune and idiopathic cytopenias and lymphoproliferations,

2) the improvement of our knowledge about clinically overlapping phenotypes through the genetic characterization of the corresponding patients, and

3) the optimization of the diagnostic work-up in order to administer disease specific treatments to patients.

Indeed, the correct identification of the pathogenic mechanisms underlying these disorders represents a challenge with many implications for effective diagnostic work-up, relevant treatment and correct follow-up.

3. Materials and Methods

3.1 Patient recruitment and DNA extraction

Patients were selected by the Hematology Unit of Istituto Giannina Gaslini based on a clinical history highly evocative of primary immunological defects, without distinction of ethnicity, age and sex. The inclusion criteria were:

- peripheral and/or central cytopenia and/or
- lymphoprolipheration and/or
- autoimmunity

All adult subjects provided written informed consent to participate to this study, while parental consent was obtained for children. A total of 149 patients (one of these analyzed for both panel 1 and panel 2 to increase chances for genetic definition) were involved in three different sets of analysis:

1. The first cohort included 51 patients that were analyzed for 146 genes associated with haemato-immuno diseases.
2. The second cohort included 69 patients that were analyzed for 315 genes associated with haemato-immuno-reumato diseases.
3. The third cohort included 30 patients that were analyzed for 58 genes associated with Bone Marrow failure and immune-dysregulations.

Forty-six of these patients had had previous genetic studies based on a candidate gene approach with no identified genetic defects.

DNA was isolated from peripheral blood samples from patients, and parents when available, and extracted by using QIAamp DNA Blood Midi kit. Quality and quantity of DNA thus obtained were determined by Nanodrop.

3.2 Enrichment/amplification design and sequencing

DNA capture probes and primers were designed based on the GRCh37/hg19 build of the human references genome. The HaloPlex HS Target Enrichment System for the Ion Torrent sequencing protocol, used for the first gene panel (Version C1, December 2016) (Figure 3A), was optimized for the digestion of 50 ng of genomic DNA splitted among 8 different restriction digestion reactions. The custom HaloPlex HS probes were designed through Agilent's SureDesign tool (www.agilent.com/genomics/suredesign).

Similarly, HaloPlex Target Enrichment System for the Ion Torrent sequencing protocol was also used for the second gene panel (Version E1, July 2015) (Figure 3B). The custom HaloPlex probes were designed through the above mentioned Agilent's SureDesign tool.

The third custom panel was designed through the Ion AmpliSeq™ Designer Tool (https://www.ampliseq.com/) that allows simultaneous multiplexed PCR amplification of thousands of genomic target regions in 2-Pool panel (Revision A.0, May 2017) (Figure 3C). This latter set of genes was ordered through the *on demand* procedure that guarantee optimized conditions of use, as each gene is pre-manufactured, tested and verified, allowing to build custom panels from over 5,000 pretested genes that are most relevant in the research of inherited germline diseases (e.g. hereditary cancer, primary immunodeficiency, hearing loss, muscular dystrophy). The genes not available in the *on demand* offer were synthesized through the *Spike-in* system, used to extend the target range to be sequenced.[20]

In each case, a list of candidate genes was submitted to the corresponding online software, with the request to design primers able to capture the coding and splice-site

regions of each selected gene.

Figure 3: workflows to prepare samples to undergo A. HaloPlex HS target-enriched sequencing; B. HaloPlex target-enriched sequencing; C. Ion Ampliseq LibraryKit Plus

For each panel, missed regions in targeted genes were covered by Sanger sequencing.

Amplicon libraries have been obtained from each DNA sample according to the

protocol specific to each corresponding gene panel. Whatever the initial capture protocol, sample libraries sequencing was carried out through the Ion Torrent™ Personal Genome Machine™ (PGM) System. Ion semiconductor sequencing utilizes the release of hydrogen ions during the sequencing reaction to detect the sequence of a cluster. Each cluster is located directly above a semiconductor transistor which is capable of detecting changes in the pH of the solution. Therefore, during nucleotide incorporation, a single H+ is released into the solution and it is detected by the semiconductor (Figure 4).

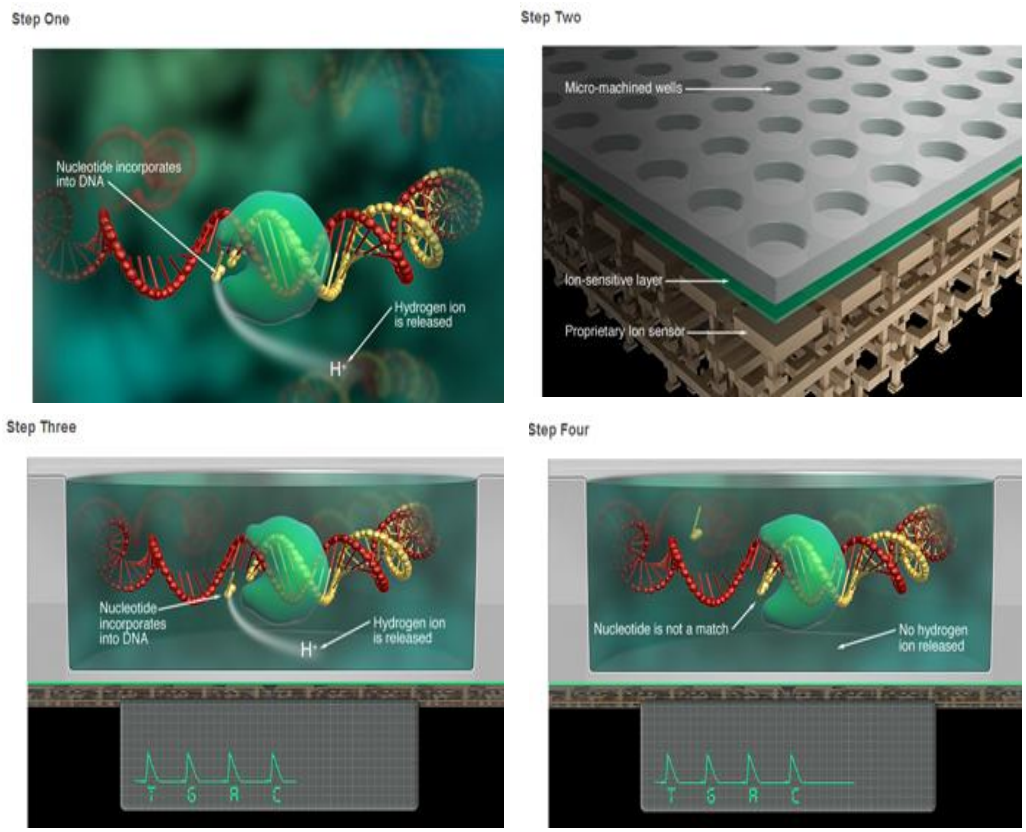


Figure 4. Step one: when a nucleotide is incorporated into a strand of DNA by a polymerase, a hydrogen ion is released as a product. Step two: Ion TorrentTM uses a high-density array of micro-machined wells to perform this biochemical process in a massively parallel way; each well holds a different DNA template; beneath the wells is an ion-sensitive layer and beneath that a proprietary ion sensor. Step three: when a nucleotide is added to a DNA template and is then incorporated into a strand of DNA, a hydrogen ion will be released: the charge from that ion will change the pH of the solution, which can be detected by the ion sensor. Step four: PGM sequencer then sequentially floods the chip with one nucleotide after another.

3.3 Bioinformatic analysis and Sanger validation

Raw data, in FastQ format, were analyzed by the Ion Reporter 5.0 (https://ionreporter.thermofisher.com/ir/). The total variants were filtered based on their frequency in the general population, as reported in the 1000 Genomes (http://www.internationalgenome.org) and Exome Aggregation Consortium (ExAC; http://exac.broadinstitute.org/) databases. Variants were assessed by the Ion Reporter™ Software 5.0 and a custom bioinformatics pipeline was additionally optimized to filter-in significant variants. In particular, variants were selected based on their frequency in the general population (lower than 5% or unreported), impact on the encoded protein (missense, stop loss and stop gain, frameshift, and splicing variants at ±2bp from the exon ends), and prediction of functional effects through *in silico* analysis using different online softwares, such as those available at https://www.varsome.org; https://www.ensembl.org/info/docs/tools/vep/index.html; https://www.ncbi.nlm.nih.gov/clinvar/; https://www.ncbi.nlm.nih.gov/projects/SNP/, as well as previous pathogenicity classification at http://genetics.bwh.harvard.edu/pph2/; http://sift.bii.a-star.edu.sg/; http://www.mutationtaster.org/; https://cadd.gs.washington.edu/.

Variants thus selected were assessed on the basis of the clinical phenotype of probands and validate by standard Sanger sequencing whenever unreported or reported as potentially damaging/damaging. To confirm the presence of the selected variants, new primers were designed by the Primer3Plus online tool (https://primer3plus.com/) and a PCR protocol was set up for each variant. PCR products were purified by ExoSAP-OT (GE Healthcare) and directly sequenced by using

Big Dye V.1.1 through an ABI3130 automated sequencer (Applied Biosystems, Foster City, California, USA). Once validated, variants segregation was finally checked in parents, when available (Figure 5).

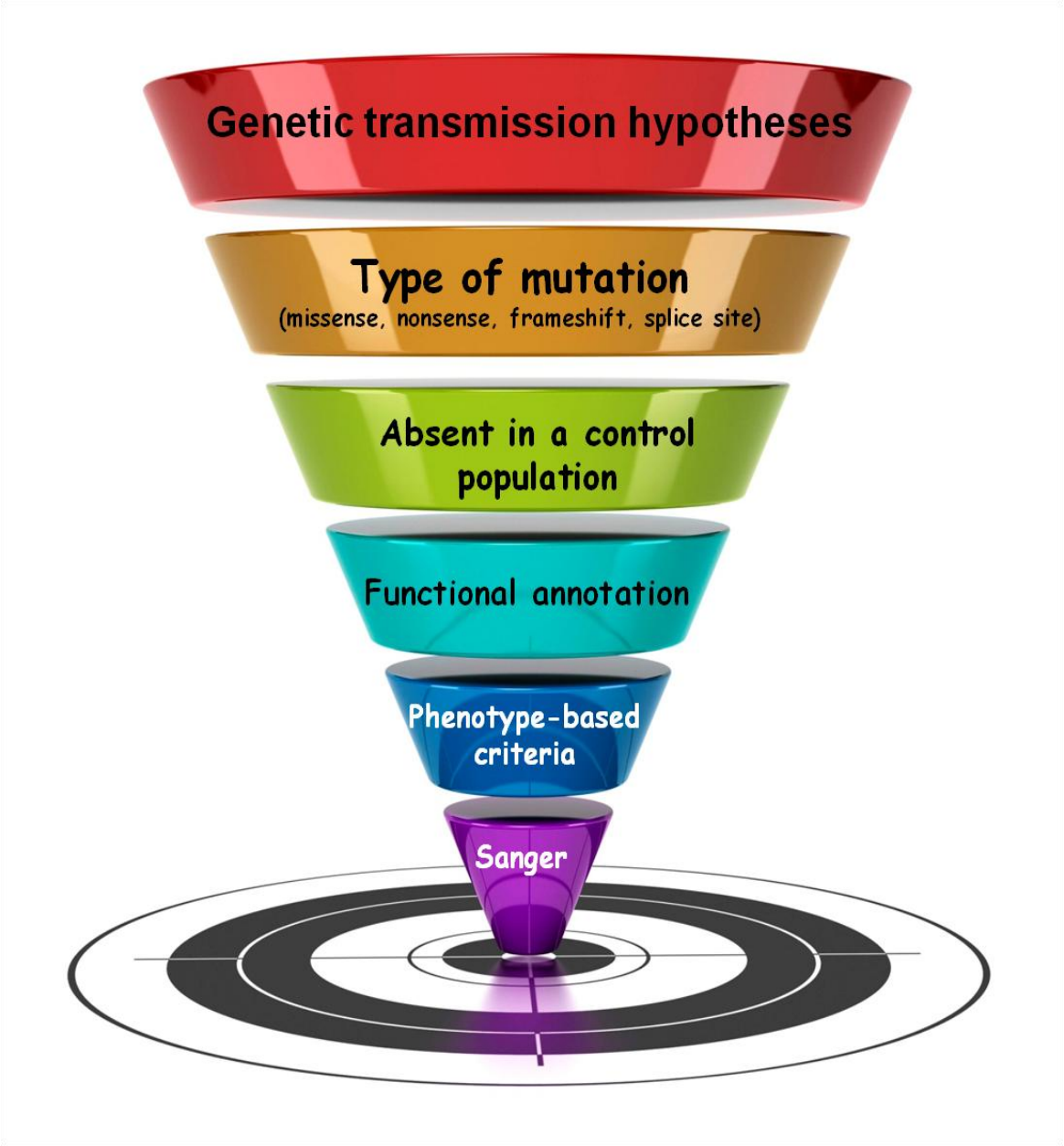


Figure 5: schematic flowchart of the filtering approach to identify potentially causative mutations.

3.4 Western Blot

B lymphocytes from four patients with variants in CASP10 gene were immortalized with the Epstein Barr virus according to the Miller protocol (1982).[21] Cell lines thus obtained were grown in culture medium Roswell Park Memorial Institute (RPMI) with antibiotics, glutamine and fetal bovine serum. The cells were then treated or not with TNF-related apoptosis-inducing ligand (TRAIL) to induce apoptosis. In parallel, a cell line from a healthy donor was treated in the same way to be used as a control. After 4 hours of treatment, cells were collected, washed and treated with a lysing solution for the production of a protein extract. For each cellular extract, proteins were quantified and equal amounts were loaded on an acrylamide gel and subjected to electrophoretic run. The proteins thus separated were transferred onto a membrane of nitrocellulose and those of interest, namely PARP-1, CASP-8, CASP-10, CASP-3 and actin, were detected with specific antibodies. Assessment of cell death rates was performed within 24 hours of the TRAIL treatment, quantifying by cytofluorimetry the number of dead cells after staining with propidium iodide.

3.5 Plasmid generation and transfection

Three PIK3CD cDNA clones were purchased from TwinHelix (www.twinhelix.eu), one to be used as control carrying the reference sequence while the two other clones carried the PIK3CD variants found in patients ID 10 and ID 2391, p.His273Tyr and p.Ser312Cys respectively. The reference sequence (NM_005026) was optimized by the company to improve its clonability. All the cDNA sequences were cloned into the pcDNA 3.1 vector

and fused in frame with C-Myc. Constructs were transferred to the HIT Competent Cells through Non-heat Shock Transformation (RBC Bioscence Corp, UK) and plasmid DNA was then extracted, after overnight growth in selective medium, through QIAprep Spin Miniprep Kit (Qiagen). A digestion with the BamHI and HindIII restriction enzymes was carried out to check the plasmid DNAs. cDNA construct DNAs were then used to transiently transfect a lymphoblastoid cell line with Lipofectamine™, according to the manufacturer's protocol (www.thermofisher.com). The activity of hyperphosphorylation of downstream proteins (AKT-mTOR pathway) was controlled by Western Blot in the transfected samples.

4. Results

4.1 Gene panel 1

4.1.1 Technical results

The first gene panel was composed of 146 genes (Table 1). The run metrics for 8 runs/51 samples using the HaloPlex HS Target Enrichment System protocol are summarized in Table 2. The total target region was represented by amplicons with an average length of 206 bases, including 337.24kbp designed out of 341.07kbp input sequence, namely all the target exons each accompanied by 10 bases of flanking regions. Therefore the target representativeness in each library undergoing NGS sequencing was 337.24/341.07 => 98.88% (see Appendix 1).

Table 1: list of genes included in the first panel

AID	AIRE	AK2	AP3B1	ATM	BCL10	BLM	BLOC1S6	BTK	C16ORF57
CARD11	CASP10	CASP8	CD19	CD20	CD21	CD27	CD3D	CD3E	CD3G
CD40	CD40LG	CD81	CD8A	CECR1	CORO1A	CSF3R	CTC1	CTLA4	CTPS1
CXCR4	CYBA	CYBB	DCLRE1C	DKC1	DOCK2	DOCK8	ELA2	EXTL3	FADD
FAS	FASLG	FOXN1	FOXP3	G6PC	GATA2	GFI1	HAX1	HOIP	ICOS
IKBKB	IL10	IL10RA	IL10RB	IL21	IL21R	IL2RA	IL2RG	IL7R	ITCH
ITK	JAGN1	JAK3	KRAS	LAMTOR2	LCK	LIG4	LRBA	LYST	MAGT1
MALT1	MAP3K14	MPL	NBN	NCF1	NCF2	NCF4	NEMO	NFKB2	NHEJ1
NOLA2	NOLA3	NRAS	ORAI1	OX40	PGM3	PIK3CD	PIK3R1	PLCG2	PNP
PRF1	PRKCD	PTPRC	RAB27A	RAC2	RAG1	RAG2	RBCK1	RFX5	RFXANK
RFXAP	RPL11	RPL26	RPL35A	RPL5	RPS10	RPS17	RPS19	RPS24	RPS26
RPS7	RTEL	RUNX1	SBDS	SH2D1A	SLC37A4	SLC7A7	SMARCAL1	STAT1	STAT3
STAT5B	STIM1	STK4	STX11	TACI	TAP1	TAP2	TAPBP	TAZ	TBX1
TERT	TINF2	TNFRSF13	TPP2	TWEAK	TYK2	UNC119	UNC13D	UNG	VPS13B
VPS45	WAS	WIPF1	WRAP53	XIAP	ZAP70				

Table 2: run metrics panel 1 (51 samples, 8 runs). In grey the details per chips, in white the average details among all the samples.

	Chip density %	Total reads per chip	Mean raw accurancy	Q20 bases	Mean read lengh	Mapped reads*	On target %*	Mean depth*	Uniformity %*
Median	81%	2823353	98	51695122	159	361531	91,17	65,77	77,45
Min	68%	1913494	91	4053953	131	22900	86,8	4,249	65,14
Max	93%	3746411	99	339587222	179	2003927	92,94	370,59	83,39

*Details available only for 5 runs on 8.

4.1.2 Cohort of patients

Fifty-one patients, already undergone conventional clinical evaluations but without a genetic confirmation of the possible diagnosis, were analyzed for the first set of genes. They were 25 male and 26 female. The average age at the time of analysis was 15.6 years. Thirteen had already undergone Sanger sequencing for at least one gene. These patients came to our attention as their clinics could be classified into one of five categories: Neutropenia diasease (n=4), Immunodeficiency (n=1), Lymphohistiocytosis HLH disease (n=1), Bone marrow failure (n=7) and ALPS-like disease (n=38).

4.1.3 Variants

Table 3 shows a summary of the genetic results thus obtained. On average, 641 variants were called for each sample. Different filters were applied to assess the variants significance. First, only exonic and splice-site variants were filter out. Excluding synonymous variants, in the coding regions only missense, stoploss, stopgain, and frameshift mutations were considered further. The third filter concerned the frequency and the predicted effect of the variants. Annotation was achieved through the use of the major databases of known variants dbSNPv.144, ClinVar, 1000Genome Browser, Varsome, VEP, and the prediction scores calculated by Polyphen-2, SIFT, Mutation Taster2, and CADD. On average, 14 variants were still deserving consideration after the filters applied, and only those believed to contribute to the phenotype were validated by Sanger sequencing and reported afterward, i.e. those affecting genes already known to be involved in the patients' respective phenotype.

Fifteen patients turned out to carry pathogenic variants that correlate with the respective clinical phenotype. In particular, five patients (IDs 5, 6, 22, 23, 53) carried

five heterozygous pathogenic variants of TNFRSF13B (two alleles p.Arg202His, one allele p.Cys104Arg, one allele p.Gln57His, and one allele p.Cys193Ter were detected),[22] each inherited by one of the parents. Each of these patients is affected by Common Variable Immunodeficiency (CVID) and the symptoms are autoimmune cytopenia, lymphoproliferation, hypogammaglobulinemia, susceptibility to infections, failure to respond to vaccinations. Four patients (IDs 11, 12, 30, 38) were found to be heterozygous for the p.His159Tyr missense variant in the TNFRSF13C gene (ID 38 presented a complex allele with a p.Pro21Arg *in cis* variant inherited by father).[23] The symptoms reported are autoimmune cytopenia, lymphoprolipheration and failure to respond to vaccinations. ID 2 carried a *de novo* PRKCD variant (p.Gly248Ser) and this correlates with the ALPS-like phenotype. Another patient (ID 32) was likely a compound heterozygous for 2 ADA2 variants (p.[Thr187Pro];[Leu188Pro]) (the father's DNA was not available, unfortunately, and the mother carried only one of the two variants) presenting a phenotype consistent with ADA2deficiency. Two patients were homozygous for gene variants responsible of autosomal recessive disorders: ID 39 carried two LRBA p.Arg655Ter alleles,[24] associated with autoimmune enteropathy, hypogammaglobulinemia, lymphoprolipheration, autoimmunity, and autoimmune hepatitis, while ID 15 carried two RAG1 p.Arg507Gln alleles and presented with immunodeficiency, susceptibility to infections, autoimmune cytopenia, Autoimmune Hemolytic Anemia (AEA), neutropenia, and thrombocytopenia. In each case, both parents were asymptomatic heterozygotes for the corresponding alleles.[25] ID 35 carried a frameshift mutation (p.Cys58fs) in the CTLA4 gene, and this consistently correlated with autoimmune cytopenia, AEA, arthritis, lymphoprolipheration, and hypogammaglobulinemia. Lastly, one patient (ID 51) carried a splice-site variant in the

ELANE gene (c.597+1G>A) that led to severe congenital neutropenia.

Four patients carried variants of uncertain significance (IDs 19, 24, 40, 45) in genes related to their respective clinical phenotypes. For two probands (IDs 24, 45) a functional test on the activity of CASP10 was performed due to clinics evocative for ALPS/ALPS-like (see below: Western Blot results).

In the remaining 32 patients we could detect 1) rare variants of uncertain significance/probably damaging but not related to their phenotype, 2) rare heterozygous variants in autosomal recessive genes or 3) absence of any significant variant. One of these patients (ID 9), lacking significant genetic variants, presented with such a complex phenotype that prompted us to perform Whole Exome Sequencing (data not available).

Table 3: results of gene panel 1.

ID	YEAR OF BIRTH	TOTAL VARIANT CALLED	FILT VAR*	VARIANTS**	#rs	PARENTAL SEGREGATION /DE NOVO***	OTHER GENETIC / FUNCTIONAL TEST
2	2003	631	11	PRKCD p.Gly248Ser	rs144320413	*de novo*	
3	2003	746	36	RNF31 p.Glu771Lys	rs142436858		
4	2006	697	10	ADA2 p.Met309Ile	rs146597836	F	
5	2013	720	18	TNFRSF13B p.Arg202His	rs104894649	M	MPL via Sanger
6	1998	592	17	TNFRSF13B p.Cys104Arg	rs34557412	M	
7	1993	657	11				
8	1998	714	13	RAG1 p.Arg332Gln; ADA2 p.Met309Ile	rs762022709; rs146597836	F M	
9	1968	865	24	STK4 p.Pro416Leu	rs33963346		TERC, TERT, TINF2 and DKC1 via Sanger; WES
11	2011	673	10	TNFRSF13C p.His159Tyr	rs61756766		MVK via Sanger
12	2002	661	19	RAG1 p.Gln242Arg; TNFRSF13C p.Gly64Val; TNFRSF13C p.His159Tyr	rs76897604; rs547352394; rs61756766	M M F	
13	1965	719	11				ELA2 via Sanger

14	2000	571	15	RAC2 p.Arg68Gln	na		
15	2009	516	16	RAG1 p.Arg507Gln HOMO	rs143969029	M, F	
17	2006	654	27	RAG1 p.Asp887Asn	rs4151034		
19	1993	497	14	IKBKG p.Glu125Lys	rs148695964	M WT, (F na)	X-inactivation test ongoing
21	2000	758	23	RAG1 p.Ser982Tyr; ADA2 p.Met309Ile	rs1245287257; rs145040665	M WT; M (F na)	
22	1995	528	12	TNFRSF13B p.Cys193Ter	rs72553885	F	
23	2001	601	20	TNFRSF13B p.Arg202His	rs104894649	F	
24	2000	607	14	CASP10 p.Val410Ile	rs13010627		Western Blot for CASP10
26	2008	620	23	RAG1 p.Arg449Lys; RAG2 p.Phe386Leu; LRBA p.Ala1090Gly	rs4151031; rs34629171; rs1782360		
27	2010	532	13				
28	2009	819	5	CD19 p.Leu285Pro	rs764208673		
29	1987	716	11	WRAP53 p.Gly481Ser	rs763828661		TERC and TERT via Sanger
30	2009	782	19	TNFRSF13C p.His159Tyr	rs61756766		
32	1990	594	11	STAT3 p.Lys658Arg (mosaicism); ADA2 p.Leu188Pro; ADA2 p.Thr187Pro	na; rs760102576; rs752890414	M WT; M; M WT (F na)	
33	1994	700	14	ADA2 p.Met243Val	rs1355940322	F	
34	1993	617	18				TERC via Sanger
35	1996	640	13	CTLA4 p.Cys58fs; LRBA p.Asp2294Asn	na; rs939898061	M M	FAS via Sanger
36	1996	549	10				Western Blot for CASP10
37	2004	602	14				
38	2001	632	15	LRBA p.Thr2686Ile; TNFRSF13C p.His159Tyr; TNFRSF13C p.Pro21Arg	rs202244838; rs61756766; rs77874543	F F F	Western Blot for CASP10
39	2009	661	8	LRBA p.Arg655Ter HOMO	rs199750191	M, F	
40	2013	534	13	CARD11 p.Arg967Cys	rs149857605	M	
41	1998	622	15				TERC via Sanger
42	2003	555	16	CD19 p.Met16Thr	rs745681190		
44	2005	648	10				
45	2008	561	13				FAS via Sanger; Western Blot for CASP10
46	2006	601	9				
47	2005	622	9	ATM p.Lys1964Glu	rs201963507		

48	1999	594	11	LYST p.Arg2624Trp	rs150306354	M	
49	2010	584	16	SLC7A7 p.Ala91Val	rs11568438		
50	1986	599	11				HAX1, G6PC3 and exon 9 of WAS via Sanger
51	1987	655	11	ELANE c.597+1	rs878855318		
52	na	600	12				
53	2013	570	14	TNFRSF13B p.Gln57His	rs149084717	M	FAS via Sanger
55	2002	730	16	ADA2 p.Lys481Asn	na		
59	2010	539	6	NCF2 p.Arg523Gln	rs139108402		MPL and TERC via Sanger
61	2004	666	10	TNFRSF4 p.Arg10Cys	rs35304565		
62	2007	690	17	RTEL1 p.Arg708Gln; LRBA p.Arg1997Cys	rs35640778; rs35879351		
65	2004	745	9				
66	1987	683	10	RUNX1 p.Leu56Ser; SMARCAL1 p.Arg499Trp	rs111527738; rs1302790588		
	MEAN	641	14				

*Filtered variants: location: exonic and splicesite; function: missense, frameshift, stoploss, stopgain; frequency: MAF≤0.05 and EMAF≤0.05. ** Variants: only validated (true positive) variants are reported; variants not validated (false positive) are not reported. *** Parental segregation: F= father; M = mother; na = not available

4.2 Gene panel 2

4.2.1 Technical results

The second gene panel was composed of 315 genes (Table 4). The run metrics for 13 runs/69 samples using the HaloPlex Target Enrichment System protocol are summarized in Table 5. The target regions was represented by amplicons with an average length of 207 bases, included 750.99kbp designed out of 769.99kbp input sequence, namely all the target exons each accompanied by 10 bases flanking regions. Therefore the target representativeness in each library undergoing NGS sequencing was 750.99/769.99 => 97.53% (see Appendix 2).

Table 4: list of second panel of genes

A20	ACP5	ACT1	ACTB	ADAR1	AICDA	AIRE	AK2	AK2	AP1S3
AP3B1	APOL1	ARPC1b	ATM	BCL10	BLM	BLNK	BLOC1S6	BOD1L1	BRCA2
BRCA1	BRIP1	BTK	C1NH	C1QA	C1QB	C1QC	C1R	C1S	C2
C3	C4A	C4B	C5	C6	C7	C8A	C8B	C8G	C9
CARD11	CARD14	CARD9	CASP10	CASP8	CD19	CD20	CD21	CD27	CD3D
CD3E	CD3G	CD3Z	CD40	CD40LG	CD46	CD59	CD70	CD79A	CD79B
CD81	CD8A	CEBPE	CECR1	CENPS	CENPX	CFB	CFD	CFH	CFHR1
CFHR3	CFI	CFP	CHD7	CIITA	COLEC11	COPA	CORO1A	CSF2RA	CSF3R
CTC1	CTLA4	CTPS1	CTSC	CXCR4	CYBA	CYBB	DCLRE1C	DKC1	DNASE1
DNASE1L3	DNASE2	DNMT3B	DOCK2	DOCK8	ELANE	ERCC4	EVER1	EVER2	EXTL3
FAAP100	FAAP20	FAAP24	FADD	FAN1	FANCA	FANCB	FANCC	FANCD2	FANCE
FANCF	FANCG	FANCI	FANCL	FANCM	FAS	FASLG	FCN3	FOXN1	FOXP3
FPR1	FUCT1	G6PC	G6PC3	GATA2	GFI1	GIMAP5	HAX1	HOIP	ICOS
IFIH1	IFNGR1	IFNGR2	IGLL1	IKAROS	IKBA	IKBKB	IKBKG	IKZF1	IL10
IL10RA	IL10RB	IL12B	IL12RB1	IL17F	IL17RA	IL1RN	IL21	IL21R	IL2RA
IL2RG	IL36RN	IL7R	IRAK4	IRF8	IRF8	ISG15	ITCH	ITGB2	ITK
JAGN1	JAK1	JAK3	KIND3	KRAS	LACC1	LAMTOR2	LCK	LIG4	LPIN2
LRBA	LYST	MAGT1	MALT1	MAP3K14	MASP1	MASP2	MCM4	MDA5	MEFV
MPL	MRE11	MTHFD1	MVK	MYD88	NBN	NCF1	NCF2	NCF4	NFKB2
NFKBID	NHEJ1	NLRC4	NLRP12	NLRP3	NLRP7	NOD2	NOLA2	NOLA3	NRAS
ORAI1	OTULIN	OX40	PALB2	PAX5	PGM3	PI3K	PIK3CD	PIK3R1	PLCG2
PMS2	PNP	POLE1	PRF1	PRF1	PRKCD	PSMA3	PSMB4	PSMB8	PSMB9
PSTPIP1	PTPRC	RAB27A	RAC2	RAD51	RAD51C	RAG1	RAG2	RASGRP1	RBCK1
RFX5	RFXANK	RFXAP	RHOH	RNASEH2A	RNASEH2B	RNASEH2C	RNF168	RPL11	RPL26
RPL35A	RPL5	RPS10	RPS17	RPS19	RPS24	RPS26	RPS7	RPSA	RTEL1
RUNX1	SAMHD1	SBDS	SEMA3E	SERPING1	SH2D1A	SH3BP2	SLC29A3	SLC37A4	SLC46A1
SLC7A7	SLX4	SMARCAL1	SP110	SPINK5	STAT1	STAT2	STAT3	STAT5B	STIM1
STK4	STN1	STX11	STXBP2	STXBP2	TAP1	TAP2	TAPBP	TAZ	TBK1
TBX1	TCF3	TCF3	TCN2	TERT	THBD	TINF2	TLR3	TMEM173	TNFAIP3
TNFRSF11A	TNFRSF13B	TNFRSF13C	TNFRSF1A	TPP2	TRAF3	TREX1	TRIF	TTC7A	TWEAK
TYK2	UAF1	UBE2T	UNC119	UNC13D	UNC93B1	UNG	USB1	USP1	VPS13B
VPS45	WAS	WIPF1	WRAP53	WDR1	XIAP	ZAP70	ZBTB24		

Table 5: run metrics panel 2 (63 samples, 13 runs). In grey the details per chips, in white the average details among all the samples.

	Chip density %	Total reads per chip	Mean raw accurancy	Q20 bases	Mean read lengh	Mapped reads	On target %	Mean depth	Uniformity %
Median	85	3306935	99,7	88714997	172	521507	97,42	42,73	75,18
Min	70	484336	99,4	1674265	122	10615	94,13	0,874	41,17
Max	94	5785648	99,8	225595060	198	1231245	98,91	110,1	99,71

4.2.2 Cohort of patients

Sixty-nine patients, already undergone conventional clinical evaluations but without a genetic confirmation of the possible diagnosis, were analyzed for this second set of genes. They were 35 male and 34 female. The average age at the time of analysis was

14 years. Twenty-four had already undergone Sanger sequencing for at least one gene and, in particular, 5 of these had also been analyzed for a custom panel of genes specific for auto-inflammatory disorders (2 for a 41 gene panel, 3 for a 11 gene panel). The 69 patients belonging to this second set came to our attention as their clinics could be classified into two categories: ALPS and ALPS-like disease (n=40), Bone marrow failure (n=10), Autoinflammatory disease (n=12), complement disease (n=1) and immunodeficiency (n=6)

4.2.3 Variants

Table 6 shows a summary of the genetic results thus obtained. The mean coverage/sample was 164,59X. On average, 1413 variants were called for each sample. After applying the filters previously described, an of average 25 variants per sample was obtained. Only those variants believed to contribute to the phenotype were validated by Sanger sequencing and reported afterward. Sixteen patients turned out to carry pathogenic variants that correlate with the respective clinical phenotype. Four of them (IDs 88, 100, 114 and 120) carried pathogenic heterozygous variants of TNFRSF13B (p.Cys104Tyr; p.Glu117fs; p.Leu69fs; p.Ser194Ter). Only the p.Glu117fs could be assessed in family members and transmission found from the father. Each of these is affected by Common Variable Immunodeficiency (CVID) showing symptoms such as autoimmune cytopenia, lymphoproliferation, hypogammaglobulinemia, susceptibility to infections, failure to respond to vaccinations. ID 38, run also in the previous panel, confirmed the already detected TNFRSF13C variants and nothing else of significant meaning came out. Another patient (ID 1176) was a compound heterozygous for 2 MVK variants (p.[Leu168_Asp170delinsHis];[Val377Ile]), presenting

with a Hyper-IgD Syndrome. One male (ID 80) was hemizygous for a X-linked IKBKG variant, inherited by the asymptomatic mother, with typical ALPS symptoms. ID 97 was homozygous for IL7R gene variants responsible for an autosomal recessive disorder (OMIM #608971: Severe Combined Immuno Deficiency). ID 2391 and ID 111 presented the same mutation in the PIK3CD gene of ID 16 for whom direct Sanger sequencing for PIK3CD gene was performed in our lab. A functional test on the activity of PIK3CD was performed also for ID 10 (see plasmid results below). They suffer from Activated PI3K-delta Syndrome (APDS), characterized by onset of recurrent sinopulmonary and other infections in early childhood. ID 109, ID 131, ID 145 and ID 64, presented variants in TMEM173, STAT3, CASP10, and RPS19 genes, respectively, already shown pathogenic in literature, even through functional studies, and leading to consistent corresponding clinic phenotypes.[26-29] In particular, STING-associated vasculopathy with onset in infancy (SAVI) is caused by TMEM173 mutation; STAT3 leads to autoimmune thrombocytopenia, lymphoprolipheration ALPS-like; ID 145 presents a typical ALPS phenotype caused by CASP10 mutation; RPS19 conduce to Anemia of Blackfan Diamond with typical clinic.

ID 90 and ID 92 carried a stopgain variant in the NHEJ1 gene and a missense variant in the CTLA4 gene, respectively, both unreported so far. ID 90 shows AEA but without signs of bone marrow failure or dysmorphic features; ID 92 exhibits a phenotype consistent with his genotype, including autoimmune enteropathy, autoimmune cytopenia, hypogammaglobulinemia, and lymphoprolipheration.

Six patients carried variants of uncertain significance (IDs 10, 58, 94, 110, 113, 2130) in genes related to their respective clinical phenotypes. In the remaining 49 patients we could detect 1) rare variants of uncertain significance/probably damaging but not

related to their phenotype, 2) rare heterozygous variants in autosomal recessive genes or 3) absence of any significant variant.

Table 6: results of gene panel 2.

PATIENT ID	DATE OF BIRTH	COVERAGE (X)	TOT VAR CALLED	FILT VAR*	VARIANTS**	#rs	PARENTAL SEGREG /DE NOVO***
10	2014	82,68	1335	21	PIK3CD p.His273Tyr; C9 p.Cys125Ter; C2 p.Asn467His	na; na; na	F; M; M
25	2005	231,25	1811	18			
38	2009	319,28	1844	38	LRBA p.Thr2686Ile; TNFRSF13C p.His159Tyr; TNFRSF13C p.Pro21Arg	rs202244838; rs61756766; rs77874543	F, F, F
54	2005	244,66	1924	21			
58	1991	262,63	1031	23	NFKBID p.Arg169Leu; WIPF1 p.Asn388His; VPS13B p.Ala3716Thr; CFP p.Arg159His	na; na; rs142476821; rs200131215	F; M; M; F HOMO
63	2010	95,24	1407	28			
64	2011	237,86	1928	33	RPS19 p.Arg62Trp	rs104894711	
71	1996	312,92	1762	28	RAG2 p.Gly509Asp	rs779267024	
72	2002	317,41	1944	22			
73	2002	165,26	996	18	RAG1 p.Arg219Gln; TNFRSF13B p.Arg122Gln; WDR1 p.Thr478Met	rs764179803; rs755343222; rs186889066	
75	2006	123,84	718	12	RAG1 p.Gln407Glu	na	M
80	1992	183,09	1225	14	IKBKG p.Glu125Lys HEMIZYGOUS	rs148695964	M
81	2004	240,29	1842	21			
82	2001	27,44	914	30	FANCA p.Ala628Thr	rs766422868	
84	2009	182,72	750	24	IFIH1 p.Arg186Cys; TNFRSF13C p.Gly64Val; MVK p.Arg388Gln	rs180843163; rs547352394; rs886048934	
85	2010	30,29	1019	35			
86	2006	95,55	1470	21	C8B p.Arg428Ter; FANCG p.Asp362Gly; ATM p.Arg2912Gly; FAN1 p.Met86fs	rs41286844: na; rs376676328; rs758406790	
87	2000	320,76	1827	21	AIRE p.Arg356Trp; BLNK p.Gly30Arg; TCF3 p.Arg158Gln	rs376901046; rs143109144; rs554419240	
88	1992	96,38	1386	22	TNFRSF13B p.Cys104Tyr; ATM p.Tyr67Cys	rs72553879; rs754033733	
90	2016	155,91	1105	16	CXCR4 p.Leu125Val; NHEJ1 p.Arg57Ter	rs1001278766; rs118204451	
92	2017	147,97	1196	18	CTLA4 p.Leu180Pro; C7 p.Arg521Ser; UNC13D p.Ile712Met	na; rs121964920; rs112245411	F, F, F
94	2000	170,57	1292	22	NCF1 p.Phe275Phe; NCF1 p.Ala308Val	na; na	
97	2016	189,83	815	28	IL7R p.Cys118Tyr HOMO	rs193922641	

99	2008	147,79	1744				
100	1989	175,37	1614	22	AIRE p.Glu517Ter; DNASE1 p.Gly127Arg; TNFRSF13B p.Glu117fs	na; rs8176919; na	F, F, F
102	1998	343,89	1961	34	PIK3CG p.Met514Val; RAG2 p.Leu279Pro; LIG4 p.Ile767Val	rs199845412; na; rs758471169	
103	1982	169,33	1509	39	TNFAIP3 p.Gly519Arg; FANCA p.Leu1138Val; FANCA p.Ala430Val	rs762149390; rs138417003; rs772567344	
105	2008	125,85	1517	26	AIRE p.Arg9Trp; AIRE p.Val484Met	na; rs367966318	
106	2000	20,54	703	25	RNASEH2B p.Ala177Thr	rs75184679	
109	2013	209,88	1603	26	TMEM173 p.Val155Met	na	
110	1993	235,65	1349	16	NLRC4 p.Arg492Trp; STAT5B p.Arg100Cys	rs1317272776; rs199894785	
111	2015	197,66	1870	28	PIK3CD p.Ser312Cys	rs61755420	
112	2008	169,25	1039	19	RAG1 p.Asn968Lys	rs193922463	
113	2002	179,23	1780	35	CASP8 p.Arg494Ter	rs1368296717	
114	na	145,43	1123	10	TNFRSF13B p.Leu69fs	rs72553875	
116	1989	131,77	816	12			
117	2013	152,04	1785	32			
120	1997	119,36	1736	33	TNFRSF13B p.Ser194Ter; DNASE1 p.Arg207Cys	rs121908379; rs148373909	
124	2009	91,23	1443	19	CHD7 p.Ser1406Arg; NOD2 p.Arg684Trp	na; rs5743276	F, M
126	2005	89,58	1549	32	CASP8 p.Val31Glu	na	
127	2016	69,35	953	22			
128	1994	87,02	1505	39			
129	2013	101,59	1158	26	TINF2 p.Pro214Ser	rs372610524	
130	2004	102,07	1507	18	PRF1 p.Asn252Ser		
131	2006	100	1450	25	STAT3 p.Arg152Trp	rs869312890	
132	2003	115,9	1607	23	C1S p.Arg534Trp	rs121909582	
133	1990	125,3	1613	20			
134	2002	137,58	1391	30			
135	2014	60,64	1095	28	SH3BP2 p.Thr531Ile	rs746860671	
144	2013	102,01	1011	27	IL21R p.Pro220Leu	rs780311714	
145	2006	125,39	743	32	TMEM173 p.Ala97Thr; TMEM173 p.Ala21Thr; CFB p.Thr400Ala HOMO; CASP10 p.Ile406Leu	na; rs140011636; na; rs80358239	
146	2004	167,81	942	17			
1030	2001	179,78	1656	21			
1176	2001	222,18	832	16	MVK p.Leu168_Asp170delinsHis; MVK p.Val377Ile	rs104895375; rs28934897	M, F
1461	2002	144,57	1565	29			
1704	1998	106,46	1242	23			
1838	2000	150,19	1701	26	NFKBID p.Pro258Leu	rs748957539	
1980	2003	77,57	932	21			
2060	2006	136,77	1721	23			
2300	2002	232,15	1834	16			

2130	na	245,27	1590	35	NOD2 p.Trp709Ter; MPL p.Arg537Gln	rs776701942; rs3820551	F, F WT
2391	1999	353,35	1860	25	PIK3CD p.Ser312Cys	rs61755420	F
2421	1989	74,85	1017	25			
2584	2001	61,89	1462	26			
2724	1998	143,59	1517	31	TBK1 p.Leu508Ile	rs144424516	
2746	2000	246,19	1694	25			
2802	1999	306,77	1729	32	LPIN2 p.Ala331Ser	rs80338805	
2582	1991	206,10	1703	34	STXBP2 p.Ile74Phe	na	
2896	2000	236,98	1810	26	GATA2 p.Ala286Pro	rs775661802	
	MEAN	164.59	1413	25			

*Filtered variants: location: exonic and splicesite; function: missense, frameshift, stoploss, stopgain; frequency: MAF≤0.05 and EMAF≤0.05. ** Variants: only validated (true positive) variants are reported; variants not validated (false positive) are not reported. *** Parental segregation: F = father; M = mother; na = not available

4.3 Gene panel 3

4.3.1 Technical results

The third gene panel was composed of 58 genes (Table 7). The run metrics for 3 run/30 samples using the Ion Ampliseq Library Kit Plus protocol are summarized in Table 8. The target region was covered by a total of 1035 amplicons (average length 202 bases) including 197.35 kbp designed out of 209.54 kbp of the input, namely target representativeness in each library undergoing NGS sequencing was 197.35/209.54 => 94.18% (see Appendix 3).

Table 7: list of third panel of genes

AP3B1	C16orf57	CARD11	CASP10	CASP8	CD19	CD20	CD40	CD40L	CSF3R
CTC1	CTLA4	CXCR4	DKC1	ELA2	FADD	FAS	FASL	G6PC	GFI1
HAX1	ITK	JAGN1	KRAS	LAMTOR2	LRBA	LYST	MAGT1	NHP2	NOP10
NRAS	PIK3CD	PIK3R1	PRKCD	RAB27A	RAC2	RPL11	RPL26	RPL35A	RPL5
RPS10	RPS17	RPS19	RPS24	RPS26	RPS7	RTEL1	SBDS	SLC37A4	TAZ
TERT	TINF2	TNFRSF13B	TNFRSF13C	VPS13B	VPS45	WAS	WRAP53		

Table 8: run metrics panel 3 (30 samples, 3 runs). In grey the details per chips, in white the average details among all the samples.

	Chip density %	Total reads per chip	Mean raw accurancy	Q20 bases	Mean read lengh	Mapped reads	On target %	Mean depth	Uniformity %
Median	87	2550372	99,4	44120544	200	278374	92,45	214,92	97,29
Min	85	684791	99	15053101	171	84965	83,99	69,51	95,45
Max	91	3673701	99,7	114146579	205	1406575	97,3	568,3	97,84

4.3.2 Cohort of patients

Thirty patients, already undergone conventional clinical evaluation but without a genetic confirmation of the possible diagnosis, were analyzed for this third set of

genes. They were 16 male and 14 female. The average age at the time of analysis was 14 years. Nine of them had already undergone Sanger sequencing for at least one gene. These patients came to our attention as their clinics could be classified into 3 categories: ALPS-like disease (n=19), Bone marrow failure (n=6), Neutropenia (n=5).

4.3.3 Variants

Table 9 shows a summary of the results of the third gene panel. The mean coverage/sample was 208.30X. On average, 127 variants were called for each sample. The filters applied were those previously described , obtaining an average of 4 variants to follow-up per sample. Indeed, only those believed to contribute to the phenotype were validated by Sanger sequencing and reported afterward. Two patients turned out to carry pathogenic variants that correlate with the respective clinical phenotype and reported in literature: ID 157 showed a variant in CASP10 (p.Ile406Leu) with typical ALPS phenotype while ID 162, studied for bone marrow failure, had a variant in TNFRSF13B (p.Cys104Arg) but not presented CVID clinical features. Four patients carried variants of uncertain significance (IDs 143, 147, 163, 175) in genes related to their respective clinical phenotypes, though variable expressivity led to lack of some symptom and limited clinical spectra, like in the case of ID 163 presenting with neutropenia but without dysmorphic feature. ID 139 and ID 178 showed the same splice-site mutation (c.258+2T>C) in heterozygosity in the SBDS gene, already reported as a susceptibility variant to aplastic anemia (OMIM #609135) and inconsistent with their ALPS-like phenotype. In the remaining 22 patients we could detect 1) rare variants of uncertain significance/probably damaging but not related to their phenotype, 2) rare heterozygous variants in autosomal recessive genes or 3) absence

of any significant variant.

Table 9: results of gene panel 3.

PATIENT ID	YEAR OF BIRTH	COVERAGE (X)	TOT VAR CALLED	FILT VAR*	VARIANTS**	#rs	OTHER GENETIC/ FUNCTIONAL TEST
EI139	1996	256,54	174	11	LYST p.Arg988Gln; LRBA p.Pro644Ser; SBDS c.258+2T>C	rs150953050; na; rs113993993	
EI143	2011	325,67	163	11	GFI1 p.Pro107Ala; CTLA4 p.Met90Val	rs149914857; rs370443546	CTLA4 via Sanger
EI147	na	107,24	133	2	WAS p.His180Asn HOMO	rs145040665	WAS via Sanger
EI154	2000	143,83	137	1			
EI155	2015	133,09	124	4			
EI156	2002	69,19	100	4			
EI157	1998	124,97	126	5	CASP10 p.Ile406Leu; CASP8 p.Lys207Arg;	rs80358239; rs148697064	
EI158	2007	162,35	105	2			
EI159	2010	118,54	134	3	G6PC p.Thr267Met;	rs145296477	
EI160	2010	96,14	125	2			
EI161	2013	156,02	125	2			
EI162	2016	152,58	106	4	TNFRSF13B p.Cys104Arg;	rs34557412	
EI163	2007	166,30	130	3	VPS13B p.Asp3057Tyr; VPS13B p.Ala3716Thr	rs140095832; rs142476821	
EI164	1995	129,31	118	3	CASP10 p.Pro501Leu	rs148939095	
EI165	1996	123,36	130	5	CTC1 p.Gly414Ala	rs62624978	
EI166	2003	118,19	136	4			
EI167	2003	100,16	136	6	LRBA p.Lys2298Glu; LYST p.Arg988Gln;	rs950337550; rs150953050	
EI168	1989	138,00	96	5	WRAP53 p.Gly521Trp	rs967111874	
EI169	2012	176,69	111	4			
EI170	1998	158,37	121	3			
EI171	2003	98,37	122	4			
EI172	2002	174,63	133	4	LRBA p.Arg2862Cys	rs145709687	
EI173	1991	165,17	137	6	CASP8 p.Lys207Arg	rs148697064	
EI174	2007	82,38	134	6	AP3B1 p.Val315Ala	na	
EI175	2002	544,50	114	2	TERT p.Gly406Arg	rs866101734	
EI176	2001	458,13	134	3	WAS p.Glu131Lys	rs146220228	
EI178	2011	500,03	111	5	FAS p.Thr319Ile; SBDS c.258+2T>C	rs372459755; rs113993993	
EI179	2004	483,69	145	6			
EI180	2005	363,85	108	3	USB1 p.Ile171Thr	rs149725439	
EI181	na	421,80	144	3			
	MEDIAN	208,30	127	4			

*Filtered variants: location: exonic and splicesite; function: missense, frameshift, stoploss, stopgain; frequency: MAF≤0.05 and EMAF≤0.05. ** Variants: only validated (true positive) variants are reported; variants not validated (false positive) are not reported. na = not available

4.4 Western Blot results

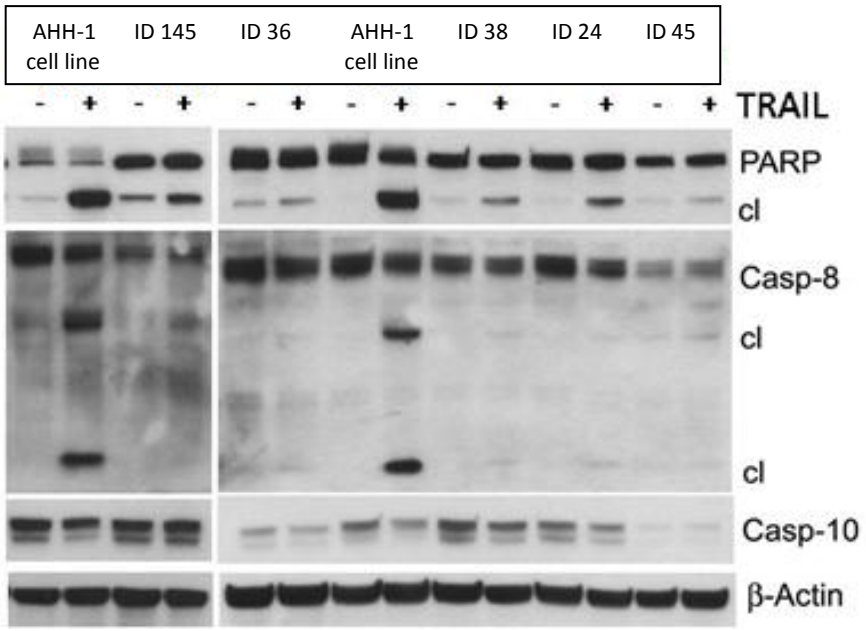
Among patients enrolled, four (1 males and 3 females) showed different genotypes for common variants in CASP10 and CASP8 genes associated with suggestive features of definitive/probable ALPS or ALPS-like. Lymphoblastoid cell lines from these four patients were used to deepen into the genotype-phenotype correlation, namely on the possible involvement of CASP-genes in the development of ALPS and ALPS-like clinics.

Four hours after cell treatment with TRAIL, Western blot analysis of apoptosis cascade proteins showed a lack of cleavage of CASP8 and PARP proteins, supposed to result in defective apoptosis, as observed in the positive control/patient ID145 who carries the CASP10 mutation p.Ile406Leu, a pathogenic mutation. Consistently, in the same test the AHH-1 cell line, used as healthy control, did not show any cleavage defect (Figure 6). We have also calculated genotype and allele frequency of these variants in our cohort (n=149) and found they are in line with the frequencies reported on the ExAC Browser (Table 10).

Table 10: genotype and allele frequency of CASP10 and CASP8 common variant

GENE	COMMON VARIANT		GENOTYPE FREQUENCY (%)			ALLELE FREQUENCY (%)		VARIANT ALLELE FREQUENCY EXAC %
CASP10	c.1228 G>A	p.V410I	133 GG (0,89)	16 GA (0,11)	0 AA (0)	282 G (0,95)	16 A (0,05)	0,04
	c.1337 A>G	p.Y446I	142 AA (0,95)	7 AG (0,05)	0 GG (0)	291 A (0,98)	7 G (0,02)	0,03
	c.1564 T>A	p.L522I	59 TT (0,40)	59 TA (0,40)	31 AA (0,20)	177 T (0,60)	121 A (0,40)	0,41
CASP8	c.2 T>C	p.M1T	138 TT (0,92)	10 TC (0,07)	1 CC (0,01)	286 T (0,97)	11 C (0,03)	0,04
	c.41 A>G	p.K14R	24 AA (0,16)	76 AG (0,51)	49 GG (0,33)	124 A (0,42)	174 G (0,58)	0,67
	C.1030 G>C	p.D344H	118 GG (0,79)	28 GC (0,19)	3 CC (0,02)	264 G (0,89)	34 C (0,11)	0,09

Figure 6: Western Blot analysis of apoptotic cascade proteins after treatment with TRAIL for 4 patient and relative genotype for CASP10 and CASP8 common variants.

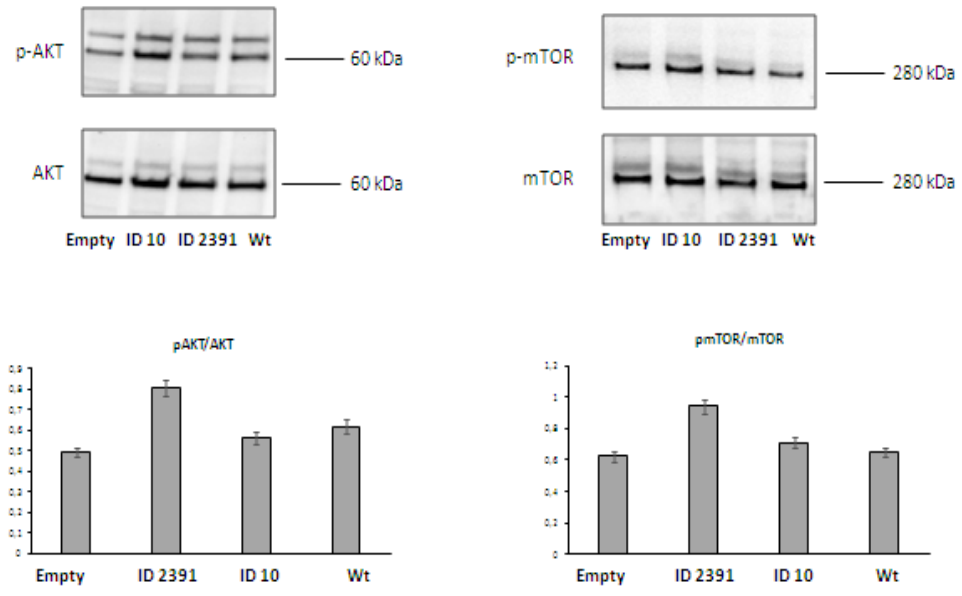


SNP p.V410I	GG	GG	GA	GG
SNP p.Y446I	AA	AA	AA	AG
SNP p.L522I	TA	TA	TA	AA
SNP p.M1T	CC	TT	TT	TT
SNP p.K14R	GG	AG	GG	GG
SNP p.D344H	GG	GG	GC	GC

4.5 Plasmid results

Four PI3KCD gene mutations have been known so far to impair T and B-cells homeostasis through an AKT-mediated hyperactivation of mTOR (OMIM #602839: p.Glu1021Lys, p.Asn334Lys, p.Glu525Lys, p.Cys416Arg). This causes APDS characterized by impaired immunoglobulin production, respiratory infections, gut/pulmonary infiltrates, and lymphoproliferation.[30] Among others, here we describe two patients harboring two novel PI3KCD mutations whose dominant clinical feature is the erythroid and myeloid precursors Marrow Failure (MF), respectively. Therefore, the activity of hyperphosphorylation of downstream proteins (AKT-mTOR pathway) has been checked by Western Blot in *in vitro* transfected samples (Figure 7). Data show an increase in protein phosphorylation in the PI3K/AKT/mTOR pathway for the plasmid carrying the p.S312C (ID 2391) mutation, compared to the wild type and to the plasmid carried the p.H273Y (ID 10) mutation.

Figure 7: Western Blot of AKT and mTOR proteins phosphorilation



5. Discussion

Primary Immunodeficiency Disorders (PIDs) are clinically heterogeneous disorders that arise from genetic defects in genes involved in immunity. The clinical effects of mutations in PID genes extend well beyond susceptibility to infection with bacteria, virus and opportunistic organisms. Immune dysregulation phenotype of PID are common and include multiorgan autoimmunity, haematological malignancy and autoinflammatory conditions. These pathologies can coexist and are often seen in combination. Furthermore, different mutations in the same gene can lead to different PID presentations, depending on whether the net effect is gain or loss of function at the protein level or on the affected domain. Most PIDs have a simple Mendelian inheritance and cause symptoms in early life, while adult presentations tend to reflect polygenic diseases, such as common variable immunodeficiency disorder (CVID) or diseases in which an environmental component reveals the underlying immunological phenotype. The wide phenotypic heterogeneity of PIDs, with often blurred and overlapping phenotypes between different clinical entities, has been preventing an effective genetic definition, including genotype-phenotype correlation, and ultimately a diagnostic assessment for many of these disorders. This has prompted us, in collaboration with the Hematology Unit and the Rheumatology Unit of IRCCS Istituto Giannina Gaslini, to develop and validate different NGS based gene panels for such haematological and immunological diseases, exploring the effectiveness and reliability of different gene combinations and various protocol options, in patients presenting with tricky phenotypes.

Given the possibility to investigate multiple genes at the same time, the advent of NGS has revolutionized clinical immunology, by allowing detailed characterization of the genetic architecture of the immune system in patients with significant and complex immunological defects, increasing our knowledge on the pathogenesis of these genetic disorders (Figure 8).[3]

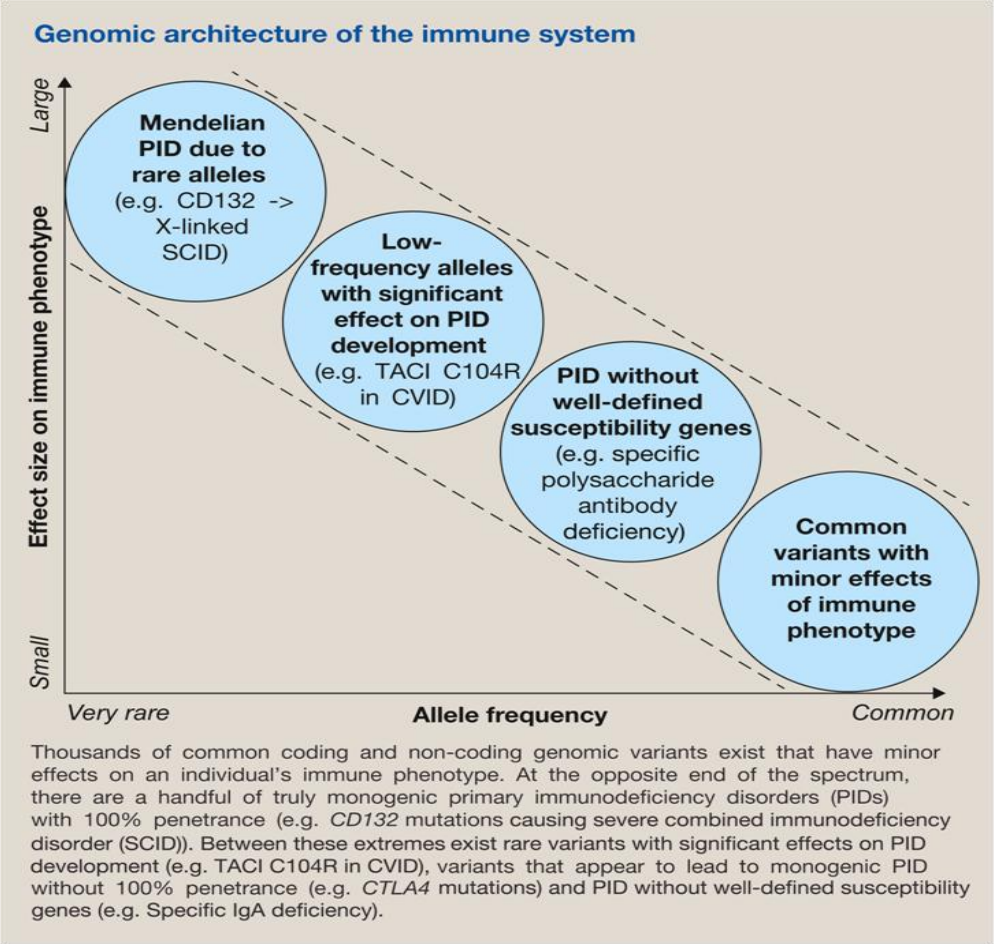


Figure 8: genomic architecture of the immune system (from Shields and Patel 2017).

In this study, we used NGS technologies to identify potential disease-causing mutations in patients affected with clinical phenotypes highly suggestive of PID, many of them yet undiagnosed after using traditional Sanger sequencing protocol, and to

genetically define the overlap between the different disorders. Some cases might represent phenotypic expansion of a known disorder to include immunodeficiency. For some autosomal dominant traits, we noted reduced penetrance associated with variant alleles, while for some autosomal recessive traits, we found predisposing variants in outbred populations. Technically, the procedure has excellent performance in evaluating large genes, features the possibility to screen up tents of patients in a single run (i.e. 22 patients for the third panel, the smallest) and its versatility allows the gene panel to be updated timeliness, depending on the selected cohorts and the new genes discovered, as we could experience in these latest three years.

One key difficulty in the design of these panels is defining the appropriate list of genes for a given phenotype because of the remarkable variability in clinical presentations. If a panel is designed to focus on certain PID phenotypes, there is a high probability that atypical presentations will be missed. On the other hand, a larger panel can lead to identify variants never seen before and / or in genes that are still little known.

Results achieved in the present study illustrate the effectiveness of the assays developed for detecting point mutations in known PID genes. The mutation rates of these panels were 15/51 (29%), 16/68 (23%) and 2/30 (6%) from the first to the third ones, respectively. The disorders that had a more accurate diagnosis were ALPS/ALPS-like (13/15; 12/16; 1/2), followed by BMF (1/15; 1/16; 1/2). Lastly, one diagnosis for neutropenia (panel 1), one for immune-deficiency and two for autoinflammatory diseases (panel 2) were also made. Albeit with different efficiencies, most of the clinical categories included in the study had a genetic diagnosis among the sequenced genes. The demand for NGS-based testing has grown rapidly without a corresponding

increase in the rate of detection of causative mutations, a circumstance that has had a strong impact on the yield of NGS panels, in terms of the proportion of patients that have been diagnosed through the molecular approach. Indeed, in the literature, the yield of the most focused disease panels varies from a higher yield (40–50%) to a lower yield (15–25%), likely depending on the phenotype and genetic heterogeneity of the included disorders, as well as, for recessive diseases, on the level of consanguinity of parents (correlated to the degree of inbreeding of the populations which patients belong to). For instance, different targeted gene panels for PID including between 162 and 173 genes allowed a definitive diagnosis in 15–25% in a total of 285 patients,[10,12,13] while the diagnostic yield was much more heterogeneous in patients with epilepsy ranging from 10 to 48.5% using panels including from 35 to 265 genes.[31]

As also reflected in the literature, in this study variants of unknown significance in potentially causative genes were found in 14 patients (4/51; 4/69; 6/30). For some of these, we attempted to study the effect of the variant through functional tests and to classify them all.

Excluding ALPS-related disorders with known genetic mutations, our clinical hematologists arbitrarily considered ALPS-like, and addressed to the genetic analysis all the conditions in which patients had symptoms and laboratory alterations similar to ALPS patients, with no definitive characterization and/or incomplete diagnosis, according to 2009 NIH revised diagnostic criteria.[32] Indeed, our preliminary study shows that all the CASP10 gene variants observed, even known as polymorphic or uncertain, similarly to mutations already established as pathogenic and regardless of the clinical phenotype, when tested by Western Blot resulted in defective cleavage of

CASP8 and PARP proteins. This contradictory observation may be explained by hypothesizing that specific CASP10 mutations have an incomplete penetrance, and/or possible epistatic effects on other genes and/or unknown genetic and epigenetic factors that would be infact the primary effectors of FAS-mediated apoptosis.[33] Moreover, the allele frequency of these variants estimated in our affected population is identical to the one known for the healthy population: this allows us to conclude that these polymorphisms neither predispose to pathology nor do they have a protective effect.

Similarly, we investigated also the variants of the PI3KCD gene found among our patients. Patient ID 2391 (and also ID 111) carried the mutation p.S312C of the PI3KCD gene. Clinically, presented an acute and very severe Pure Red Cell Aplasia (PRCA) characterized by life-threatening haemoglobin levels (2.1 gr/dl), absence of retyculocytes, and hepatomegaly. Bone marrow evaluation showed severely reduced erythroid precursors at any stage of maturation. The variant is present in this patient and in her mother, who had history of severe seizures during childhood but not of anemia. The levels of AKT protein phosphorilation tested by Western-blot has been found increased compared to an age-matched healthy control. PI3Kδ protein is broadly expressed in mice brain but its role in human nervous system is still unclear, although neurological symptoms (autism and neurodevelopmental delay) were reported in APDS patients.[28] Based on these observations and on the known APDS clinical heterogeneity within family members carrying the same mutation, we are tempted to speculate that both seizures and the severe encephalopathy may also be part of the APDS phenotype. Therefore, the p.S312C mutation of the PI3KCD gene may generate a so far unreported clinical phenotype of APDS, whose most striking feature is hypo-

regenerative cytopenia. This finding widen the spectrum of genetic diseases underlying Marrow Failure (MF) in children and highlights the possible overlap between these disorders and Combined Immunodeficiencies (CIDs), as reported in other syndromes, thus outlining the need to include CID in the differential diagnosis of MF, particularly if a single-lineage is involved.

Patient ID 10 carried the mutation p.H273Y of the PIK3CD gene, which was also found in her healthy father. She presented severe neutropenia, mild anemia and thrombocytopenia and bone marrow aspirate showed absence of myeloid precursors at any stage of maturation and presence of megakaryocytes highlighting features of a Pure White Cell Aplasia (PWCA). The levels of AKT protein phosphorilation tested by Western-blot has not been found increased compared to an age-matched healthy control.

Our attention was also attracted by the considerable frequency of heterozygous cases of RAG1 variants (8/149=5.37%). Three of these fall in the Zinc Binding Domain (Zn-BD) (p. Asp887Asn; p.Asn968Lys; p.Ser982Tyr) and two in the Nonamer Binding Domain (NBD) (p.Gln407Glu; p.Arg449Lys). In particular, p.Asn968Lys is reported on Clinvar as likely pathogenic and is very close to the conserved catalytic amino acid E965, which may alter the structure of the catalytic centre and the DNA-binding capability. Notarangelo *et al* collected all the mutations of RAG1 gene and associated them with the various possible phenotypes.[34] Disease-associated missense mutations have predominantly been detected in the Zinc-Binding region of RAG1 core domain; however, when normalized to the length of each domain, a higher mutation rate is observed in the NBD, followed by the Carboxy-Terminal Domain.[34] Our suspicion is that some of these variants can have a biological meaning even if in heterozygosis and,

for the most evocative cases, there is the possibility of a null allele, which might be demonstrated by checking the gene transcript to look for loss of heterozygosity.

Among patients unsolved, as left without a genetic diagnosis, we cannot rule out the possibility of novel genetic/clinical entities, our cohort including many atypical cases. Indeed, negative cases are likely to be enriched in novel genetic cause of PIDs.

In the majority of the patients evaluated with the present targeted NGS approach, we were not able to find a genetic defect definitively explaining the clinical phenotype, a circumstance primarily attributed to the phenotypic and genotypic heterogeneity of PIDs. Moreover, the two main technical causes able to account for negative cases might be 1) defects in genes not included in our panel because not yet described in literature and/or 2) defects located in regulatory regions not sequenced by targeted panels. The second-line diagnostic tool, as possible solutions for these unsolved cases, might be represented by the use of the Whole Exome Sequencing, covering all the existing genes, or the Whole Genome Sequencing, detecting alterations in regulatory regions or structural variations of DNA.

6. Conclusion

The Next Generation Sequencing approach, applied for Primary Immunodeficiency Disorders (PIDs), has demonstrated excellent performances in the 1) evaluation of large genes and mutation detection, 2) possibility to screen up dozens of patients in a single run, despite diseases rarity, 3) overall timeliness of the gene panels, relying on continuous literature updates, and 4) definition of different disease clinical entities characterized by overlapping phenotypes.

On the other hand, due to the remarkable variability in clinical presentations, defining the appropriate list of genes for a given phenotype represents one key difficulty in the design of these panels: a small panel targeted to genes known to be involved in well-characterized patients may be effective as a primary diagnostic screening test. In our experience, however, the best results have been obtained from the widest panels, as expected being the Unit of Hematology of Istituto Giannina Gaslini a National referent for rare and still genetically undefined hematological diseases.

Based on the present work, in the near future we need to focus on the functional study of the many variants, especially those of uncertain significance, that have emerged in a massive study like ours.

References

1. Gallo V., Dotta L., Giardino G., Cirillo E., Lougaris V., D'Assante R., Prandini A., Consolini R., Farrow E.G., Thiffault I., Sauders C.J., Leonardi A., Plebani A., Badolato R., Pignata C. "Diagnostic of Primary Immunodeficiencies through Next-Generation Sequencing" *Front. Immunol.* 2016

2. McCusker C., Upton J. and Warrington R. "Primary immunodeficiency" *Allergy Asthma Clin immunol* 2018

3. Shield AM., Patel SY. "The primary immunodeficiency disorders" *Medicine* 2017

4. Picard C., Gaspar H.B., Al-Herz W., Bousfiha A., Casanova J-L., Chatila T., Crow Y.J., Cunningham-Rundkes C., Etzioni A., Franco J.L., Holland S.M., Klein C., Morio T., Ochs H.D., Oksenhendler E., Puck J., Tang M.L.K., Tangye S.G., Torgerson T.R., Sullivan K.E. "International Union of Immunological Societies: 2017 Primary Immunodeficiency Diseases committee report on inborn errors of immunity" *J Clin Immunol* 2017

5. Ochs HD., Smith E., Puck JM. "Primary immunodeficiency Diseases: a molecular and cellular approach. Third edition" *Oxford University Press* 2013

6. Alberts B, Johnson A, Lewis J, et al. "Molecular Biology of the Cell. 4th edition" New York *Garland Science* 2002

7. Azizi G., Yazdani R., Rae W., Abolhassani H., Rojas M., Aghamohammadi A., Anaya J.M. "Monogenic polyautoimmunity in primary immunodeficiency diseases" *Autoimm Reviews* 2018

8. Bride K. and Teachey D. “Autoimmune lymphoproliferative syndrome: more than a FAScinating disease” *F1000Research* 2017

9. Oliveira J.B. and Fleisher T.A. “Laboratory evaluation of primary immunodeficiencies” *J Allergy Clin Immunol* 2010

10. Al-Mousa H., Abouelhoda M., Monies D.M., al-Tassan N., Al-Ghonaium A., Al-Saus B., Al-Dhekri H., Arnaout R., Al-Muhsen S., Aes N., Elshorbagi S., Al Gazlan S., Sheikh F., Dasouki M., El-Baik L., Elamin T., Jaber A., Kheir O., El-Kalioby M., Subhani S., Al Idrissi E., Al-Zahrani M., Alhelale M., Alnader N., Al-otaibi A., Kattan R., Al Abdelrahman K., Al Breacan M., Bin Humaid F.S., Wakil S.M., Alzayer F., Al-Dusery H., Faquih T., Al-Hissi S., Meyer B.F., Hawwari A. “Unbiased targeted next-generation sequencing molecular approach for primary immunodeficiency diasease” *American Acad Allergy, Asthma & Immunol* 2016

11. Ceccherini I, Rusmini M, Arostegui JI. Genetic Aspects of Investigating and Understanding Autoinflammation. In R. M. Laxer et al. (eds.): Textbook of Autoinflammation, © Springer Nature Switzerland AG. 2019, in press (https://doi.org/10.1007/978-3-319-98605-0_2)

12. Nijman I.J., van Montfrans J.M., Hoogstraat M., Boes M.L., van de Corput L., Renner E.D., van Zon P., van Lieshout S., Elferik M.G., van der Burg M., Vermont C.L., van der Zwaag B., Jason E., Cuppen E., van Amstel J.K.P., van Giji M.E. “Targeted next-generation sequencing: a novel diagnostic tool for primary immunodeficiencies” *American Acad Allergy, Asthma & Immunol* 2013

13. Stoddard JL., Niemela JE., Fleisher TA., Rosenzweig SD “Targeted NHS: a cost-

effective approach to molecular diagnosis of PIDs" *Front Immun* 2014

14. Samorodnitsky E., Jewell B.M., Hagopian R., Miya J., Wing MR., Lyon E., Damodaran S., Bhatt D., Reeser J.W., Datta J., Roychowdhury S. "Evalluation of hybridization capture versus amplicons-based methods for Whole-Exome Sequencing" *Hum Mutat* 2015

15. Salzer U., Bacchelli C., Buckridge S., Pan-Hammarstrom Q., Jennings S., Lougaris V., Bergbreiter A., Hagena T., Birmelin J., Plebani A., Webster A.D.B., Peter H.H., Suez D., Chapel H., McLean-Tooke A., Spickett G.P., Anover-Sombke S.,Ochs H.D., Urschel S., Belohradsky B.H., Ugrinovic S., Kumarantne D.S., Lawrence T.C., Holm A.M., Franco J.L., Schule I., Schneider P., Gertz E.M., Schaffer A.A, Hammarstrom L., Thrasher A.J., Gaspar H.B., Grimbacher B. "Relevance of biallelic versus monoallelic TNFRSF13B mutations in distinguishing disease-causing from risk-increasing TNFRSF13B variants in antibody deficiency syndromes" *Blood* 2009

16. Dong L., Wang W., Li A., Kansal R., Chen Y., Chen H., Li X. "Clinical Next Generation Sequencing for Precision Medicine in Cancer" *Curr Genomics* 2015

17. HL. Rehm "Disease-targeted sequencing: a cornerstone in the clinic" *Nature Review Genetics* 2013

18. Shendure J, Ji H. "Next-Generation DNA Sequencing" Nat Biotech 2008

19. Quail MA., Smith M., Coupland P., Otto TD., Harris SR., Connor TR., Bertoni A., Swerdlow HP., Gu Y. "A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers"

BMC Genomics 2012

20. Hylans F., Manivannan M., Krishnaswami B., Zhu Y., Tian Y. et al "Developing custom Next Generation Sequencing panels using pre-optimezed assays: an integrated approach from disease research area to functionally annotated variants with Ion Ampliseq On-Demand panel"

21. Miller G. "Immortalization of human lymphocytes by Epstein-Barr virus" *Yale J Biol Med* 1982

22. Mardis E. "Next-Generation DNA Sequencing Platforms" *Annu Rev Anal Chem* 2013

23. Bogaert D.J.A, Dullaers M., Lambrecht B.N., Vermaelen K.Y., De Baere E., Haerynck F. "Genes associated with Common Variable immunodeficiency: one diagnosis to rule them all?" *J Med Genet* 2016

24. Lo B., Zhang K., Lu W., Zhang L., Zhang Q., Kanellopoulou C., Zhang Y., Liu Z., Fritz J.M., Marsh R., Husami A., Kissell D., Nortman S., Chaturvedi V., Haines H., Young L.R., o J., Filipovich A.H., Blessing J.J., Mustillo P., Stephens M., Rueda C.M., Chougnet C.A., Hoebe K., McElwee J., Hughes J.D., Karakoc-Aydiner E., Matthews H.F., Price S., Su H.C., Rao V.K., Lenardo M.J., Jordan M.B. "Patients with LRBA deficiency show CTLA4 loss and immune dysregulation responsive to abatacept therapy" *Science* 2015

25. Westermann-Clark E.[1], Grossi A.[1], Fioredda F., Giardino S., Cappelli E., Terranova P., Palmisani E., Farmer J.R., Foldvari Z., Yamazaki Y., Faraci M., Lanino E., Notarangelo L.D., Dufour C., Ceccherini I., Walter J.E., Miano M. "RAG

deficiency with ALPS features successfully treated with TCRαβ/CD19 cell depleted haploidentical stem cell transplant" *Clin Immunol* 2018 ([1] co-author)

26. Cerboni S., Jeremiah N., Gentili M., Gehrmann U., Conrad C., Stolzenberg MC., Picard C., Neven B., Fischer A., Amigorena S., Rieux-Laucat F., Manel N. "Intrinsic antiproliferative activity of the innate sensor STING in T lymphocytes" *J Exp Med* 2017

27. Milner J.D., Vogel T.P., Forbes L., Ma C.A., Stray-Pedersen A., Niemela J.E., Lyons J.L., Engelhardt K.R., Zhang Y., Tpocagic N., Roberson E., Matthews H., Versky J., Dasu T., Vargas-Hernandez A., Varghese N., McClain K.L., Karam L.B., Nah,od K., Makedonas G., Mace E., Sorte H., Perminow G., Rao V.K., O'Connel M.P., Price S., Su H., Butrick M., McElwee J., Hughes J.D., Willet J., Swan D., Xu Y., Santibanez-Koref M., Slowik V., Dinwiddie D.L., Ciaccio C.E., Saunders C.J., Septer S., Kingsmore S.F., White A.J., Cant A.J., Hamleton S., Cooper M.A. "Early-onset lymphoproliferation and autoimmunity caused by germline STAT3 gain-of-function mutations" *Blood* 2015

28. Zhu S., Hsu A.P., Vacek M.M., Zheng L., Schaffer A.A., Dale J.K., Davis J., Fischer R., Straus S.E., Boruchov D., Saulsbury F.T., Lenardo M.J., Puck J.M. "Genetic alterations in caspase-10 may be causative or protective in autoimmune lymphoproliferative syndrome" *Hum Genet* 2006

29. Angelini M., Cannata S., Mercaldo V., Gibello L., Santoro C., Dianzani I., Loreni F. "Missense mutations associated with Diamond–Blackfan anemia affect

the assembly of ribosomal protein S19 into the ribosome" *Hum Molec Genet* 2007

30. Coulter TI, Chandra A, Bacon CM, et al. "Clinical spectrum and features of activated phosphoinositide 3-kinase δ syndrome: A large patient cohort study" *J Allergy Clin Immunol.* 2016

31. Dunn P, Albury CL, Maksemous N, Benton MC Sutherland HG, Smith RA, et al. "Next generation sequencing methods for diagnosis of epilepsy syndromes" *Front Genet* 2018

32. Oliveira, J.B., Bleesing, J.J., Rao, V.K. "Revised diagnostic criteria and classification for the autoimmune lymphoproliferative syndrome (ALPS): report from the 2009 NIH International workshop" *Blood* 2010

33. Rao VK., Price S., Similuk M., Niemela J., Milner JD., Rosenzweig S. "Clinical spectrum of Autoimmune Lymphoproliferative Syndrome associated with Caspase 10 mutations" *58th ASH Annual Meeting and Expositions* 2016

34. Notarangelo LD., Kim MS., Walter JE., Lee YN. "Human RAG mutations: biochemistry and clinical implications" *Nature Reviews Immun* 2016

Appendix

Appendix 1: list of genes included in panel 1 with proportion of target covered in the design (cover.)

Target ID	Cover.	Target ID	Cover.	Target ID	Cover.	Target ID	Cover.	Target ID	Cover.
AID	100	**CTPS1**	100	**IL7R**	100	**PIK3R1**	100	**SLC7A7**	100
AIRE	100	**CXCR4**	100	**ITCH**	100	**PLCG2**	100	**SMARCAL1**	100
AK2	100	**CYBA**	100	**ITK**	100	**PNP**	100	**STAT1**	100
AP3B1	100	**CYBB**	100	**JAGN1**	100	**PRF1**	100	**STAT3**	100
ATM	100	**DCLRE1C**	100	**JAK3**	100	**PRKCD**	100	**STAT5B**	90,23
BCL10	100	**DKC1**	100	**KRAS**	100	**PTPRC**	97,54	**STIM1**	100
BLM	100	**DOCK2**	100	**LAMTOR2**	100	**RAB27A**	100	**STK4**	100
BLOC1S6	100	**DOCK8**	100	**LCK**	100	**RAC2**	100	**STX11**	100
BTK	100	**ELA2**	100	**LIG4**	100	**RAG1**	100	**TACI**	100
C16ORF57	100	**EXTL3**	100	**LRBA**	99,94	**RAG2**	100	**TAP1**	100
CARD11	100	**FADD**	100	**LYST**	99,99	**RBCK1**	100	**TAP2**	99,71
CASP10	100	**FAS**	100	**MAGT1**	99,4	**RFX5**	100	**TAPBP**	100
CASP8	100	**FASLG**	100	**MALT1**	100	**RFXANK**	93,94	**TAZ**	100
CD19	100	**FOXN1**	100	**MAP3K14**	100	**RFXAP**	100	**TBX1**	100
CD20	100	**FOXP3**	100	**MPL**	100	**RPL11**	100	**TERT**	100
CD21	100	**G6PC**	100	**NBN**	100	**RPL26**	100	**TINF2**	100
CD27	100	**GATA2**	100	**NCF1**	42,96	**RPL35A**	100	**TNFRSF13C**	100
CD3D	100	**GFI1**	100	**NCF2**	100	**RPL5**	100	**TPP2**	100
CD3E	100	**HAX1**	100	**NCF4**	100	**RPS10**	100	**TWEAK**	100
CD3G	100	**HOIP**	100	**NEMO**	46,03	**RPS17**	15,97	**TYK2**	100
CD40	100	**ICOS**	100	**NFKB2**	100	**RPS19**	100	**UNC119**	100
CD40LG	100	**IKBKB**	100	**NHEJ1**	100	**RPS24**	100	**UNC13D**	100
CD81	100	**IL10**	100	**NOLA2**	100	**RPS26**	100	**UNG**	100
CD8A	100	**IL10RA**	100	**NOLA3**	100	**RPS7**	100	**VPS13B**	100
CECR1	100	**IL10RB**	100	**NRAS**	100	**RTEL**	100	**VPS45**	100
CORO1A	95,62	**IL21**	100	**ORAI1**	100	**RUNX1**	99,94	**WAS**	100
CSF3R	100	**IL21R**	100	**OX40**	100	**SBDS**	92,88	**WIPF1**	100
CTC1	100	**IL2RA**	100	**PGM3**	100	**SH2D1A**	100	**WRAP53**	100
CTLA4	100	**IL2RG**	100	**PIK3CD**	100	**SLC37A4**	100	**XIAP**	100
								ZAP70	100

Appendix 2: list of genes included in panel 2 with proportion of target covered in the design (cover.)

Target ID	Cover.	Target ID	Cover.	Target ID	Cover.	Target ID	Cover.	Target ID	Cover.
A20	99,49	**CENPS**	100	**HAX1**	100	**NHEJ1**	100	**SERPING1**	100
ACP5	100	**CENPX**	100	**HOIP**	100	**NLRC4**	100	**SH2D1A**	100
ACT1	100	**CFB**	100	**ICOS**	100	**NLRP12**	100	**SH3BP2**	100
ACTB	61,77	**CFD**	100	**IFIH1**	100	**NLRP3**	100	**SLC29A3**	100
ADAR1	100	**CFH**	100	**IFNGR1**	100	**NLRP7**	99,77	**SLC37A4**	100
AICDA	100	**CFHR1**	95,24	**IFNGR2**	100	**NOD2**	100	**SLC46A1**	100
AIRE	100	**CFHR3**	88,54	**IGLL1**	100	**NOLA2**	100	**SLC7A7**	100
AK2	100	**CFI**	100	**IKAROS**	100	**NOLA3**	100	**SLX4**	100
AP1S3	100	**CFP**	100	**IKBA**	100	**NRAS**	100	**SMARCAL1**	100
AP3B1	100	**CHD7**	100	**IKBKB**	100	**ORAI1**	100	**SP110**	100
APOL1	100	**CIITA**	100	**IKBKG**	46,03	**OTULIN**	100	**SPINK5**	100
ARPC1B	100	**COLEC11**	100	**IKZF1**	100	**OX40**	100	**STAT1**	100
ATM	100	**COPA**	100	**IL10**	100	**PALB2**	100	**STAT2**	100
BCL10	100	**CORO1A**	95,62	**IL10RA**	100	**PAX5**	100	**STAT3**	100
BLM	100	**CSF2RA**	22,43	**IL10RB**	100	**PGM3**	100	**STAT5B**	90,23
BLNK	99,46	**CSF3R**	100	**IL12B**	100	**PI3K**	99,63	**STIM1**	100
BLOC1S6	100	**CTC1**	100	**IL12RB1**	100	**PIK3CD**	100	**STK4**	100
BOD1L1	100	**CTLA4**	100	**IL17F**	100	**PIK3R1**	100	**STN1**	100
BRCA1	100	**CTPS1**	100	**IL17RA**	100	**PLCG2**	100	**STX11**	100
BRCA2	99,97	**CTSC**	100	**IL1RN**	100	**PMS2**	74,9	**STXBP2**	98,89
BRIP1	100	**CXCR4**	100	**IL21**	100	**PNP**	100	**TAP1**	100
BTK	100	**CYBA**	100	**IL21R**	100	**POLE1**	99,89	**TAP2**	99,85
C1NH	100	**CYBB**	100	**IL2RA**	100	**PRF1**	100	**TAPBP**	100
C1QA	100	**DCLRE1C**	100	**IL2RG**	100	**PRKCD**	100	**TAZ**	100
C1QB	100	**DKC1**	100	**IL36RN**	100	**PSMA3**	100	**TBK1**	100
C1QC	100	**DNASE1**	100	**IL7R**	100	**PSMB4**	100	**TBX1**	100
C1R	100	**DNASE1L3**	100	**IRAK4**	100	**PSMB8**	100	**TCF3**	100
C1S	100	**DNASE2**	100	**IRF8**	100	**PSMB9**	100	**TCN2**	100
C2	99,06	**DNMT3B**	100	**ISG15**	100	**PSTPIP1**	99,53	**TERC**	100
C3	100	**DOCK2**	100	**ITCH**	100	**PTPRC**	97,54	**TERT**	100
C4A	21,47	**DOCK8**	100	**ITGB2**	100	**RAB27A**	100	**THBD**	100
C4B	27,75	**ELANE**	100	**ITK**	100	**RAC2**	100	**TINF2**	100
C5	100	**ERCC4**	100	**JAGN1**	100	**RAD51**	100	**TLR3**	100
C6	100	**EVER1**	100	**JAK1**	100	**RAD51C**	100	**TMEM173**	100
C7	100	**EVER2**	100	**JAK3**	100	**RAG1**	100	**TNFAIP3**	100
C8A	100	**EXTL3**	100	**KIND3**	100	**RAG2**	100	**TNFRSF11A**	100
C8B	100	**FAAP100**	100	**KRAS**	100	**RASGRP1**	100	**TNFRSF13B**	100
C8G	100	**FAAP20**	100	**LACC1**	100	**RBCK1**	100	**TNFRSF13C**	100
C9	96,03	**FAAP24**	100	**LAMTOR2**	100	**RFX5**	100	**TNFRSF1A**	98,34
CARD11	100	**FADD**	100	**LCK**	100	**RFXANK**	93,94	**TPP2**	100
CARD14	100	**FAN1**	100	**LIG4**	100	**RFXAP**	100	**TRAF3**	100
CARD9	100	**FANCA**	97,66	**LPIN2**	100	**RHOH**	100	**TREX1**	100
CASP10	100	**FANCB**	100	**LRBA**	99,94	**RMRP**	100	**TRIF**	100

CASP8	100	**FANCC**	100	**LYST**	99,99	**RNASEH2A**	100	**TTC7A**	100
CD19	100	**FANCD2**	99,42	**MAGT1**	99,4	**RNASEH2B**	100	**TWEAK**	100
CD20	99,57	**FANCE**	100	**MALT1**	100	**RNASEH2C**	100	**TYK2**	100
CD21	100	**FANCF**	100	**MAP3K14**	100	**RNF168**	100	**UAF1**	100
CD27	100	**FANCG**	100	**MASP1**	100	**RPL11**	100	**UBE2T**	100
CD3D	100	**FANCI**	100	**MASP2**	100	**RPL26**	100	**UNC119**	100
CD3E	100	**FANCL**	100	**MCM4**	100	**RPL35A**	100	**UNC13D**	100
CD3G	99,65	**FANCM**	100	**MDA5**	100	**RPL5**	100	**UNC93B1**	95,43
CD3Z	100	**FAS**	100	**MEFV**	100	**RPS10**	100	**UNG**	100
CD40	100	**FASLG**	100	**MPL**	100	**RPS17**	15,97	**USB1**	100
CD40LG	100	**FCN3**	100	**MRE11**	100	**RPS19**	100	**USP1**	100
CD46	100	**FOXN1**	100	**MTHFD1**	100	**RPS24**	100	**VPS13B**	100
CD59	100	**FOXP3**	100	**MVK**	100	**RPS26**	100	**VPS45**	100
CD70	99,52	**FPR1**	100	**MYD88**	100	**RPS7**	100	**WAS**	100
CD79A	100	**FUCT1**	100	**NBN**	100	**RPSA**	100	**WDR1**	100
CD79B	100	**G6PC**	100	**NCF1**	42,96	**RTEL1**	100	**WIPF1**	100
CD81	100	**G6PC3**	100	**NCF2**	100	**RUNX1**	99,94	**WRAP53**	100
CD8A	100	**GATA2**	100	**NCF4**	100	**SAMHD1**	100	**XIAP**	100
CEBPE	100	**GFI1**	100	**NFKB2**	100	**SBDS**	92,88	**ZAP70**	100
CECR1	100	**GIMAP5**	100	**NFKBID**	100	**SEMA3E**	100	**ZBTB24**	100

Appendix 3: list of genes included in panel 3 with proportion of target covered in the design (cover.)

Target ID	Coverage	Target ID	Coverage
AP3B1	94,51	**NOP10**	100
C16orf57	100	**NRAS**	100
CARD11	99,89	**PIK3CD**	98,44
CASP10	99,27	**PIK3R1**	99,66
CASP8	99,37	**PRKCD**	99,71
CD19	100	**RAB27A**	100
CD20	88,4	**RAC2**	100
CD40	100	**RPL11**	100
CD40LG	99,31	**RPL26**	100
CSF3R	100	**RPL35A**	100
CTC1	93,27	**RPL5**	100
CTLA4	100	**RPS10**	100
CXCR4	100	**RPS17**	99,13
DKC1	99,47	**RPS19**	100
ELA2	90,61	**RPS24**	97,85
FADD	92,77	**RPS26**	100
FAS	99,38	**RPS7**	100
FASLG	100	**RTEL1**	97,94
G6PC	98,99	**SBDS**	100
GFI1	97,52	**SLC37A4**	99,36
HAX1	99,97	**TAZ**	83,98
ITK	98,35	**TERT**	88,74
JAGN1	100	**TINF2**	100
KRAS	100	**TNFRSF13B**	89,38
LAMTOR2	88,75	**TNFRSF13C**	33,85
LRBA	99,94	**VPS13B**	96,48
LYST	99,97	**VPS45**	96,7
MAGT1	100	**WAS**	86
NHP2	100	**WRAP53**	98,8