Genetic diagnosis of autoinflammatory disease patients using clinical exome sequencing

Laura Batlle-Masó\textsuperscript{a,b}, Anna Mensa-Vilaró\textsuperscript{c,d}, Manuel Solís-Moruno\textsuperscript{a,b}, Tomàs Marquès-Bonet\textsuperscript{a,e,f,g}, Juan I. Arostegui\textsuperscript{c,d,h}, Ferran Casals\textsuperscript{b,*}

\textsuperscript{a} Institut de Biologia Evolutiva (CSIC-UPF), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Doctor Aiguader 88, Barcelona, Spain
\textsuperscript{b} Genomics Core Facility, Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Parc de Recerca Biomèdica de Barcelona, 08003, Barcelona, Spain
\textsuperscript{c} Department of Immunology, Hospital Clínic, Barcelona, Spain
\textsuperscript{d} Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain
\textsuperscript{e} Catalan Institution of Research and Advanced Studies (ICREA), Passeig de Lluís Companys, 23, 08010, Barcelona, Spain
\textsuperscript{f} CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Balmes i Reixac 4, 08028 Barcelona, Spain
\textsuperscript{g} Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Edifici ICTA-ICP, c/ Columns s/n, 08193, Cerdanyola del Vallès, Barcelona, Spain
\textsuperscript{h} Universitat de Barcelona, Barcelona, Spain

\textbf{ARTICLE INFO}

\textbf{Keywords:}
Autoinflammatory diseases
Primary immunodeficiency
Targeted sequencing
Clinical exome
Gene panels

\textbf{ABSTRACT}

Autoinflammatory diseases comprise a wide range of syndromes caused by dysregulation of the innate immune response. They are difficult to diagnose due to their phenotypic heterogeneity and variable expressivity. Thus, the genetic origin of the disease remains undetermined for an important proportion of patients. We aim to identify causal genetic variants in patients with suspected autoinflammatory disease and to test the advantages and limitations of the clinical exome gene panels for molecular diagnosis. Twenty-two unrelated patients with clinical features of autoinflammatory diseases were analyzed using clinical exome sequencing (~4800 genes), followed by bioinformatic analyses to detect likely pathogenic variants. By integrating genetic and clinical information, we found a likely causative heterozygous genetic variant in\textit{NFKBIA} (p.D31N) in a North-African patient with a clinical picture resembling the deficiency of interleukin-1 receptor antagonist, and a heterozygous variant in\textit{DNASE2} (p.G322D) in a Spanish patient with a suspected lupus-like monogenic disorder. We also found variants likely to increase the susceptibility to autoinflammatory diseases in three additional Spanish patients: one with an initial diagnosis of juvenile idiopathic arthritis who carries two heterozygous\textit{UNC13D} variants (p.R727Q and p.A59T), and two with early-onset inflammatory bowel disease harbouring\textit{NOD2} variants (p.L221R and p.A728V respectively). Our results show a similar proportion of molecular diagnosis to other studies using whole exome or targeted resequencing in primary immunodeficiencies. Thus, despite its main limitation of not including all candidate genes, clinical exome targeted sequencing can be an appropriate approach to detect likely causative variants in autoinflammatory diseases.

\section{1. Introduction}

Autoinflammatory diseases (AID) constitute a family of immune disorders characterized by recurrent episodes of sterile inflammation that may appear without a known trigger. AID are considered primary immunodeficiency diseases (PID), a wide classification that includes different subgroups, all of them caused by dysregulations of specific pathways of the immune system. Although reliable epidemiological data is only available for some of the subgroups (i.e. inflammosomopathies), AID are considered rare diseases (Wekell et al., 2017). Clinically, they are characterized by recurrent episodes of fever, systemic inflammation and other symptoms such as skin rashes, abdominal pain, chest pain, lymphadenopathy or arthritis (Wekell et al., 2016). The severity of the disease is variable among patients, supporting the observations of clinical heterogeneity that makes a definitive diagnosis difficult, especially in early-onset patients without a positive family history (Almeida de Jesus and Goldbach-Mansky, 2013).

Over the past decades, several efforts were focused on better
definition and classification of the wide variety of phenotypes and clinical manifestations included in AID group (Ciccarelli et al., 2014; Wekell et al., 2016). Many molecular mechanisms underlying these diseases have been discovered, most of them affecting genes related to the inflammasome (TNFRSF1A, NLRP3 and MEFV among others (Hoffman and Broderick, 2016)). All these advances represented a significative improvement in diagnosis rates. However, there is still a significant proportion of patients with clinical features suggestive of AID that remain without a clear molecular diagnosis (Martorana et al., 2017; Notarangelo and Sorensen, 2008).

Current massive sequencing technologies have raised the possibility of analysing the whole genome or whole exome of patients. Alternatively, in targeted sequencing studies the analysis is restricted to a set of previously defined candidate genes, thus reducing economic costs and data analysis efforts (Seleman et al., 2017). In recent years, targeted sequencing panels have become a useful tool to identify the molecular causes of autoinflammatory diseases (Karacan et al., 2019; Omoyinmi et al., 2017; Papa et al., 2019; Yao et al., 2016) and other primary immunodeficiencies (reviewed in Yska et al., 2019).

In this project we assess the suitability and performance of a gene panel approach, the TruSight One Sequencing Panel (Illumina), to detect the genetic origin in presumably monogenic AID. We first analyzed the proportion of candidate genes for autoinflammatory disorders included and properly sequenced in the panel. We discussed their applicability and its benefits or limitations compared to other approaches, such as whole genome or whole exome sequencing. Second, we performed targeted sequencing in 22 patients, and exhaustively analyzed the data to detect potentially causative variants, considering different genetic models for the disease and the available clinical information for each patient.

2. Materials and methods

2.1. Study cohort

This study includes 22 unrelated patients (named A1-A22) with severe symptoms and clinical suspicion of AID. Written informed consent to participate in the study in accordance with the Declaration of Helsinki was obtained before the extraction of blood samples. The protocol was approved by the Comité Étic d’Investigació Clínica - Parc de Salut Mar (Barcelona). Genomic DNA was extracted from whole blood using AllPrep DNA/RNA kit (Qiagen, Hilden, Germany) following the manufacturer’s instructions. DNA quality and concentrations were measured using Nanopore spectrophotometer (Thermo Fischer Scientific, Waltham, Massachusetts, USA), and Qubit (Thermo Fisher Scientific, Waltham, Massachusetts, USA) or PicoGreen (Invitrogen, Carlsbad, California, USA) fluorimetric quantification methods.

2.2. DNA library preparation and sequencing

Targeted sequencing capture was performed using the TruSight One Sequencing Panel (Illumina, San Diego, California, USA) which includes 4813 genes related to disease. After the enrichment, quality control of the pooled library was performed with the Bioanalyzer System (Agilent, Santa Clara, California, USA) and quantified with a specific qPCR for library adaptors (Kappa Biosystems-Roche, Basel, Switzerland). Sequencing was performed on a NextSeq 500 System (Illumina, San Diego, California, USA) in a 2 × 75 paired-end cycles high-output run.

2.3. Data analysis

Bcl output files were converted to Fastq and demultiplexed with bcl2fastq (Illumina, San Diego, California, USA). Mapping and alignment against the human reference genome (GRCh38/hg38) was performed with BWA. Picard (http://www.picard.sourceforge.net) was used to reorder and sort SAM files before marking PCR duplicates and replace read groups. Following the Genome Analysis Toolkit (GATK (McKenna et al., 2010)) best practices we performed base quality score recalibration and local indel realignment before the variant calling. The resulting VCF file was annotated using ANNOVAR (Wang et al., 2010).

We used VarScan2 (Koboldt et al., 2012) to analyze somatic variation and XHMM (Fromer et al., 2012; Fromer and Purcell, 2014) for copy number alterations. The study of synonymous variants was performed using the Transcript-Inferred Pathogenic Score from http://trap-score.org/about.jsp (Gelfman et al., 2017).

2.4. Genetic variant filtering and prioritizing

A first quality filter discarded variants with DP < 5 and QUAL < 30 (depth and quality given by GATK). We also excluded variants present in more than ten patients to avoid possible sequencing artifacts. We removed indels up to 10 base pairs next to another indel due to its high probability of being produced by mapping errors. Intronic variants not involved in splicing procedures were also discarded. We used the bed file provided by the manufacturer to exclude off-target regions and a public dataset to detect genes which are known to be enriched in sequencing errors (Fuentes Fajardo et al., 2012).

As shown in Fig. 1 we applied different filters considering the different groups of variants and gene properties. For loss of function variants, we applied a filter of MAF < 1% in gnomAD (Lek et al., 2016) database. For these variants, we also considered the constraint metrics of gnomAD such as the precalculated probability of being loss of function intolerant (pLI) for each gene. We retrieved information about the function of the gene and its implication in human disease using the Gene Ontology (GO (Ashburner et al., 2000; The Gene Ontology Consortium, 2019)). Missense variants were filtered by allele frequency (MAF < 1%) and GO terms, as well as conservation and deleteriousness scores such as GERP (Chong et al., 2015; Cooper et al., 2005) and CADD (Kircher et al., 2014), retaining variants with values higher than 2 and 10, respectively. Genes with missense variants were additionally annotated with their haploinsufficiency scores (Huang et al., 2010; Steinberg et al., 2015), intolerance to genetic variation (RVIS (Petrovski et al., 2015)) and essentiality scores (Khurana et al., 2013) to detect single heterozygous variants which caused a gain of function effect. This mechanism has been widely described in AID and it has been related to the variable clinical expression of the phenotype (Martorana et al., 2017). Finally, other in silico predictors of the protein effect were also used as a proxy of pathogenicity (PolyPhen2 (Adzhubei et al., 2010), Provean (Choi et al., 2012) and SIFT (Kumar et al., 2009)). Synonymous variants with MAF < 0.1% were considered as potential candidate variants. We annotated it using TraP score (Gelfman et al., 2017) and filtered according to the recommended threshold (Benign if TraP < 0.446) and MAF < 0.1%.

Copy number variation was studied using XHMM (Fromer et al., 2012; Fromer and Purcell, 2014) and filtered afterwards according to the best practices described by the developers. We also used the dataset provided by Audano et al. (2019) to filter out variants described at high frequencies in the population. Finally, possible somatic variants were called using VarScan2 (Koboldt et al., 2012) and filtered considering its variant allele frequency (VAF) and overall quality of the calling (p-value given by VarScan2).

As candidate genes for AID and PID, we used those in the list provided by the International Union of Immunology Societies (IUIS) (Picard et al., 2018). For each patient, we performed an extensive clinical review of the phenotype and family history. Finally, candidate variants were visually inspected in the respective alignment using Integrative Genome Viewer (IGV (Robinson et al., 2011)). We filtered out variants only supported with poor quality reads or low mapping quality.
Quality filtering (8000 variants/patient)
Discard: intronic, off target, QUAL<30, DP<5, <10bp closer to indels, >10 patients

Genetic filtering (630 variants/patient)
Keep: variants accomplishing the requirements in each group

- gnomAD<1% - GO terms - gnomAD pLI score
- gnomAD<1% - GERP>2 - CADD>10 - HI vaies (AD) - GO terms
- gnomAD<0.1% - TraP score
- gnomAD<0.1% - Calling quality
- VarScan2
- gnomAD<1%
- VAF
- P-value

Clinical filtering (10 variants/patient)
Consider: Candidate genes, phenotype, symptoms, clinical history, other findings

3. Results

3.1. AID-associated genes in the TruSight One Sequencing Panel

The TruSight One Sequencing Panel, also referred to as clinical exome gene panel, consists of 4813 genes associated with human disease. With the aim of testing its applicability and performance to study AID, we analyzed the ability to capture and produce reliable sequencing data for candidate genes. We first assessed the proportion of genes related to autoinflammatory disorders included in this panel. For that, we used 35 genes described to cause AID by the IUIS (Picard et al., 2018) and OMIM (http://www.omim.org/). We used essentiality and haploinsufficiency scores to better de-
homozgyous and compound heterozygous genotypes. No candidate homozgyous variant was found in any candidate gene. For the com-pound heterozygotes analysis, we considered together misssense and other functional variants. Patient A10 has two variants located in UNC13D, a gene associated with hemophagocytic lymphohistiocytosis. Sieni et al. (2011) described a patient with these same variants and suggested that they could be considered as susceptibility factors. Due to the lack of parental samples we were not able to test if the variants are inherited from the same progenitor. We also described several variants located at NOD2 in patients A19 and A20, a gene which plays an important role in autoinflammatory disorders (Negroni et al., 2018). A19 and A20 show a clinical picture concordant with alterations in this gene (Table 2). Based on pathogenicity scores and literature search, most of the NOD2 variants described in patients A19 and A20 were considered to be most probably benign (Hugot et al., 2001), except the two misssense variants shown in Table 1.

We also analyzed the presence of synonymous changes in candidate genes. Synonymous variants are traditionally excluded from disease genetics analysis, and are discarded when filtering for functional var-iants. However, synonymous variants have been reported to be related to disease in some cases, especially if they have extremely low allele frequency in databases and provoke the appearance of novel, cryptic splicing sites (Simons et al., 2015). As an example, a novel synonymous pathogenic variant has been reported as causal for IL-7R deficiency (Simons et al., 2015). As an example, a novel synonymous pathogenic variant has been reported as causal for IL-7R deficiency (Simons et al., 2015). As an example, a novel synonymous pathogenic variant has been reported as causal for IL-7R deficiency (Simons et al., 2015). As an example, a novel synonymous pathogenic variant has been reported as causal for IL-7R deficiency (Simons et al., 2015). As an example, a novel synonymous pathogenic variant has been reported as causal for IL-7R deficiency (Simons et al., 2015). As an example, a novel synonymous pathogenic variant has been reported as causal for IL-7R deficiency (Simons et al., 2015).

s.3. Somatic variation

Recent studies corroborated that gene mosaicism is an important mechanism in the pathogenesis of AID (Mensa-Vilaró et al., 2019). Thus, we considered the possibility of the disease being originated by a somatic variant. While germline variants are expected to be present in approximately 50% of the reads, somatic variants will typically depart from these expectations and will show lower variant allele frequencies (VAF). Because of the low number of reads supporting them, these variants are commonly discarded by variant calling software. With the aim of detecting possible somatic variants, we use VarScan2 which has shown good performance on these kind of analyses and allows reduc-tion in the allelic imbalance thresholds (Suzuki et al., 2017).

We generated a table with 12 variants located in Acg or OPcg (Table 56), two of them present in two of the patients. We performed ADS (see Methods) to validate these variants, and to calculate their VAF. We included two previously reported somatic variants as positive controls (CP1 and CP2) (Mensa-Vilaró et al., 2019). We validated the presence of somatic variants in CP1 and CP2 with similar VAF as the previously

| Positiona | Ind | GT | Ref | Alt | Qualb | Gene | Function | Change | Freqc | PolyPhen2 | CADD | dbSNP ID |
|-----------|-----|----|-----|-----|-------|------|----------|--------|------|-----------|------|---------|
| 17:75834443 | A10 | het | C   | T   | 1830.44 | UNC13D | missense | p.R727Q | 0.0001 | P   | 17.36     | rs747390615 |
| 14:5404554 | A11 | het | C   | T   | 645.44 | NFKBIA | missense | p.D31N | 0.0142 | A   | 16.13     | rs9904366  |
| 16:50710654 | A19 | het | T   | G   | 1735.44 | NOD2   | missense | p.L221R | 0.0002 | D   | 26.6      | rs104895423 |
| 16:50712175 | A20 | het | C   | T   | 1844.44 | NOD2   | missense | p.R727Q | 0.0003 | D   | 23.5      | rs61477625  |
| 19:12876108 | A21 | het | C   | T   | 587.44 | DNEASE2 | missense | p.G322D | 0.1543 | A   | 16.13     | rs9904366  |

Table 1: Candidate genetic variants in patients with autoinflammatory disorders.

*a GRCg38/hg38 human reference genome; b GATK scores; c gnomAD population frequencies.
### Table 2
Clinical information of patients harboring variants likely to be causing the autoinflammatory disorder.

| Sample | A10 | A11 | A19 | A20 | A21 |
|--------|-----|-----|-----|-----|-----|
| Gender | Female | Male | Female | Female | Male |
| Country of origin | Morocco | Spain | Spain | Spain | Spain |
| Year of birth | 2012 | 1989 | 1985 | | |
| Age of onset | Infancy | Birth | Childhood | Childhood | Infancy |
| Affected family members | No, healthy sibling | Early onset | No, healthy sibling | No, healthy sibling | No, healthy sibling |
| Initial diagnosis | Systemic-onset JIA | DIRA-like phenotype | Early onset IBD | Early onset IBD | Monogenic lupus |
| Fever | Yes | Yes | Yes | Yes | No |
| Musculoskeletal phenotype | None | No | Severe IBD | Severe IBD | No |
| Cutaneous phenotype | Erythematous skin rash | Pustular skin rash, soft tissue swelling | No | No | Chilblain lupus |
| Ocular phenotype | No | No | No | No | No |
| Treatment | Anti-inflammatory drugs (including Corticoids, hydroxychloroquine) | | | | |
| Complications, other features | Multiorgan failure | Unknown | Alveolar hemorrhage | Unknown | |
| Outcome | Alive | Alive | Alive | Alive | Alive |

* Systemic Lupus Erythematosus; **Juvenile Idiopathic Arthritis; *Deficiency of Interleukin-1 Receptor Antagonist; 4 Inflammatory Bowel Disease.

**Table 2 continued:**

| Sample | A10 | A11 | A19 | A20 | A21 |
|--------|-----|-----|-----|-----|-----|
| Gender | Female | Male | Female | Female | Male |
| Country of origin | Morocco | Spain | Spain | Spain | Spain |
| Year of birth | 2012 | 1989 | 1985 | | |
| Age of onset | Infancy | Birth | Childhood | Childhood | Infancy |
| Affected family members | No, healthy sibling | Early onset | No, healthy sibling | No, healthy sibling | No, healthy sibling |
| Initial diagnosis | Systemic-onset JIA | DIRA-like phenotype | Early onset IBD | Early onset IBD | Monogenic lupus |
| Fever | Yes | Yes | Yes | Yes | No |
| Musculoskeletal phenotype | None | No | Severe IBD | Severe IBD | No |
| Cutaneous phenotype | Erythematous skin rash | Pustular skin rash, soft tissue swelling | No | No | Chilblain lupus |
| Ocular phenotype | No | No | No | No | No |
| Treatment | Anti-inflammatory drugs (including Corticoids, hydroxychloroquine) | | | | |
| Complications, other features | Multiorgan failure | Unknown | Alveolar hemorrhage | Unknown | |
| Outcome | Alive | Alive | Alive | Alive | Alive |

* Systemic Lupus Erythematosus; **Juvenile Idiopathic Arthritis; *Deficiency of Interleukin-1 Receptor Antagonist; 4 Inflammatory Bowel Disease.

**4. Discussion**

We performed targeted sequencing of ~4800 genes of medical interest in 22 samples from patients with suspected AID, mostly with early-onset. We propose a candidate genetic variant in two of the patients, based on stringent filtering of DNA variants and the patients’ clinical phenotypes. For three additional patients, we also propose candidate variants which can be acting as susceptibility factors of the disease.

Patient A11 was a Spanish male that suffered since birth from a pustular and systemic inflammatory disease resembling the deficiency of interleukin-1 receptor antagonist (DIRA), he unfortunately died soon after birth because of multiorgan failure. Our analysis revealed the patient carried a novel missense heterozygous variant at NFKBIA (p.D31N). NFKBIA is an inhibitor of the NFKB family which acts as a regulatory element of the inflammatory response (Hayden and Ghosh, 2008). When proinflammatory stimuli are detected NFKBIA is phosphorylated and degraded, leading to the nuclear translocation of NFKB, which correlates with our patient’s phenotype.

The patient presented with sterile periostitis, pustular skin rash, and soft tissue swelling. We suspect that the variant p.D31N could be contributing to the disease in patient A11. This variant fits to some of the standard criteria (Richards et al., 2015) to be classified as pathogenic: absence in databases; presence in a critical functional domain with reported pathogenic variants; in a gene where missense variation is a common mechanism of disease; a significant lack of missense and LoF variants; deleterious effect according to in silico predictors; and specific gene phenotype. We reported the variant to ClinVar (SCV001147010), unfortunately, we were not able to perform segregation studies to validate the possible de novo origin due to the non availability of additional patient and parental samples. Of interest, this patient also harbours a synonymous variant in homozygosity with a TraP score close to the pathogenicity threshold, a high CADD value of 17.04 and located in a promoter region (Ensembl ID: ENSR00000067656).

Patient A21 is a Spanish male with early-onset lupus-like symptomatology. A novel missense heterozygous variant at DNA2E2 was found in the patient (p.G322D), this variant may be pathogenic (Richards et al., 2015) to be classified as pathogenic: absence in databases; presence in a critical functional domain with reported pathogenic variants; in a gene where missense variation is a common mechanism of disease; a significant lack of missense and LoF variants; deleterious effect according to in silico predictors; and specific gene phenotype. We reported the variant to ClinVar (SCV001147010), unfortunately, we were not able to perform segregation studies to validate the possible de novo origin due to the non availability of additional patient and parental samples. Of interest, this patient also harbours a synonymous variant in homozygosity with a TraP score close to the pathogenicity threshold, a high CADD value of 17.04 and located in a promoter region (Ensembl ID: ENSR00000067656).
et al., 2015) due to its absence in databases, its in silico predictors, conservation values (Table 1), and reported pathogenic variants in the gene (Rodero et al., 2017). DNA2 has recently been associated to inflammatory diseases (Rodero et al., 2017), and mutations in DNA2 are well known to be causing monogenic forms of systemic lupus erythematosus (Martínez Valle et al., 2008; Yasutomo et al., 2001), with a clinical phenotype partially overlapping with that described in this patient. In our patient, the disease seems to follow an autosomal dominant pattern of inheritance, as some close relatives are also affected (father, uncle and grandmother). This diverges from the described autosomal recessive cases. We performed Sanger sequencing in the patient (Fig. S2) and we propose the variant p.G322D as a good candidate to underly the disease in patient A21. We reported the variant to ClinVar (SCV001147011).

In addition to the analysis to detect disease-causing variants, we also considered the possibility of mutations contributing to the phenotype although not responsible for the whole clinical picture. Patient A10 is a 12-year-old girl from North Africa with systemic-onset juvenile idiopathic arthritis (JIA). She suffered episodes of macrophage activation syndrome (MAS), which are associated with UNC13D mutations. We found two variants at UNC13D which have been already reported by Sieni et al. (2011) in the analysis of haemophagocytic lymphohistiocytosis. Both variants are heterozygous and missense (p.R727Q and p.A59T), with low population allele frequencies (0.05% and 1.6% respectively) and high predicted impact values (CADD > 20). Although the patient's main clinical diagnosis is not MAS, her phenotype resembles the ones described in Sieni's work. This led us to suspect that maybe part of the phenotype is caused by these missense variants. Of note, the presence of two or more causal genes for different Mendelian disorders has been reported in 4.9% of the 2076 molecularly diagnosed patients in an exhaustive analysis of more than 7000 cases (Posey et al., 2016). The case reported here would fit the situation where the two disease phenotypes in the same individual are partially overlapping.

We also described many variants in NOD2 gene in our cohort. NOD2 has a key role in immune function and inflammatory pathways. Numerous variants have been described in the gene in recent years, mainly associated with Blau syndrome and Crohn disease (Negroni et al., 2018). Of interest, two of the patients with NOD2 variants (A19 and A20) have early-onset inflammatory bowel disease, a phenotype which has already been associated to NOD2 gene (Cho and Abraham, 2007; Negroni et al., 2018). Each of them carries one missense variant of uncertain significance, p.L221R in patient A19 and p.A728V on A20. Both variants have been reported by Hugot et al. (2001) in inflammatory bowel disease patients accompanied by other coding variants. We hypothesize a possible additive effect which would contribute to the final phenotype.

Of interest, in this work, we show that beyond the detection of germline genetic variants, targeted sequencing also allows the detection of somatic variation if relatively high sequencing depths are achieved. Thus, while no important differences are seen in the success rate of whole exome sequencing versus candidate genes approaches (see below), reducing the number of targeted genes allows to increase the depth of coverage at still reasonable costs, which is a critical aspect for somatic variant detection. It can be of special interest in the case of PID in general and AID in particular, where somatic variants will probably be present in the blood or even saliva, the two most common starting materials for DNA extractions. We recommend using software allowing the reduction of the thresholds on the number of reads supporting the alternative allele (i.e. VarScan2) followed by ADS to validate candidate variants. ADS experiments increase the coverage of the targeted region from the range of hundreds to thousands, allowing us to discard false-positives.

The proportion of patients with a proposed molecular diagnosis in this work is similar to other previous analyses in PID, both from whole genome and exome sequencing (Arts et al., 2019; de Valles-Ibáñez et al., 2018; Stray-Pedersen et al., 2017) and targeted sequencing approaches (see Introduction). However, the number of diagnosed patients is barely comparable across studies because of the use of different filters and stringency criteria to consider a genetic variant as causal. This would ultimately require functional validation experiments, which are mostly performed in a reduced subset of the candidate variants. Also, the proportion of monogenic cases in a cohort depends on other features, such as disease severity, the age of onset and the disease prevalence (de Valles-Ibáñez et al., 2018). Thus, the different inclusion criteria of patients, can also partially explain the differences in the success rates in detecting causal genes across studies. In addition to the clinical exome approach being limited to a particular gene set, other technical constraints refer to the non-inclusion of regulatory regions and limited power for structural variant detection. We note some of these factors are also common to whole exome or even whole genome strategies. Beyond technical and analytical reasons, we should always consider the possibility that the disease follows non-monogenic models, such as a digenic pattern of inheritance (Ameratunga et al., 2018; de Valles-Ibáñez et al., 2018; Hoyos-Bachiloglu et al., 2017) or gene mosaicism (Mensa-Vilaró et al., 2019).

Overall, we propose that targeted sequencing approaches of candidate genes, such as clinical exome, can be an appropriate tool for the genetic diagnoses of AID. The main limitation of targeted approaches is that the analysis is restricted to particular regions. However, this limitation to specific regions is also seen in several whole genome and exome data studies, where in spite of generating more complete data the analysis is often finally restricted to a set of candidate genes, at the cost of reducing sequencing depth (Meyts et al., 2016). In this sense, the design of specific gene panels for patients with similar disorders can represent an efficient strategy for diagnosis, which may even be able to capture the presence of somatic mosaicism which is relevant in diseases such as AID. Of importance, sequencing coverage and depth analyses to detect genes that have not been properly covered should be performed to avoid false-negative results. Also, the creation of an adequate candidate genes list is critical for this approach and should be based on a strong knowledge of the disease pathophysiology and the use of available tools to describe functionally related genes. However, generating a comprehensive but relatively short list of candidate genes list can be almost impossible for situations like this study, with an important level of clinical heterogeneity among patients even if their syndromes are included in the same group of disorders. In such situations, a more inclusive approach as the analysis of the ~4800 genes more probably related to genetic disease, can be an efficient strategy for maximizing the number of diagnosed patients while reducing the costs of sequencing and analysis in whole genome or exome approaches.

CRediT authorship contribution statement

Laura Batlle-Masó: Conceptualization, Methodology, Software, Investigation, Writing - original draft. Anna Mensa-Vilaró: Conceptualization, Methodology, Investigation, Writing - review & editing. Manuel Solís-Moruno: Software, Formal analysis, Writing - review & editing. Tomàs Marquès-Bonet: Resources, Funding acquisition. Juan I. Arostegui: Conceptualization, Resources, Writing - review & editing, Supervision. Ferran Casals: Conceptualization, Resources, Writing - original draft, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
Acknowledgements

We thank all participants in the study. This study was funded by grants SAF2012-35025 and SAF2015-68472-C2-2-R from the Ministerio de Economía y Competitividad (Spain) and FEDER (EU) to FC, RTI2018-096824-B-C22 grant from the Spanish Ministry of Science, Innovation and Universities co-financed by ERDF (FC), by Dirección General de Recerca, Generalitat de Catalunya (2014SGR-866 and 2017SGR-702) to FC, the CERCA Programme, Generalitat de Catalunya (JIA), SAF2015-68472-C2-1-R grant from the Spanish Ministry of Economy and Competitiveness co-financed by European Regional Development Fund (ERDF) (JIA), RTI2018-096824-B-C21 grant from the Spanish Ministry of Science, Innovation and Universities co-financed by ERDF (JIA), and AC15/00027 grant from the Instituto de Salud Carlos III, Translational Research Projects on Rare Diseases (JIA). LBM is supported by a Formación de personal Investigador fellowship from Generalitat de Catalunya (2018_F1_B00072). MSM is supported by the María de Maetzu Programme (MDM-2014-0370-16-3).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejmg.2020.103920.

References

Adzhubei, I. a, Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., 2010. A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–249. https://doi.org/10.1038/nmeth0410-248.

Aldrich, M., Beerman, T., Gehrke, H., 2019. An update on antiphospholipid syndrome: clinical features, diagnostic criteria, and therapeutic approaches. J. Immune. Med. 58, 401–416. https://doi.org/10.1146/annurev-med.58.061705.145024.

Aldridge, J., Chong, J.X., Buckingham, K.J., Jiangangi, S.N., Boehm, C., Sobreiza, N., Smith, J.D., Harrell, T.R., McMullin, M.J., Wixsniewski, W., Gambin, T., Coban Akdemir, Z.H., Dohney, K., Scott, A.F., Avramopoulos, D., Chakravarti, A., Hoever-Fong, J., Mathews, D., Witmer, P.D., Ling, H., Heink, K., Watkins, L., Patterson, K.E., Reiner, F., Flint, E., Muney, D., Leal, G., van de Pol, B., Fry, R., Bucan, M., Caskey, C., Yilmaz, E., 2012. Genetic basis of mendelian phenotypes: discoveries, challenges, and opportunities. Am. J. Hum. Genet. 97, 199. https://doi.org/10.1016/j.ajhg.2015.06.009.

Carenzzi, F., De Martini, M., Ginaldi, L. 2014. An update on autoinflammatory diseases. Curr. Med. Chem. 21, 261–269. https://doi.org/10.2174/0929867131006660030.

Cooper, G.M., Stone, E.A., Asimenos, G., Program, S., Green, E.D., Batzoglou, S., Sidow, A., 2005. Distribution and intensity of Constraint in Mammalian Genomic Sequence. Curr. Med. Chem. 21, 261–269. https://doi.org/10.2174/09298673113206660303.

DiDonato, J., Mercurio, F., Rosette, C., Wu-Li, J., Suyang, H., Ghosh, S., Karin, M., 1996. The p65 component of the NF-kappaB transcription factor physically associates with the kinase I kappaB alpha. Cell 87, 295–305. https://doi.org/10.1016/0092-8674(96)80023-6.

Endo, Y., Takahasi, T., Diogo, D., del Valle-Álvarez, G., Esteve-Solé, A., Piquer, M., González-Navarro, E.A., Hernandez-Ortega, J.L., 2019. Exome sequencing in patients with autoinflammatory diseases. Annu. Rev. Med. 70, 311–332. https://doi.org/10.1146/annurev-med-051018-014714.

Fael, A., van de Veerdonk, F.L., Netea, M.G., Hoischen, A., 2019. Exome sequencing identifies a shared genetic background in autoinflammatory diseases and bronchiolitis. J. Immunol. 202, 1108–1115. https://doi.org/10.1172/JCI18714.

Gallego-Bustos, F., Gotes, V., Ramos-Amador, J.T., Rodríguez-Pena, R., Gil-Herrera, J., Sastre, A., Delmiro, A., Potiez, J.J., Alvear, L.M., Albumin, M., Allende, L.M., 2016. A case of IL-7R deficiency caused by a novel synonymous mutation and implications for mutation screening in SCID diagnosis. J. Immunol. 197, 4474–4483. https://doi.org/10.4049/jimmunol.1600919.

Gelfand, M., Wang, Q., McInerney, K.M., Ren, Z., La Carpa, F., Halvorsen, M., Schoch, K., Ratzen, F., Heinen, E.L., Boland, M.J., Petrovski, S., Goldstein, D.B., 2017. Annotating pathogenic non-coding variants in genomic regions. Nat. Genet. 49, 23–30. https://doi.org/10.1038/nature.2016.20033.

Hoffman, D.W., Ghosh, S., 2009. General principles in NF-kappaB signaling. Cell 132, 346–362. https://doi.org/10.1016/j.cell.2008.01.020.

Hoffman, D.W., Broderick, L. 2016. Clinical reviews in allergy and immunology the role of the inflammasome in patients with autoinflammatory diseases. https://doi.org/10.1097/01.cri.0000563668.70305.d8.

Housi-Bachaloglu, R., Chou, J., Sodroski, C.N., Beano, A., Bainter, W., Angelova, M., Al Idrissi, E., Habazi, M.K., Alghamdi, H.A., Almanjomi, F., Al Shehri, M., Ehsid, N., Eldin, M.A., Kline, D.M., Alzahrani, M., Geha, R.S., 2017. A digenic human immunodeficiency characterized by IFNAR1 and ING5R mutations. J. Clin. Invest. 127, 4415–4420. https://doi.org/10.1172/JCI93948.

Huang, N., Lee, I., Marcotte, E.M., Hurles, M.E., 2010. Characterising and predicting haploinsufficiency in the human genome. PLoS Genet. 6, e1–11. https://doi.org/10.1371/journal.pgen.1000969.

Hugot, J.P., Chamaillard, M., Zouali, H., Lesage, S., Cézard, J.P., Belaiche, J., Almers, J., Syrko, C., Morfin, A., Gasull, M., Binder, V., Finkel, Y., Cortot, A., Modigliani, R., Laurent-Puig, P., Gower-Rousseau, C., Marcy, J., Colombel, J.F., Sabato, M., Thomas, C., 2001. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 411, 599–603. https://doi.org/10.1038/35079107.

Kasarskis, A., Lewis, S., Matse, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene Ontology: tool for the unification of biology. Nat. Genet. 25, 25–29. https://doi.org/10.1038/350556.

Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging non-synonymous variants. Nucleic Acids Res. 38, 654–658. https://doi.org/10.1093/nar/gkq1142. 23.21.

Kožuch, J., Trzcińska, A., Drsgo, P., Kaczmarek, M., Zamojski, M., Zemlicki, M., 2012. The frequency of FLCN gene alterations in the group of patients with sporadic von Hippel-Lindau disease. Med. Sci. Monit. 18, 1670–1674. https://doi.org/10.12659/MSM.804306.
variants from high-throughput sequencing data. Nucleic Acids Res. 38, e164. https://doi.org/10.1093/nar/gkq603. e164.

Wekell, P., Berg, S., Karlsson, A., Fasth, A., 2017. Toward an inclusive, congruent, and precise definition of autoinflammatory diseases. Front. Immunol. 8, 6–10. https://doi.org/10.3389/fimmu.2017.00497.

Wekell, P., Karlsson, A., Berg, S., Fasth, A., Wekell, C.P., 2016. Review of autoinflammatory diseases, with a special focus on periodic fever, aphthous stomatitis, pharyngitis and cervical adenitis syndrome. https://doi.org/10.1111/apa.13531.

Yao, Q., Lacbawan, F., Li, J., 2016. Adult autoinflammatory disease frequency and our diagnostic experience in an adult autoinflammatory clinic. Semin. Arthritis Rheum. 45, 633–637. https://doi.org/10.1016/j.semarthrit.2015.10.012.

Yasutomo, K., Horiuchi, T., Kagami, S., Tsukamoto, H., Hashimura, C., Urushihara, M., Kuroda, Y., 2001. Mutation of DNASE1 in people with systemic lupus erythematosus. Nat. Genet. 28, 313–314. https://doi.org/10.1038/91070.

Yska, H.A.F., Elsink, K., Kuipers, T.W., Frederix, G.W.J., van Gijn, M.E., van Montfrans, J.M., 2019. Diagnostic yield of next generation sequencing in genetically undiagnosed patients with primary immunodeficiencies: a systematic review. J. Clin. Immunol. 39, 577–591. https://doi.org/10.1007/s10875-019-00656-x.