Next Generation Sequencing Data Analysis in Primary Immunodeficiency Disorders – Future Directions

Mingyan Fang[1,2] · Hassan Abolhassani[1,3] · Che Kang Lim[1,4] · Jianguo Zhang[2] ·
Lennart Hammarström[1]

Abstract Primary immunodeficiency diseases (PIDs) comprise a group of highly heterogeneous immune system diseases and around 300 forms of PID have been described to date. Next Generation Sequencing (NGS) has recently become an increasingly used approach for gene identification and molecular diagnosis of human diseases. Herein we summarize the practical considerations for the interpretation of NGS data and the techniques for searching disease-related PID genes, and suggest future directions for research in this field.

✉ Lennart Hammarström
   lennart.hammarstrom@ki.se

   Mingyan Fang
   fangmingyan@genomics.cn

   Hassan Abolhassani
   hassan.abolhassani@ki.se

   Che Kang Lim
   che.kang.lim@ki.se

   Jianguo Zhang
   zhangjg@genomics.cn

1   Department of Laboratory Medicine, Division of Clinical Immunology and Transfusion Medicine, Karolinska University Hospital Huddinge, -141 86 Stockholm, SE, Sweden

2   BGI-Shenzhen, Shenzhen 518083, China

3   Research Center for Immunodeficiencies, Pediatrics Center of Excellence, Children's Medical Center, Tehran University of Medical Sciences, Tehran, Iran

4   Department of Clinical Research, Singapore General Hospital, Singapore 169856, Singapore

Abbreviations
NGS   Next generation sequencing
PID   Primary immunodeficiency diseases
MDs   Mendelian disorders
WGS   Whole genome sequencing
WES   Whole exome sequencing
QC   Quality control

Primary immunodeficiency diseases (PIDs) comprise a group of highly heterogeneous genetic disorders caused by defects in the immune system and can be categorized as lymphocyte deficiencies (T and/or B cell and NK cells defects), phagocytic defects, complement deficiencies and innate immunodeficiencies [1]. The prevalence of each individual PID varies in the population [2], ranging from 1/600 for selective IgA deficiency [3] to approximately 1/250,000 for chronic granulomatous disease [4]. However, the prevalence of various forms of PID is probably underestimated [5]. There are currently close to 300 forms of PID as defined by the International Union of Immunological Societies (IUIS) [6, 7]. However, around 3110 genes may be potentially PID-causing based on their biological functions and human gene connectome analysis [8] suggesting the existence of a large number of hitherto unrecognized diseases. Some of these diseases are monogenic, but a growing proportion are caused by a digenic/polygenic mechanism [9, 10].

Since the first methodology paper on next generation sequencing (NGS) was published [11], this technology has rapidly advanced the discovery of genetic variants underpinning human Mendelian disorders (MDs). Rapid advances in NGS facilitate, at ever decreasing costs, processing and analysis of genomic regions, ranging from targeted gene sets to high coverage of whole human genomes, including linkage analysis, homozygosity mapping and candidate gene approaches.

Whole exome sequencing (WES) and whole genome sequencing (WGS) have become increasingly widespread approaches for the identification of MDs associated genes, by sequencing trio or quartet [12] families, by sequencing several individuals in a pedigree [13], a small patient cohort [14], or even by only sequencing the proband [15].

Gene discovery and understanding of the molecular basis of disease are essential starting points for making a molecular diagnosis and providing genetic counseling [16, 17] and may provide clues to the development of new therapeutic approaches for different diseases. In the past several years, the molecular basis of several forms of PID and their clinical consequences has been well documented [18–25]. However, despite intensive work during the last decades, many forms of PID still do not have a defined underlying genetic defect and further exploration is needed to push this frontier forward. There are several publicly accessible tools or bioinformatic analysis pipelines available, which house numerous NGS analysis scripts for data quality control (QC), as well as tools for alignment, variant-calling, and annotation. However, a comprehensive pipeline designed specifically for PID candidate gene screening has not yet been developed. Herein we summarize practical considerations for the interpretation of the NGS data from PID patients and provide an overview of data and findings from our own PID cohort.

Pipeline for Screening the Candidate Gene

Next Generation Sequencing is a promising strategy for the study of human monogenic disorders. However, pinpointing the causal mutations in a small number of samples is still a major challenge. A large number of variants can be detected by WES and WGS and approximately 100,000 SNVs and 15,000 Indels (Fig. 1) have been identified per sample using the Agilent SureSelect Human All Exon V4 kit in our PID cohort. They include polymorphisms, sequencing artifacts and non-pathogenic rare mutations, and finding the causative gene remains a challenge (Fig. 1).

There are several steps that may help us to reduce the number of candidate genes. Firstly, we can prioritize variants according to their impact on the sequence of the protein product. The majority of disease-causing variants are believed to be amino-acid-sequence-changing, including non-synonymous substitutions, splice site mutations and insertion/deletions as well as truncation of proteins due to a premature stop codon. To avoid the possibility of failure to find disease-causing mutations due to the incompleteness of the available annotation databases, using multiple databases [26] to annotate the variants to select the canonical one and the one with the most severe consequence is recommended. However, this method of annotation will not identify large deletions, even in a homozygous form; therefore, additional bioinformatics tools

such as Exondel [27] (for homozygous large deletion) and FishingCNV [28] (for homozygous as well as heterozygous large deletion) using cross-samples comparisons are required in separate analysis step.

Secondly, as most of the PIDs are relatively uncommon, and supposedly caused by rare mutations, mutations found at polymorphic sites can be excluded by removal of mutations found in public or in-house databases and have a frequency of more than 1 %. The available public databases include the 1000 Genomes Project (KG) (2500 samples; http://www.1000genomes.org/), the Exome Variant Server (ESP) (6500 WES samples; http://evs.gs. washington.edu/EVS/) and the Exome Aggregation Consortium (ExAC) (60,706 samples; http://exac.broadinstitute.org/). It should be noticed that the ESP and ExAC database are mix of healthy individuals and patients, which should be carefully used when there is any patient affected by the studied disease in the database. In-house databases are often helpful to exclude platform-specific false positives or background noise. Analysis of the frequency of PID-related variants downloaded from public databases, including 5206 PID causative variants (selected substitutions or Indels from 7561 records), the Resource of Asian Primary Immunodeficiency Diseases (RAPID) http://web16. kazusa.or.jp/rapid/ database [29] and 4203 variants in PID genes [6] from The Human Gene Mutation Database (HGMD) http://www.hgmd.cf.ac.uk/ac/index.php database, show that most causative PID variants have an allele frequency of less than 1 % (Fig. 2 a,b). This also provides the rationale for exclusion of polymorphisms in the candidate mutation prioritization.

Examining the consistency of the inheritance model of disease and zygosity of the mutation is another important step to identify causative variants. It is rational to examine known genes with biallelic variants for a recessive model, hemizygous variants for an X-linked model in male patients, and heterozygous variants for a dominant model. Classification update of PID from IUIS in 2015 shows that, approximately 70 % are inherited as autosomal recessive, 20 % as autosomal dominant and 5 % as X-linked disorders [6] and the percentage of new autosomal dominant etiologies is increasing more than other types, however, it should be noted that sporadic PID cases are not uncommon [30]. These cases could be due to *de novo* dominant heterozygous mutations. We hypothesize that the penetrance of the disease is high for a family only has single or limited affected members. Hence, it is rational to hypothesize that unaffected members carry no disease predisposing/causing variant that is consistent with an inheritance model. Some complicated algorithms based on genetic rules could help us to narrow down the candidate gene list. For instance, WES data enable haplotype analysis based on identical by descent (IBD) information, which could be employed to pinpoint chromosomal segments that may harbor disease-causing mutations and performing homozygosity mapping (using HomSI, AutoSNPa and AgileVariantMapper et al. [21, 31]) will facilitate disease gene identification in consanguineous families as they are likely to be located in homozygous regions. In

Fig. 1 Pipeline of NGS-based gene identification in PID. The bar chat indicates the remaining number of variants after each prioritization step
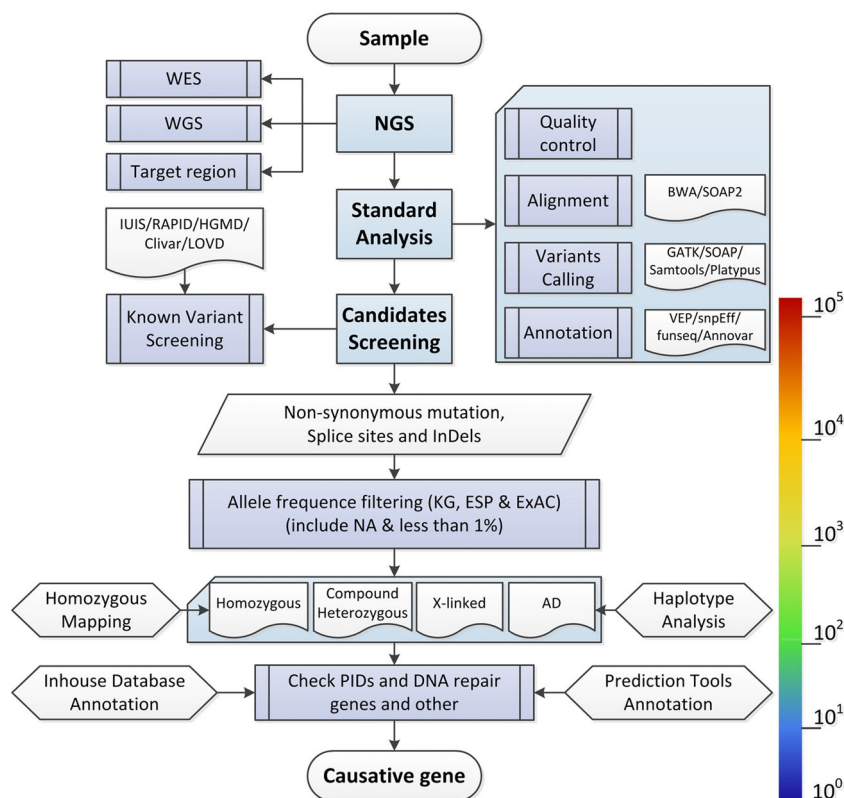
addition, multiple software tools such as pVAAST [32] and VarScan trio family calling [33] can be used to identify genetic variant(s) that directly influence disease risk in both consanguineous and non-consanguineous family data. VarScan trio support single family analysis while pVAAST provide probabilistic predictions of the disease- risk variants from single to multiple families' data.

Most of the MDs causing mutations are supposed to be functionally critical and located in sites that are evolutionarily conserved. Several software tools have been developed based on this theory and may provide helpful information on the severity of impact of the variants. These tools include CADD, SIFT, PolyPhen2, GERP, GWAWA and MutationTaster, all of which predict the biological effect of the variants. These tools are powerful, but they do not categorize all variants correctly and sometimes they are inconsistent with each other. Both SIFT and PolyPhen2 have a difficulty in predicting stop-gain variants. Prediction of the severity and functional impact of known PID-causing variants show that a majority are predicted to be damaging, but some are predicted to be benign while others cannot be evaluated (Fig. 2c, d). Known PID genes can be utilized for training for some prediction models, which prioritize candidate genes according to the connection between the function, pathway, expression information on candidate genes and phenotype, including VASST, eXtasy, PHEVOR, Phen-Gen, Phenolyzer, Phenomizer, GeneCards and PosMed [8, 34–36].

Some genes are prone to enrich phenotype-irrelevant mutations due to their long coding sequences [37] and evolutionary pressure [38]. It should be noted that these genes are the source of many false positive results. Fortunately, there are some tools or databases available to reduce noise, including Gene Damage Index (GDI) [38], a tool to indicate the degree of mutation enrichment in diseases under different inheritance models. Another available tool is Frequently Mutated Genes (FLAGS) [37], which provides a list of genes that are enriched for neutral mutations. For a given sequencing platform, a database of genes that are enriched in neutral mutations could provide helpful information in order to reduce platform-specific false variants.

On the other hand, internal homozygous exon deletion may be investigated using Exondel [27], a tool that use exons as a unit to examine deletions at the gene level. By comparison with multiple controls and in-house databases, the pipeline is able to identify disease-associated large exon deletions within a gene region.

In our cohort study, sequencing reads were aligned to the human reference genome (NCBI build 37.1, hg19) using both SOAPaligner and BWA with default parameter. Approximately 1,000,000 SNV and 100,000 Indels for the cohort have been detected by using SOAPsnp and GATK. After several steps of filtering such as removal of polymorphic site with frequency more than 1% in public database, keep the amino acid changing variants, there were still some mutations that were found in a majority of the sequenced samples and were predicted to be highly deleterious. These included 61
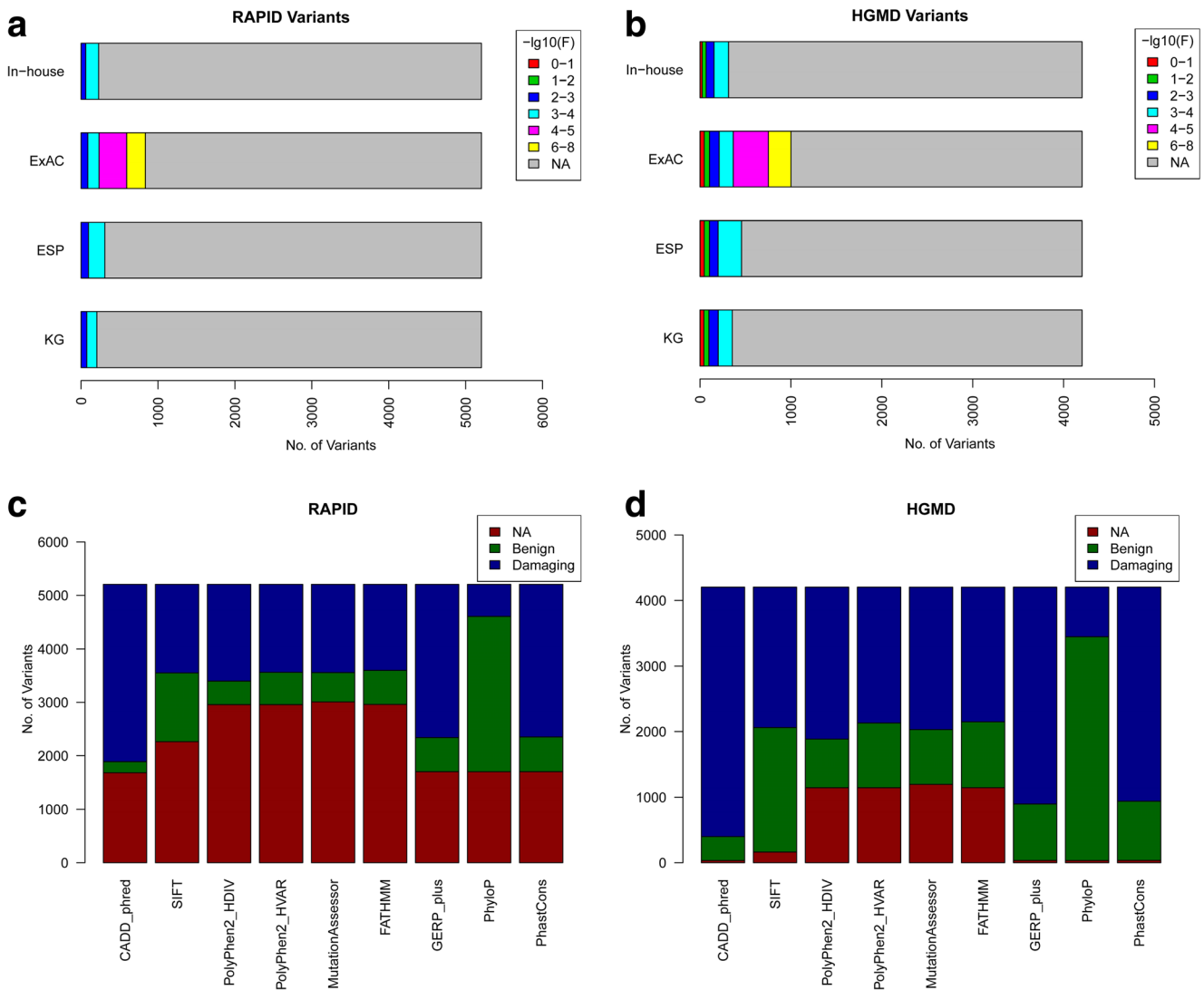
Fig. 2 Frequency and predicted functional impact of the known PID variants. (**a**) (**b**) Number of PID variants with different ranges of frequencies where variants were downloaded from the RAPID (http://web16.kazusa.or.jp/rapid/) database and the HGMD (http://www.hgmd.cf.ac.uk/ac/index.php) database, respectively. The frequency data are based on our in-house database, ExAC (http://exac.broadinstitute.org/), ESP (http://evs.gs.washington.edu/EVS/) and KG (http://www.1000genomes.org/) databases, 'NA' represents variants not present in the corresponding database. (**c**) (**d**) Number of variants fall into different categories of functional impact based on prediction using different software, 'NA' represents the severity of functional impact of the variants that could not be predicted by the software

SNVs and 217 InDels and were carried by more than 50 % of the samples (206 patients) and 50 % of our PID controls (27 normal individuals from the patients' families) in our cohort (data not shown), suggestive of platform association errors.

Here we summarize some of the variants that were found among our patients, including variants reported in our previous studies (Table 1), hypomorphic variants causing an incomplete phenotype (F3, F9, F12, F15, F16), a new or unexpected clinical presentation in known PID genes (F10, F19), variants in known genes but technically failed in Sanger screening (F11, F13, F14, F16), a large deletion mutation spanning an exonic region (F17) and mutations in potential modifier genes result in different phenotypes in two siblings with a same causative mutation (F7, F12, Table 1).

The strategies described above to prioritize detected gene variants are commonly used. Although for many rare diseases, filtering variants with a frequency of greater than 1 % is reasonable, it may not be suitable for every PID, especially for recessive diseases with a relatively high incidence [5]. It should also be noted that their frequency may vary in different ethnic groups and some databases, for instance, dbSNP, should be used with caution, since it is built from a myriad of sources which do not only include polymorphisms but also disease-causing mutations. In-house databases could provide a suitable filter to remove platform or laboratory-specific artifacts. However, the sample size of the database and the population structure of samples has an impact on the sensitivity and specificity of the filter. For instance, if the in-house database only contains samples from a

Table 1 Reported mutations in known PID genes in our cohort

Gene	Nucleotide change	Amino acid change	Function	Sample information	Exon No.	Allele frequency in			Patients[#]	Normal individuals[#]	Computational prediction						Ref
						KG	ESP	ExAC			CADD	SIFT	PolyPhen2 HDIV	Gerp	GWAWA	MutationTaster	
CD27	c.232C > T	p.Arg78Trp	missense	F1, Hom	'2/6'	NA	NA	1.65E-05	1	0	34	0.01	1	3.64	-	0.828247	[18]
CD27	c.287G > A	p.Cys96Tyr	missense	F2, Hom	'3/6'	NA	NA	NA	2	0	28.2	0.04	1	3.87	-	0.951396	[18]
JAK3	c.3308G > A	p.Arg1103Gln	missense	F3, Hom	'23/23'	NA	NA	1.65E-05	1	0	13.95	1	0.485	−0.537	-	-	[19]
LRBA	c.4814C > G	p.Ser1605Ter	stop_gained	F4, Hom	'30/57'	NA	NA	NA	1	0	49	-	-	4.67	-	-	[20]
LRBA	c.1014 + 1G > A	-	splice_region_variant	F5, Hom	'8–9/57'	NA	NA	NA	1	0	26.5	-	-	5.41	-	-	[20]
LRBA	c.175G > T	p.Glu59Ter	stop_gained	F6, Hom	'2/57'	NA	NA	0	2	0	37	0.65	-	5.34	0.61	1	[20]
LRBA	c.4729 + 2dupT	-	splice_region_variant	F7, Hom	'29–30/57'	NA	NA	NA	2	0	-	-	-	-	-	-	[20]
LRBA	c. 544C > T	p.Arg182Ter	stop_gained	F8, Hom	'4/57'	NA	NA	0	1	0	38	0	1	-	-	-	[20]
RAG1	c.1073G > A	p.Cys358Tyr	missense	F9, Hom	'8/10'	NA	NA	0	1	0	25.4	0	1	5.39	-	0.999998	[21]
RAC2	c. 167G > A	p.Trp56Ter	stop_gained	F10, Hom	'3/7'	NA	NA	0	2	0	39	0	1	-	-	-	[22]
BTK*	c.982C > T	p.Gln328Ter	stop_gained	F11, Hem	'12/16'	NA	NA	NA	1	0	37	1	-	5.18	0.31	1	[39]**
BTK*	c.82C > T	p.Arg28Cys	missense	F12, Hem	'2/19'	NA	NA	NA	2	0	32	0	1	5.35	0.41	0.999941	[40]**
BTK*	c.1263G > A	p.Trp421Ter	stop_gained	F13, Hem	'14/19'	NA	NA	NA	1	0	39	1	-	5.44	0.41	1	[41]**
BTK*	c.1237T > C	p.Phe413Leu	missense	F14, Hem	'14/19'	NA	NA	NA	1	0	23.5	0.04	0.997	5.44	-	0.999989	[42]**
BTK*	c.1750 + 5G > A	-	splice_region_variant	F15, Hem	'17–18/19'	NA	NA	NA	1	0	-	-	-	-	-	-	[43]**
CD40LG*	c.496C > T	p.Gln166Ter	stop_gained	F16, Hem	'4/4'	NA	NA	NA	3	0	-	0.65	-	4.65	0.48	1	[44]**
LRBA*	Exon 41 del	40 amino acid del	large deletion	F17, Hom	'41/57'	NA	NA	NA	2	0	-	-	-	-	-	-	-
DCLRE1C*	c.41G > T	p.Gly14Val	missense	F18, Hom	'5/16'	NA	NA	NA	1	0	33	0	1	5.65	0.42	0.999979	-
ATM*	c.5552_5553insC	p.Gln1852ProfsTer5	frameshift variant	F19, Hom	'36/63'	NA	NA	NA	1	0	-	-	-	-	-	-	-

Hom: Homozygous according to an autosomal recessive model. Hem: Hemizygous according to an X-linked model. NA: not applied

* Patients from our cohorts that have not been previously reported

** First report on a primary immunodeficiency patient with the same mutation in the literature from other cohorts

In-house database from 183 families sequenced at BGI for 206 patients and 27 healthy relatives

few big families, it could be enriched for rare, disease-associated heterozygous variants, which may lead to the exclusion of recessive disease-causing homozygous variants.

Disease-causing genes may not be found in all studied families, for instance, in our study, disease-causing candidate genes were only identified in 134 of 183 families sequenced, (134 patients in a total of 206 cases). Failure in identification of causative genes in the remaining families could be due to many reasons, one being that some genes or exons are not well covered or included in the capture arrays, including some known PID genes (*LAMTOR2*, *EPG5*, *IGKC*, *IGHG1*, *IGHG2*, *IGHG3*, *IGHG4*, *IGHA1*, *IGHA2*, *IGHE and IL36RN*). Large insertions or deletions, copy number variations or structural variations are also not easily traceable using short read approaches. In addition, some under-investigated disease-associated variants like SNPs in the UTR region, long non-coding RNAs and non-canonical or deep intronic splicing elements may not be detected by WES or targeted region sequencing. The advance of sequencing technologies and the improvement of capture may narrow this gap, and carefully making choices according to the characteristics and preferences of the different platforms may also be helpful in order to improve the success rate of genetic studies [45]. Beside all these technical limitation, new genetic etiologies that may be unrelated to the previous known genes/pathways, that is beyond our hypothesis, and is not easy to be identified, especially when the subjects are sporadic or a very small number of patients are available in family.

If no good candidate is found, revisiting all assumptions in the screening pipeline and analysis of the parameters is essential for finding the missing pathogenic variants. To reduce false positive variant calling in the homogenous regions of the genome, we only use unique mapping reads to call variants, and as a result, call variants in homogenous regions may be missed. If no good candidates are found, we can review variants in the homogeneous regions by analyzing all reads, including reads mapping to multiple regions, for variant calling.

Although the variant prioritization methods described above are able to exclude most of the irrelevant mutations, accurately identifying disease-causing mutations in the remaining patients is still a challenge. Some novel methodologies to assist the variant prioritization process could be very helpful. One possible way is to prioritize genes based on their position in a pathway, the hypothesis being that genes in the same or a related pathway are more likely to be relevant to the same disorder. Mapping known MDs genes into pathways and assessing their relative position is essential to test this hypothesis. If it holds true, we will be able to relax the filtering criteria, to reduce false negative results in disease gene identification.

During the past decades, much effort has been devoted to unravelling the genetic basis of different forms of PID, and the etiology of many monogenic diseases have now been identified [6]. However, even in some disorders with a presumed monogenic origin, and in most patients with a disease of digenic or polygenic origin, little progress has been made to date. The genetic analysis is also hampered by a variable clinical picture, where patients with mutations in recognized PID genes that are known to result in a given clinical phenotype, may, in fact, have features characteristic of a broad spectrum of different PIDs [18–24]. In addition, family members carrying the same "pathogenic" mutation as the proband may be healthy, suggesting an influence of modifying genes in the former, beside, factors other than genetic one can make a contribution to the development of disease, which include environmental factor, age, diet et al.

Mutations in several genes that are potentially disease-causing have been observed in patients with common variable immunodeficiency (reviewed in [6]). However, these only account for a minority of cases (10–15 %), suggesting that additional genes contribute to the development of the defect. The same issue is noted in patients with IgA deficiency, the most common form of PID in Caucasians, where several genes [25] have been found to be associated with the deficiency, yet, the associated variants are only found in a minority of patients. Thus, a search for additional disease-causing genes is clearly warranted in patients with different forms of PID. However, although the methodology (from single gene sequencing, to exome sequencing and ultimately whole genome sequencing) has improved the success rate in finding disease-causing mutations, we strongly feel that a multi-omics approach (including epigenetics, transcriptomics, proteomics, functional pathway testing and analysis of the microbiome) is now needed to supplement the genetic data and thus, to further the PID field.

Compliance with Ethical Standards

Authorship Contributions M.F. performed bioinformatics analysis, results interpretation, made the figures and wrote the paper; H.A. performed experiments, analyzed results and wrote the paper; C.K. analyzed results and wrote the paper. J. Z. performed the WES and L.H. designed the project and edited the paper.

Disclosure of Conflicts of Interest None.

References

1. Routes J, Abinun M, Al-Herz W, Bustamante J, Condino-Neto A, De La Morena MT, et al. ICON: the early diagnosis of congenital immunodeficiencies. Journal of Clinical Immunology. 2014;34(4): 398–424.
2. Joshi AY, Iyer VN, Hagan JB, St Sauver JL, Boyce TG. Incidence and temporal trends of primary immunodeficiency: a population-based cohort study. Mayo Clinic Proceedings. 2009;84(1):16–22.

3. Hammarstrom L, Vorechovsky I, Webster D. Selective IgA deficiency (SIgAD) and common variable immunodeficiency (CVID). Clinical and Experimental Immunology. 2000;120(2):225–31.

4. van den Berg JM, van Koppen E, Ahlin A, Belohradsky BH, Bernatowska E, Corbeel L, et al. Chronic granulomatous disease: the European experience. PLoS ONE. 2009;4(4):e5234.

5. Bousfiha AA, Jeddane L, Ailal F, Benhsaien I, Mahlaoui N, Casanova JL, et al. Primary immunodeficiency diseases worldwide: more common than generally thought. Journal of Clinical Immunology. 2013;33(1):1–7.

6. Picard C, Al-Herz W, Bousfiha A, Casanova JL, Chatila T, Conley ME, et al. Primary Immunodeficiency Diseases: an Update on the Classification from the International Union of Immunological Societies Expert Committee for Primary Immunodeficiency 2015. Journal of Clinical Immunology. 2015.

7. Bousfiha A, Jeddane L, Al-Herz W, Ailal F, Casanova JL, Chatila T, et al. The 2015 IUIS phenotypic classification for primary immunodeficiencies. Journal of Clinical Immunology. 2015;35(8):727–38.

8. Itan Y, Casanova JL. Novel primary immunodeficiency candidate genes predicted by the human gene connectome. Frontiers in Immunology. 2015;6:142.

9. Germeshausen M, Zeidler C, Stuhrmann M, Lanciotti M, Ballmaier M, Welte K. Digenic mutations in severe congenital neutropenia. Haematologica. 2010;95(7):1207–10.

10. Zhang K, Chandrakasan S, Chapman H, Valencia CA, Husami A, Kissell D, et al. Synergistic defects of different molecules in the cytotoxic pathway lead to clinical familial hemophagocytic lymphohistiocytosis. Blood. 2014;124(8):1331–4.

11. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009;461(7261):272–6.

12. Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science (New York, NY). 2010;328(5978):636–9.

13. Wang JL, Yang X, Xia K, Hu ZM, Weng L, Jin X, et al. TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. Brain: a Journal of Neurology. 2010;133(Pt 12):3510–8.

14. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, et al. Exome sequencing identifies the cause of a mendelian disorder. Nature Genetics. 2010;42(1):30–5.

15. Lupski JR, Reid JG, Gonzaga-Jauregui C, Rio Deiros D, Chen DC, Nazareth L, et al. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. The New England Journal of Medicine. 2010;362(13):1181–91.

16. Bonnefond A, Durand E, Sand O, De Graeve F, Gallina S, Busiah K, et al. Molecular diagnosis of neonatal diabetes mellitus using next-generation sequencing of the whole exome. PLoS ONE. 2010;5(10):e13630.

17. Notarangelo LD. Primary immunodeficiencies. The Journal of Allergy and Clinical Immunology. 2010;125(2 Suppl 2):S182–94.

18. Alkhairy OK, Perez-Becker R, Driessen GJ, Abolhassani H, van Montfrans J, Borte S, et al. Novel mutations in TNFRSF7/CD27: clinical, immunologic, and genetic characterization of human CD27 deficiency. The Journal of Allergy and Clinical Immunology. 2015.

19. Abolhassani H, Cheraghi T, Rezaei N, Aghamohammadi A, Hammarstrom L. Common variable immunodeficiency or late-onset combined immunodeficiency: A new hypomorphic JAK3 patient and review of the literature. Journal of Investigational Allergology & Clinical Immunology. 2015;25(3):218–20.

20. Alkhairy OK, Abolhassani H, Rezaei N, Fang M, Andersen KK, Chavoshzadeh Z, et al. *Spectrum* of Phenotypes Associated with Mutations in LRBA. Journal of clinical immunology. 2015.

21. Abolhassani H, Wang N, Aghamohammadi A, Rezaei N, Lee YN, Frugoni F, et al. A hypomorphic recombination-activating gene 1 (RAG1) mutation resulting in a phenotype resembling common variable immunodeficiency. The Journal of Allergy and Clinical Immunology. 2014;134(6):1375–80.

22. Alkhairy OK, Rezaei N, Graham RR, Abolhassani H, Borte S, Hultenby K, et al. RAC2 loss-of-function mutation in 2 siblings with characteristics of common variable immunodeficiency. The Journal of allergy and clinical immunology. 2015;135(5):1380–4.e1–5.

23. Volk T, Pannicke U, Reisli I, Bulashevska A, Ritter J, Bjorkman A, et al. DCLRE1C (ARTEMIS) mutations causing phenotypes ranging from atypical severe combined immunodeficiency to mere antibody deficiency. Human Molecular Genetics. 2015;24(25):7361–72.

24. Salzer U, Bacchelli C, Buckridge S, Pan-Hammarstrom Q, Jennings S, Lougaris V, et al. Relevance of biallelic versus monoallelic TNFRSF13B mutations in distinguishing disease-causing from risk-increasing TNFRSF13B variants in antibody deficiency syndromes. Blood. 2009;113(9):1967–76.

25. Ferreira RC, Pan-Hammarstrom Q, Graham RR, Gateva V, Fontan G, Lee AT, et al. Association of IFIH1 and other autoimmunity risk alleles with selective IgA deficiency. Nature Genetics. 2010;42(9):777–80.

26. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics (Oxford, England). 2010;26(16):2069–70.

27. Guo Y, Zhao S, Lehmann BD, Sheng Q, Shaver TM, Stricker TP, et al. Detection of internal exon deletion with exon del. BMC Bioinformatics. 2014;15:332.

28. Shi Y, Majewski J. FishingCNV: a graphical software package for detecting rare copy number variations in exome-sequencing data. Bioinformatics. 2013;29(11):1461–2.

29. Keerthikumar S, Raju R, Kandasamy K, Hijikata A, Ramabadran S, Balakrishnan L, et al. RAPID: resource of Asian primary immunodeficiency diseases. Nucleic Acids Research. 2009;37(Database issue):D863–7.

30. Casanova JL, Conley ME, Seligman SJ, Abel L, Notarangelo LD. Guidelines for genetic studies in single patients: lessons from primary immunodeficiencies. The Journal of Experimental Medicine. 2014;211(11):2137–49.

31. Lamperti C, Fang M, Invernizzi F, Liu X, Wang H, Zhang Q, et al. A novel homozygous mutation in SUCLA2 gene identified by exome sequencing. Molecular Genetics and Metabolism. 2012;107(3):403–8.

32. Hu H, Roach JC, Coon H, Guthery SL, Voelkerding KV, Margraf RL, et al. A unified test of linkage analysis and rare-variant association for analysis of pedigree sequence data. Nature Biotechnology. 2014;32(7):663–9.

33. Koboldt DC, Larson DE, Wilson RK. Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection. Current protocols in bioinformatics/editoral board, Andreas D Baxevanis [et al]. 2013;44:15.4.1–.4.7.

34. Moreau Y, Tranchevent LC. Computational tools for prioritizing candidate genes: boosting disease gene discovery. Nature Reviews Genetics. 2012;13(8):523–36.

35. Yang H. Robinson PN. Phenolyzer: Phenotype-Based Prioritization of Candidate Genes for Human Diseases. 2015;12(9):841–3.

36. Javed A, Agrawal S, Ng PC. Phen-gen: combining phenotype and genotype to analyze rare disorders. Nature Methods. 2014;11(9):935–7.

37. Shyr C, Tarailo-Graovac M, Gottlieb M, Lee JJ, van Karnebeek C, Wasserman WW. FLAGS, frequently mutated genes in public exomes. BMC Medical Genomics. 2014;7:64.

38. Itan Y, Shang L, Boisson B, Patin E, Bolze A, Moncada-Velez M, et al. The human gene damage index as a gene-level

approach to prioritizing exome variants. Proceedings of the National Academy Of Sciences of the United States Of America. 2015;112(44):13615–20.

39. Vorechovsky I, Luo L, Hertz JM, Froland SS, Klemola T, Fiorini M, et al. Mutation pattern in the Bruton's tyrosine kinase gene in 26 unrelated patients with X-linked agammaglobulinemia. Human Mutation. 1997;9(5):418–25.

40. Conley ME, Mathias D, Treadaway J, Minegishi Y, Rohrer J. Mutations in btk in patients with presumed X-linked agammaglobulinemia. American Journal of Human Genetics. 1998;62(5):1034–43.

41. Holinski-Feder E, Weiss M, Brandau O, Jedele KB, Nore B, Backesjo CM, et al. Mutation screening of the BTK gene in 56 families with X-linked agammaglobulinemia (XLA): 47 unique mutations without correlation to clinical course. Pediatrics. 1998;101(2):276–84.

42. Valiaho J, Faisal I, Ortutay C, Smith CI, Vihinen M. Characterization of all possible single-nucleotide change caused amino acid substitutions in the kinase domain of Bruton tyrosine kinase. Human Mutation. 2015;36(6):638–47.

43. Zeng H, Tao Y, Chen X, Zeng P, Wang B, Wei R, et al. Primary immunodeficiency in South China: clinical features and a genetic subanalysis of 138 children. Journal of Investigational Allergology & Clinical Immunology. 2013;23(5):302–8.

44. Lee WI, Torgerson TR, Schumacher MJ, Yel L, Zhu Q, Ochs HD. Molecular analysis of a large cohort of patients with the hyper immunoglobulin M (IgM) syndrome. Blood. 2005;105(5):1881–90.

45. Clark MJ, Chen R, Lam HY, Karczewski KJ, Chen R, Euskirchen G, et al. Performance comparison of exome DNA sequencing technologies. Nature Biotechnology. 2011;29(10):908–14.