

# IEIVariantFilter: a bioinformatics tool to speed up genetic diagnosis of inborn errors of immunity patients

Juan Pereda<sup>1</sup>, Rafael Espinosa<sup>1</sup>, Blanca García-Solís<sup>2,3,4</sup>, Teresa Guerra-Galán<sup>5</sup>, Ana Van-Den-Rym<sup>2,3,4</sup>, Meltem Ece Kars<sup>6</sup>, Rocío Mena<sup>7</sup>, Víctor Galán<sup>8</sup>, Ana de Andrés-Martín<sup>9</sup>, Carlos Rodríguez-Gallego<sup>10,11,12,13</sup>, Alberto López-Lera<sup>14</sup>, Fernando Corvillo<sup>14</sup>, Antonio Pérez-Martínez<sup>8</sup>, Eduardo López-Collazo<sup>3</sup>, Silvia Sánchez-Ramón<sup>4,5</sup>, Rubén Martínez-Barricarte<sup>15,16</sup>, Lluís Quintana-Murci<sup>17</sup>, José Miguel Lorenzo-Salazar<sup>11,18</sup>, Yuval Itan<sup>6,†</sup>, Carlos Flores<sup>11,13,18,19,†</sup>, Rebeca Pérez-de-Diego<sup>11,2,3,4,11,\*</sup>

<sup>1</sup>TECH ID Solutions, Madrid 28049, Spain

<sup>2</sup>Laboratory of Immunogenetics of Human Diseases, IdiPAZ Institute for Health Research, La Paz University Hospital, Madrid 28046, Spain

<sup>3</sup>Innate Immunity Group, IdiPAZ Institute for Health Research, La Paz University Hospital, Madrid 28046, Spain

<sup>4</sup>Interdepartmental Group of Immunodeficiencies, Madrid, Spain

<sup>5</sup>Clinical Immunology Department, San Carlos Clinical Hospital, Madrid 28040, Spain

<sup>6</sup>The Charles Bronfman Institute for Personalized Medicine and Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, United States

<sup>7</sup>Institute of Medical & Molecular Genetics (INGEMM), Hospital Universitario La Paz, Universidad Autónoma de Madrid, IdiPAZ, Madrid 28046, Spain

<sup>8</sup>Translational Research in Paediatric Oncology, Haematopoietic Transplantation and Cell Therapy, IdiPAZ Institute for Health Research, La Paz University Hospital, Madrid 28046, Spain

<sup>9</sup>Department of Immunology, Ramón y Cajal Hospital, Madrid 28034, Spain

<sup>10</sup>Department of Immunology, University Hospital of Gran Canaria Dr. Negrin, Las Palmas de Gran Canaria 35010, Spain

<sup>11</sup>CIBER de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III, Madrid, Spain

<sup>12</sup>Department of Medical and Surgical Sciences, School of Medicine, University of Las Palmas de Gran Canaria, Las Palmas de Gran Canaria 35016, Spain

<sup>13</sup>Department of Clinical Sciences, University Fernando Pessoa Canarias, Las Palmas de Gran Canaria 35450, Spain

<sup>14</sup>IdiPAZ Institute for Health Research, La Paz University Hospital, CIBERER U-754, Madrid 28046, Spain

<sup>15</sup>Division of Genetic Medicine, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN 37232, United States

<sup>16</sup>Division of Molecular Pathogenesis, Department of Pathology, Microbiology, and Immunology, Vanderbilt Center for Immunobiology, Vanderbilt Institute for Infection, Immunology, and Inflammation, Vanderbilt University Medical Center, Nashville, TN 37232, United States

<sup>17</sup>Unit of Human Evolutionary Genetics, Institut Pasteur and CNRS URA3012, Paris 75015, France

<sup>18</sup>Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife 38600, Spain

<sup>19</sup>Research Unit, Hospital Universitario Ntra. Sra. de Candelaria, Instituto de Investigación Sanitaria de Canarias (IISC), Santa Cruz de Tenerife 38010, Spain

\*To whom correspondence should be addressed. Email: [rebeca.perez@idipaz.es](mailto:rebeca.perez@idipaz.es)

†These authors contributed equally to the manuscript.

## Abstract

Severe infectious diseases remain the leading cause of death in children and young adults worldwide. Monogenic inborn errors of immunity (IEIs) are traditionally defined as a heterogeneous group of rare inborn genetic diseases affecting the functioning of the immune system. Greater awareness has led to the clinical definition of 485 monogenic IEIs and whole exome sequencing (WES) is becoming increasingly relevant for IEI genetic diagnosis. The current protocol for IEI genetic studies includes manual filtering of the list of genes obtained as a WES read-out providing a short list of candidate genes. This procedure is time-consuming and can produce mistakes due to human error in manual filtering. IEIVariantFilter is a new web-based bioinformatics tool to speed up and refine the genetic diagnosis of IEI patients oriented for users in the biomedical field without needing bioinformatics expertise. IEIVariantFilter prioritizes genetic variants based on ranges of zygosity, the quality of reads, the predicted variant effect, and genes related to immunity, considering a consanguineous hypothesis whenever necessary. IEIVariantFilter facilitates gene and variant list prioritization, speeding up the identification of candidate disease-causing variants for validation by experimental studies. The software improves the genetic diagnosis of patients, thereby facilitating precision medicine and fast and proper treatment.

Received: January 17, 2025. Revised: April 25, 2025. Editorial Decision: May 12, 2025. Accepted: May 14, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of NAR Genomics and Bioinformatics.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

## Introduction

Inborn errors of immunity (IEIs) are a heterogeneous group of diseases caused by quantitative and/or functional changes in the different mechanisms involved in both the innate and the adaptive immune responses [1]. They are classified as primary immunodeficiencies (PIDs), when their origin is genetic, and secondary immunodeficiencies (SIDs), when their origin is acquired. Both types are associated with or cause a predisposition towards clinical complications, such as severe or recurrent infections, autoimmune disorders, immune dysregulation with lymphoproliferation, inflammatory disorders, lymphomas, and other types of cancer, many of which are diagnosed and treated by rheumatologists, haematologists, and oncologists [1–5]. IEIs were traditionally defined as rare in-born genetic diseases affecting the functioning of the immune system. However, it has become apparent in recent years that IEIs are much more common than initially thought, and a high percentage of patients with severe infections since childhood have an IEI [2, 3, 6]. Greater awareness and improved collaboration in recent years have led to the clinical definition of 485 monogenic IEIs [1]. IEI studies have greatly improved our understanding of human immunology, providing a framework for studies of physiologically relevant genes in the human context [4, 7]. In addition, whole exome sequencing (WES), i.e. the targeted sequencing of the protein-coding portion of the human genome, has proven to be a powerful and efficient method for identifying disease-causing variants underlying Mendelian disorders [8]. It has been estimated that the protein-coding regions of the human genome account for ~85% of all described disease-causing variants [9]. The use of WES is becoming increasingly relevant for IEI genetic diagnosis as our understanding of the functional consequences of sequence variation improves [9]. Researchers who typically confront the WES data analyses of IEI patients come from the biological, chemical, medical, or pharmaceutical fields, as it is their knowledge of the biology of the immune system that yields new candidate genes to find new genetic defects. However, the use of some software or the design of scripts for filtering requires bioinformatics skills that prevent these researchers from using bioinformatics tools, which increases the risk of errors in the first steps of WES filtering.

Despite recent advances, the genetic basis of a significant proportion of IEIs remains undefined. The continuous improvement of technologies for the genetic diagnosis of diseases is unprecedented and the impact of this exponential improvement is further amplified when these technologies coalesce into open platforms and ecosystems. For instance, new categories of technologies in molecular biology and materials science are being combined, leading to advances and radical changes in approaches in an ever-increasing number of industry functions and fields. Several algorithms have been developed to help in WES data interpretation, such as the CADD (Combined Annotation Dependent Depletion) score and the REVEL (Rare Exome Variant Ensemble Learner) that predict deleteriousness/pathogenicity of genetic variants [10, 11]. Another relevant tool to help discover a large proportion of IEI-causing genes in patients is the human gene connectome and the interactome approach, which shows the biological distance between all human genes and can be used to prioritize candidate genes by their relatedness to known disease-causing genes [12]. All these tools and algorithms, together with new high-throughput sequencing techniques, represent a powerful strategy for the identification of new genetic aetiolo-

gies of IEIs. However, even today, filtering such data is mainly performed manually, increasing the risk of human errors and slowing down the analysis. Yet in line with one of the objectives of the International Rare Diseases Research Consortium (IRDRC) [13], IEIVariantFilter is a tool with a pipeline tailored specifically for IEI, since the genetic causes of different disease groups are extremely heterogeneous. IEIVariantFilter provides easy-to-use automatic variant filtering, offering the user without needing bioinformatics expertise a shortened list of candidate genes for further detailed study by the researcher, thus achieving faster filtering, allowing a genetic diagnosis and rapid clinical decisions for IEI patients.

## Materials and methods

### Study approval

The experimental protocol was approved by the Ethics Committee of La Paz University Hospital (Madrid, Spain) and written informed consent was obtained from the patient's family for participation in this study.

### Platform architecture

We developed the IEIVariantFilter, a tool based on features of genetic variants obtained as a read-out of WES data from patients and controls that allows filtering of these variants. Due to the limitations of our server, the tool can only be used for WES files and not whole genome sequencing (WGS) files, which are much larger. The tool accepts a tabular-format input file containing annotated variants through a web-based graphic user interface with a number of annotation fields (see below). The user is asked to select the reference genome assembly used for the annotation (GRCh37.p13 or GRCh38.p14 are available at the moment). Only WES files from the same human genome assembly can be used for filtering.

The software uses samples as controls to filter 'benign' features; these samples can be healthy controls or patients with different pathology from the patient analysed (pseudo-controls).

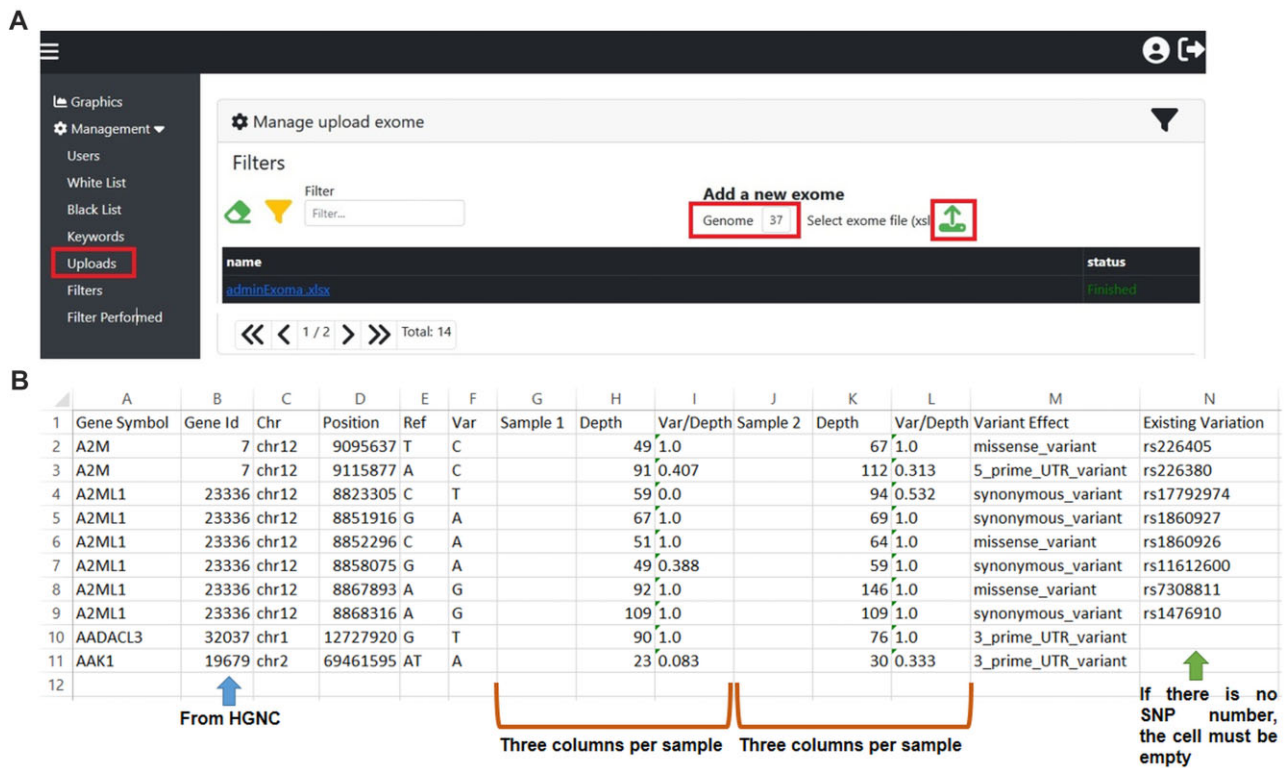
WES result files must be uploaded to the database following the instructions in the User Guide (see [supplementary material and website](#)). It is installed on a server providing security password-controlled access for users after requesting permission.

These data will not be shared between users following data protection rules and informed consent.

### WES file filtering tool

The filtering tool was developed in PHP for the front part, in server-side scripting language for the frontend (Apache server), and in Java for communications and the backend (Tomcat server) ([Supplementary Fig. S1](#)).

The user manually inputs a gene to compare or a tabular format Excel-like files with a list of genes to compare through the graphical interface provided by the tool. Subsequently, the platform searches within the database for the correlation of any gene that matches those previously entered. Due to the size of the search, the Spring Batch framework is used, which allows processing the information in batches ([Supplementary methods](#) and [Supplementary Fig. S2](#)). Before conducting the search and applying filters, the user must select the filters to apply, which include the patient or patients, the controls, and the consanguineous parents (if applicable).



**Figure 1.** Uploading WES files. **(A)** Software interface highlighting the uploading of WES files in red. **(B)** Format for .xlsx files to be uploaded. The columns should contain the following data: *Gene Symbol*; *Gene Id* (from HGNC); *Chr* (chromosome); *Position* (position number in the genome); *Ref* (reference allele); and *Var* (variant allele). Three more columns contain the data from each sample: *Sample 1* (patient identification, this column is empty); *Depth* (number of total reads of sample 1 for the variant indicated in each file); and *Var/Depth* (allele balance of sample 1 for the variant indicated in each file). Another column contains the *Variant Effect* (if the variant effect is 5' UTR, 3' UTR, or a synonymous variant, it should be indicated exactly in these formats: 5\_prime\_UTR\_variant, 3\_prime\_UTR\_variant, or synonymous\_variant) and *Existing Variation* (if there is no SNP number, the box must be left empty).

The software includes variant filtering based on (i) filtering variants by range of zygosity (heterozygous/homozygous), quality of reads, genes included in the blacklist (genes not related to immunology), 3' and 5' UTRs, and synonymous variants; (ii) highlighting compound heterozygous variants and genes in the white list, a list of genes involved or related to immunology (7393 genes, [Supplementary Table S1](#)) [14]; and (iii) considering recessive hypothesis homozygous filtering for the patient and heterozygous filtering for parents.

To refine the search results, the following approximations are leveraged:

1. **Indexing:** We have created indices on columns that are frequently used in the WHERE, JOIN, or ORDER BY clauses of queries. These allow the database engine to efficiently find the data without need to scan the entire table.
2. **Query optimization:** We have reviewed and optimized SQL queries to ensure that they are as efficient as possible. This includes selecting only the necessary columns, avoiding unnecessary subqueries, and using appropriate joins.
3. **Result caching:** We have implemented a caching solution to store the results of frequently executed queries. This helps to significantly reduce the response time for these queries by avoiding the need to access the database each time.

## Results

### Access to the software website, uploading WES files, and filtering steps

The software is accessible at the following URL: <http://185.49.184.146:8080/exome-web/login?redirect=%2Fapp%2Fdashboard>. As the web server is paid for our lab projects, its capacity is limited, and user access must be requested. Upon register via the webpage, a user will be created, and credentials will be provided by email. Users can then upload the WES files of their samples. Each user will be allowed to analyse their own uploaded data. The data entered by each user are private and can only be accessed by the user themselves, ensuring that the results remain completely confidential. Upon request to the correspondence author, users can be connected to each other confidentially so that they can share data.

Tabular format (Excel files) input files can be filtered to exclude common variants and intronic variants beforehand for faster uploading. A User Guide provides detailed instructions for uploading WES files and the recommended steps for filtering (see [supplementary material or website](#); Figs 1 and 2).

### Downloading results after filtering

After defining and submitting the required filters, the tool provides a filtered list of variants that is downloadable from the *Filter Performed* section (Fig. 3). The User Guide (see the

**Manage filters**

Title\* **Write title of the filter**

Analysed with genome 37 Show patients Only mine

Legend: P=Patient, C=Control, CP=Consanguineous Parent

Identificator	P	C	CP	User	Genome
RP0072	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	admin	37
RP0095	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	admin	37
RP0096	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	admin	37

<< < 1 / 5 > >> Total: 46

☐ Remove reference variants in patients.  
☐ Remove hom alt in controls. Conditions: Var/Depth=1 and Depth> 7 or Var/Depth> 0.9 and Depth> 20 .  
☐ Remove genes in blacklist.  
☐ Remove simple het variants common in controls. Consider heterozygote when Var/Depth is between 0.3 -> 0.7 and Depth> 20 .  
☐ Highlight in pink compound het variants. Consider heterozygote when Var/Depth is between 0.2 -> 0.9 and Depth> 7 .  
☐ Consanguineous hypotesis.  
☐ Remove 3' and 5' UTR zones.  
☐ Remove synonymous.  
☐ Match PID and II lists and highlight in green.

Submit

**Filtering options**

**Figure 2.** Filtering of WES samples. WES samples are filtered in the *Filters* option. After naming the filter, patients (P), controls (C), and consanguineous parents (CP) (the latter in the case of a Trio WES with hypothesis of consanguinity) are selected. Different checkboxes (yellow arrows in the figure) allow WES files to be filtered (see Results: 'Steps for filtering' for a description of each).

**A**

**Filter Performed**

Filters

Status File Pending Filters

Finished	adminis	
Finished	adminisab	
Finished	adminFilter.RP0095 Software 2023	

<< < 1 / 4 > >> Total: 39

**B**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	Gene Symbol	Gene Id	Chr	Position	Ref	Var	Sample 1	Depth	Var/Depth	Sample 2	Depth	Var/Depth	Variant Effects	Existing Variant
1	CR2	2336	chr1	207627693	T	C	1st	7	0,2857143		19	0,4210526	START_GAINED	rs3813946
2	KRT18	6430	chr12	53343059	C	A		25	0,24		24	0,4166667	NON_synonymous_variant	rs78343594
3	NCOA3	7670	chr20	46279836	ACAG	A		296	0,3648649		569	0,3778559	CODON_DELETION	rs147879509
4	NCOR2	7673	chr12	124887058	G	GGCT		123	0,8617886		207	0,7826087	CODON_INSERTION	rs35831183
5	RAPGEF1	4568	chr9	134585181	C	T	2nd	12	0,5833333		48	0,4166667	START_GAINED	rs11243480
6	IGSF5	5952	chr21	41151154	C	CGCTGCT	3rd	159	0,3584906		303	0,4752475	CODON_INSERTION	
7	MUC6	7517	chr11	1017084	G	A		430	0,2581395		517	0,1992263	NON_synonymous_variant	rs34649796
8	MYO5B	7603	chr18	47363931	G	C	4th	10	0,45		0	0	NA	rs77840018
10														

**Figure 3.** Filter performed. (A) Software interface for the *Filter Performed* section. Files are downloaded in Excel-like format (yellow arrow). (B) The Excel spreadsheet obtained after filtering, organized in the subsequent columns: *Gene Symbol*; *Gene Id*; *Chr*; *Position*; *Ref*; and *Var*. The next three columns show the data for each subject: *Sample* (patient identification); *Depth* (number of total sample reads); *Var/Depth* (allele balance); *Variant Effect*; and *Existing Variation*. Green boxes indicate genes in the white list and pink boxes indicate heterozygous compound variants.



A

**Manage filters**

Title\*  
Validation test\_ patient Ezr

Analysed with genome 38 Show patients Only mine

Legend: P=Patient, C=Control, CP=Consanguineous Parent

Identificator	P	C	CP	User	Genome
RP0042	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	admin	38
RP0043	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	admin	38
RP0045	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	admin	38
RP0046	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	admin	38
RP0047	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	admin	38
RP0048	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	admin	38
RP0049	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	admin	38
RP0050	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	admin	38
RP0057	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	admin	38
RP0058	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	admin	38

<< < 4 / 6 > >> Total: 55

☒ Remove reference variants in patients.

☒ Remove hom alt in controls. Conditions: Var/Depth=1 and Depth> 7 or Var/Depth> 0.9 and Depth> 15 .

☒ Remove genes in blacklist.

☒ Remove simple het variants common in controls. Consider heterozigote when Var/Depth is between 0.3 -> 0.7 and Depth> 15 .

☒ Highlight in pink compound het variants.Consider heterozigote when Var/Depth is between 0.2 -> 0.9 and Depth> 7 .

☐ Consanguineous hypotesis.

☒ Remove 3' and 5' UTR zones.

☒ Remove synonymous.

☒ Match PID and II lists and highlight in green.

Submit

B

**Filter Performed**

Filters

Filter  
Filter...

Status	File	Pending Filters
Finished	adminValidation.test_patient Ezr	

<< < 1 / 5 > >> Total: 42

**Figure 4.** Case study on a patient with IEI. (A) Software user interface for filtering data from a patient reported as a carrier of the *EZR* deficiency [15]. (B) Result of data filtering for the *EZR* carrier patient in the validation test.

**Table 1.** Case study on a patient with IEI

A										
Gene symbol	Gene Id	Chr	Position	Ref	Var	RP0047	Depth	Var/ Depth	Variant effects	Existing variant
<i>HLA-B</i>	4932	chr6	31356431	G	A	P_Homo_var	2	1	missense_variant	Null
<i>PDE4DIP</i>	15580	chr1	148870282	G	T	P_Homo_var	2	1	missense_variant	rs12118314
<i>PDE4DIP</i>	15580	chr1	148870283	T	G	P_Homo_var	2	1	missense_variant	rs12119750
<i>MAML2</i>	16259	chr11	96093510	C	T	P_Homo_var	36	1	missense_variant	rs61749254
<i>LILRB3</i>	6607	chr19	54222457	T	C	P_Homo_var	14	1	missense_variant	rs200199363
<i>EZR</i>	12691	chr6	158785391	C	T	P_Homo_var	29	1	missense_variant	rs528409234
<i>MAP3K4</i>	6856	chr6	161070652	G	T	P_Homo_var	21	1	missense_variant	rs34018542
<i>SRP9</i>	11304	chr1	225786860	A	T	P_Homo_var	1	1	splice_acceptor_variant	Null
B										
<i>MAML2</i>	16259	chr11	96093510	C	T	P_Homo_var	36	1	missense_variant	rs61749254
<i>LILRB3</i>	6607	chr19	54222457	T	C	P_Homo_var	14	1	missense_variant	rs200199363
<i>EZR</i>	12691	chr6	158785391	C	T	P_Homo_var	29	1	missense_variant	rs528409234
<i>MAP3K4</i>	6856	chr6	161070652	G	T	P_Homo_var	21	1	missense_variant	rs34018542

(A) Homozygous variants present in the white list of immune genes for the patient analysed in the validation test. (B) Homozygous variants present in the white list of immune genes, removing those with a low number of reads. Ref: reference nucleotide and Var: variant nucleotide.

supplementary material or website) describes how the list of filtered variants is returned to the user.

### Case study on a patient with IEI

We tested the software with WES data from a patient with a known homozygous disease-causing variant in *EZR* [15]. The patient has consanguineous parents, but WES data were lacking from the parents. In *Manage Filters*, we selected the patient (ID: RP0047) and available data from several samples to serve as controls (Fig. 4A). The following options were selected for filtering:

- *Remove reference variants in patients*: This removed variants that were homozygous reference in the patient analysed.
- *Remove hom alt in controls*: This removed homozygous variants in controls, as the premise establishes that a homozygous variant in unaffected cannot be responsible for the disease in the patient.
- *Remove genes in the blacklist*: This checkbox removed all variants of genes included in the blacklist.
- *Remove simple het variants common in controls*: This removed variants in heterozygosis that were present in patients and in controls; the premise is that a simple heterozygous variant in a patient that appears in unaffected is not responsible for the disease.
- *Highlight in pink compound het variants*: This highlights in pink colour the variants that are compound heterozygous at the range of Var/Depth ratio and depth established by the user.
- *Remove 3' and 5' UTR zones and remove synonymous*.
- *Match PID and II list and highlight in green*: This checks which variants are in genes that match a white list of genes related to PID, innate immunity, and infectious diseases.

After submitting the filtering, the result is downloaded from *Filter Performed* (Fig. 4B). For this patient, a total of 1770 variants were obtained (Supplementary Table S2). From these, 592 variants were highlighted in green, as they were genes related to immunology. Since the patient is consanguineous, we selected the homozygous variants highlighted in the white list and then eight variants were obtained (Table 1A). Further filtering of these eight variants in terms of quality of reads (we

removed variants with <10 reads) left us with four variants, one in each of the following genes: *MAML2*, *LILRB3*, *EZR*, and *MAP3K4* (Table 1B). The clinical features of the patient fitted the *EZR* variant better, making it the best candidate gene for downstream analyses.

### Discussion

IEIs are a heterogeneous group of different (>400) immunity disorders that impair the functions of the human immune system, causing greater susceptibility to infection, inflammation, autoimmunity, allergy, and malignancy. These diseases pose a challenge for diagnosis and treatment due to overlapping symptoms and similarities between diseases. IEIs are thought to affect at least 10 million people worldwide [16]. WES/WGS combined with the existing knowledge can provide a genetic aetiology around 50% of IEI patients assessed [17]. Therefore, IEIs should no longer be considered as rare diseases due to the number of undiagnosed patients. The International Union of Immunological Societies Inborn Errors of Immunity Committee only reports single-gene inborn errors [1]. However, IEI experts around the world suspect that the high percentage of undiagnosed patients is due to complex genetic scenarios such as oligo- or polygeny. If IEIs are left un/misdiagnosed, the immune system remains compromised, leading to chronic illness, disability, reduced working capacity, decreased quality of life for patients and families, permanent organ damage, or even death. Despite the clear relevance of these disorders, there is still an urgent need for a better understanding of their causes and early detection, to design better therapies, improve prognosis, and provide accurate genetic counselling to the family. The current procedure for IEI studies includes filtering the list of genes obtained as a WES/WGS read-out, resulting in a short list of candidate genes that are then experimentally validated. This procedure is slow and can lead to human errors associated with manual filtering.

The researchers who typically analyse the IEI patients' WES data come from the biological, chemical, medical, or pharmaceutical fields, and their knowledge of the human immune system biology is leveraged to filter the WES-provided list of variants to find new genetic defects. In many cases, the researchers have difficulties to deal with complex bioinformatics tools or to develop scripts for WES filtering, and they have to make all



dures and editing). R.P.D.: Conceptualization and Writing – original draft (head of the laboratory, experiment and software design, manuscript drafting and editing, and corresponding author).

## Supplementary data

Supplementary data is available at NAR Genomics & Bioinformatics online.

## Conflict of interest

None declared.

## Funding

This study was supported by Carlos III Health Institute (ISCIII) through the project PI22/00790 and co-funded by the European Union, Ayudas Luis \_Alvarez 2022 FIBHULP. B.G.S. was supported by PEJD2019-PRE/BMD-16556 Pre-doctoral Fellowships CAM and ESID Bridge Fellowship. A.V.D.R. received support from Carlos III Health Institute (ISCIII) through the project PI22/00790. R.M.B. was funded in part the National Institute of Allergy and Infectious Diseases and the National Cancer Institute of the National Institutes of Health (grant R21AI171466, #1R01CA269217-01A1, 1R01AI168210-01A1). C.F. was supported by the Carlos III Health Institute (CB06/06/1088 and PI23/00980) and co-financed by the European Regional Development Funds, ‘A way of making Europe’ from the European Union, ITER agreements (OA17/008 and OA23/043) and the Cabildo Insular de Tenerife local government (CGIEU0000219140 and ‘Apuestas científicas del ITER para colaborar en la lucha contra la COVID-19’). Y.I. and M.E.K. were funded by the Charles Bronfman Institute for Personalized Medicine at Icahn School of Medicine at Mount Sinai, and the New York Regional Inborn Errors of Immunity Resource Initiative League (NY-ROYAL, 1R24AI167802-01A1).

## Data availability

IEIVariantFilter is available at <http://185.49.184.146:8080/exome-web/login?redirect=%2Fapp%2Fdashboard>.

## References

1. Tangye SG, Al-Herz W, Bousfiha A *et al.* Human inborn errors of immunity: 2022 update on the classification from the International Union of Immunological Societies Expert Committee. *J Clin Immunol* 2022;42:1473–507. <https://doi.org/10.1007/s10875-022-01289-3>
2. Casanova JL, Abel L. The genetic theory of infectious diseases: a brief history and selected illustrations. *Annu Rev Genomics Hum Genet* 2013;14:215–43. <https://doi.org/10.1146/annurev-genom-091212-153448>
3. Casanova JL, Abel L. Primary immunodeficiencies: a field in its infancy. *Science* 2007;317:617–9. <https://doi.org/10.1126/science.1142963>
4. Notarangelo LD. Primary immunodeficiencies. *J Allergy Clin Immunol* 2010;125:S182–94. <https://doi.org/10.1016/j.jaci.2009.07.053>
5. Maródi L, Notarangelo LD. Education and worldwide collaboration pays off. *Nat Immunol* 2007;8:323–4. <https://doi.org/10.1038/ni0407-323>
6. Akalu YT, Bogunovic D. Inborn errors of immunity: an expanding universe of disease and genetic architecture. *Nat Rev Genet* 2024;25:184–95. <https://doi.org/10.1038/s41576-023-00656-z>
7. Notarangelo LD, Gambineri E, Badolato R. Immunodeficiencies with autoimmune consequences. *Adv Immunol* 2006;89:321–70. [https://doi.org/10.1016/S0065-2776\(05\)89008-X](https://doi.org/10.1016/S0065-2776(05)89008-X)
8. Ng SB, Turner EH, Robertson PD *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 2009;461:272–6. <https://doi.org/10.1038/nature08250>
9. Choi M, Scholl UI, Ji W *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci USA* 2009;106:19096–101. <https://doi.org/10.1073/pnas.0910672106>
10. Schubach M, Maass T, Nazaretyan L *et al.* CADD v1.7: using protein language models, regulatory CNNs and other nucleotide-level scores to improve genome-wide variant predictions. *Nucleic Acids Res* 2024;52:D1143–54. <https://doi.org/10.1093/nar/gkad989>
11. Ioannidis NM, Rothstein JH, Pejaver V *et al.* REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am Hum Genet* 2016;99:877–85. <https://doi.org/10.1016/j.ajhg.2016.08.016>
12. Itan Y, Zhang SY, Vogt G *et al.* The human gene connectome as a map of short cuts for morbid allele discovery. *Proc Natl Acad Sci USA* 2013;110:5558–63. <https://doi.org/10.1073/pnas.1218167110>
13. Boycott KM, Rath A, Chong JX *et al.* International cooperation to enable the diagnosis of all rare genetic diseases. *Am Hum Genet* 2017;100:695–705. <https://doi.org/10.1016/j.ajhg.2017.04.003>
14. Deschamps M, Laval G, Fagny M *et al.* Genomic signatures of selective pressures and introgression from archaic hominins at human innate immunity genes. *Am Hum Genet* 2016;98:5–21. <https://doi.org/10.1016/j.ajhg.2015.11.014>
15. García-Solís B, Van Den Rym A, Martínez-Martínez L *et al.* Inherited human ezrin deficiency impairs adaptive immunity. *J Allergy Clin Immunol* 2023;152:997–1009. <https://doi.org/10.1016/j.jaci.2023.05.022>
16. U.S. Department of Health and Human Services, National Institutes of Health. PAR-08-206: Investigations on Primary Immunodeficiency Diseases (R01). <http://grants.nih.gov/grants/guide/pa-files/PAR-08-206.html>, (2009, date last accessed).
17. Platt CD, Zaman F, Bainter W *et al.* Efficacy and economics of targeted panel versus whole-exome sequencing in 878 patients with suspected primary immunodeficiency. *J Allergy Clin Immunol* 2021;147:723–6. <https://doi.org/10.1016/j.jaci.2020.08.022>