Title

Expanding the potential genes of inborn errors of immunity through protein interactions

Journal

BMC Genomics, 22(1)

ISSN

1471-2164

Authors

Khan, Humza A

Butte, Manish J

Publication Date

2021-12-01

DOI

10.1186/s12864-021-07909-3

Khan and Butte *BMC Genomics* (2021) 22:618
https://doi.org/10.1186/s12864-021-07909-3

BMC Genomics

RESEARCH Open Access

Expanding the potential genes of inborn errors of immunity through protein interactions

Humza A. Khan[1,2] and Manish J. Butte[1,2*]

Abstract

Background: Inborn errors of immunity (IEI) are a group of genetic disorders that impair the immune system, with over 400 genes described so far, and hundreds more to be discovered. To facilitate the search for new genes, we need a way to prioritize among all the genes in the genome those most likely to play an important role in immunity.

Results: Here we identify a new list of genes by linking known IEI genes to new ones by using open-source databases of protein-protein interactions, post-translational modifications, and transcriptional regulation. We analyze this new set of 2,530 IEI-related genes for their tolerance of genetic variation and by their expression levels in various immune cell types.

Conclusions: By merging genes derived from protein interactions of known IEI genes with transcriptional data, we offer a new list of candidate genes that may play a role in as-yet undiscovered IEIs.

Keywords: Inborn errors of immunity, Clinical immunology, Next-generation sequencing, Protein-protein interactions, Primary immunodeficiency

Introduction

IEIs are a collection of over 400 monogenic disorders with phenotypes of recurrent, severe or unusual infections, autoimmunity, and autoinflammation. The process of making a diagnosis in these patients requires the synthesis of the clinical phenotype, the results of immunological testing, and the results of genetic testing that identify the key pathogenic variant(s). Whole exome and genome sequencing have made the process of identifying genomic variants straightforward. But when this process cannot find a known pathogenic variant or even a known IEI gene, the process switches tracks to discovering whether a new gene might explain the disease phenotype. Paring down the list of thousands of genes and tens of thousands of variants in "non-clinical" genes/genes not yet identified as being important for a human disease, is an unsolved problem. The list of variants can be filtered to eliminate those variants that occur commonly in human populations, resulting in a shorter list of few hundreds to thousands of genes. However, beyond this step lies a painstaking process of choosing genes and testing their roles. Nowadays, over three dozen new IEI genes are discovered yearly [1]. Eventually, if the right gene has been identified, the ensuing process of validating the biochemical and immunological impacts of the identified variant(s) is well defined if arduous [2].

Here we describe a list of IEI-related genes that were gathered by associating known IEI proteins with new ones. A similar approach in 2015 created a list of over 3,000 IEI-related genes by linking 229 known (at the time) IEI genes to new ones by the human gene connectome (HGC) [3, 4]. The HGC was created by calculating genetic distance from the binding portion of the STRING

* Correspondence: mbutte@mednet.ucla.edu
[1]Division of Immunology, Allergy, and Rheumatology, Department of Pediatrics, University of California Los Angeles, 10833 Le Conte Ave, MDCC Building, Room 12-430, Los Angeles, CA 90095, USA
[2]Department of Microbiology, Immunology, and Molecular Genetics, University of California Los Angeles, 10833 Le Conte Ave, MDCC Building, Room 12-430, Los Angeles, CA 90095, USA

protein database [5]. Expanding gene lists for other collections of rare diseases by such an approach has been fruitful [6]. The trade-off with creating such a large list of 3,000 + genes (more than 15 % of the genome) is that many will be expected not to actually participate in non-redundant pathways of the immune response. Regardless, the list of known IEI genes since the 2015 paper has doubled, and the field needs an updated gene list.

In this work, we used the recently described OmniPath protein-interaction meta-database to reveal novel genes that are functionally related to IEI genes [7]. Specifically, we employed two routes to include putative IEI genes: (1) we analyzed annotated interactions encompassing transcriptional regulation and post-translational modifications (PTM) between two proteins, and (2) we analyzed pathways without functional annotation between all combinations of immunodeficiency genes present in the database.

Materials and methods

Known IEI gene list

We derived a list of 403 confirmed IEI genes from the 2019 International Union of Immunological Societies (IUIS) IEI classifications at https://s3-eu-west-1.amazonaws.com/wp-iuis/app/uploads/2019/12/20113228/IUIS-IEI-list-for-web-site-December-2019-003.xlsx.
We manually filtered the list of only include monogenic, Mendelian disorders. Our analyses grouped IEI genes based on the "Major Category" parameter. Our list is available at https://github.com/humzalikhan/omnipath_IEI.

Algorithm verification

We split our list of confirmed IEI genes into training datasets (80 %) and validation datasets (20 %). We then filtered our already built pathways and PTM/TF interactions datasets to only include the training dataset and then derived a list of associated genes. The percentage of the validation dataset that was re-discovered within the associated genes was calculated. This analysis was performed 1,000 times, with randomly sampled training and validation datasets each time. We extended this to an out-of-sample methodology where all the genes in a previously defined immune function category are left out and the genes in all other categories are used to rediscover the left-out genes.

Protein-interaction databases

We refer to Discovery pathway 1 as our method of collecting new genes by examining IEI genes with respect to transcriptional regulation and post-translational modifications (PTM) between two proteins. We refer to Discovery pathway 2 as our method of collecting new genes from all the pathways arising from all pairwise combinations of immunodeficiency genes present in the database.

OmniPath is a recently described meta-database of signaling pathways and protein interactions, integrating over 100 individual databases. We ran our list of 403 IEI genes into the OmniPathR package and infrastructure to derive the proteins most related to the relevant IEI gene products [8].

Discovery path 1

For analyzing transcriptional regulation, we imported interactions from the DoRothEA_A database [9]. IEI gene products were analyzed for their activity as transcription factors as well as their activity in inducing or repressing transcription of other genes. For post-translational modifications, we imported interactions from the SIGNOR database [10]. IEI gene products were analyzed for their activity as modifiers and as recipients of modifications.

Discovery path 2

Omnipath contains protein-protein association data from over 100 different databases. We queried whether a "pathway" could exist between all known IEI genes to other known IEI genes. If the Omnipath databases could identify a chain of protein interactions (a "pathway") linking pairwise each known IEI gene, we then collected the genes along that pathway. This was implemented via the all_shortest_paths function in OmnipathR; all of the shortest possible pathways between known IEI genes were investigated. Furthermore, to avoid the risk that every protein touches every other along some hypothetical pathway, we limited consideration to pathways with a total distance of less than 6; that is, an IEI gene would have to link to another IEI gene by less than six protein-protein interactions.

For IEI genes that did not return any interactions, we utilized the HGNChelper to find alternative gene symbols to query [11]. In sum, 334 IEI genes were analyzed for related proteins.

Filtering pLI and GDI

We filtered pLIs under 0.9 using the updated pLI index available at https://gnomad.broadinstitute.org/downloads [12]. Gene damage indices (GDI) were taken from the HGC GDI server [13, 14].

RNAseq data

We used immune cell RNA sequencing data from the Human Blood Atlas [15]. This provided expression data from 18 different immune cell types (available at https://www.proteinatlas.org/humanproteome/blood). Monocyte subsets, eosinophils, basophils, and neutrophils were grouped as myeloid cells. Plasmacytoid and myeloid dendritic cells were grouped together as DCs. NK, B and T

cell subtypes were grouped as such. Genes with an average of under 1 transcript per million in their group were filtered out of our group-specific IEI candidate gene lists.

Genes present in our list of candidate genes that were not found in the RNAseq expression data ($n = 14$) automatically were included in our filtered lists.

Code

All plotting was done in ggplot2 in R. Code is available at github.com/humzalikhan/omnipath_IEI.

Results and discussion

IEIs are genetic diseases, and every patient with a genetic disease deserves a specific genetic diagnosis whenever possible. The published IEI genes offer only a limited snapshot of all the human genes that underlie immune disorders, as is shown by the rapid growth in the identification of new genes, more than 30 per year at this point [1]. We employed two methods in parallel to evaluate genes and include them in a list of potential IEI-related genes. First, we assessed protein associations through the lenses of post-translational modification and transcriptional regulation. Second, we analyzed unannotated pathways of protein-protein interactions between all combinations of IEI genes (Fig. 1). Together, these approaches identified a list of 2,530 genes that we propose should be prioritized for consideration as IEI genes (Table S1).

To validate our algorithm, we repeatedly sampled the list of *known* IEI genes, taking a random subset of 80 %

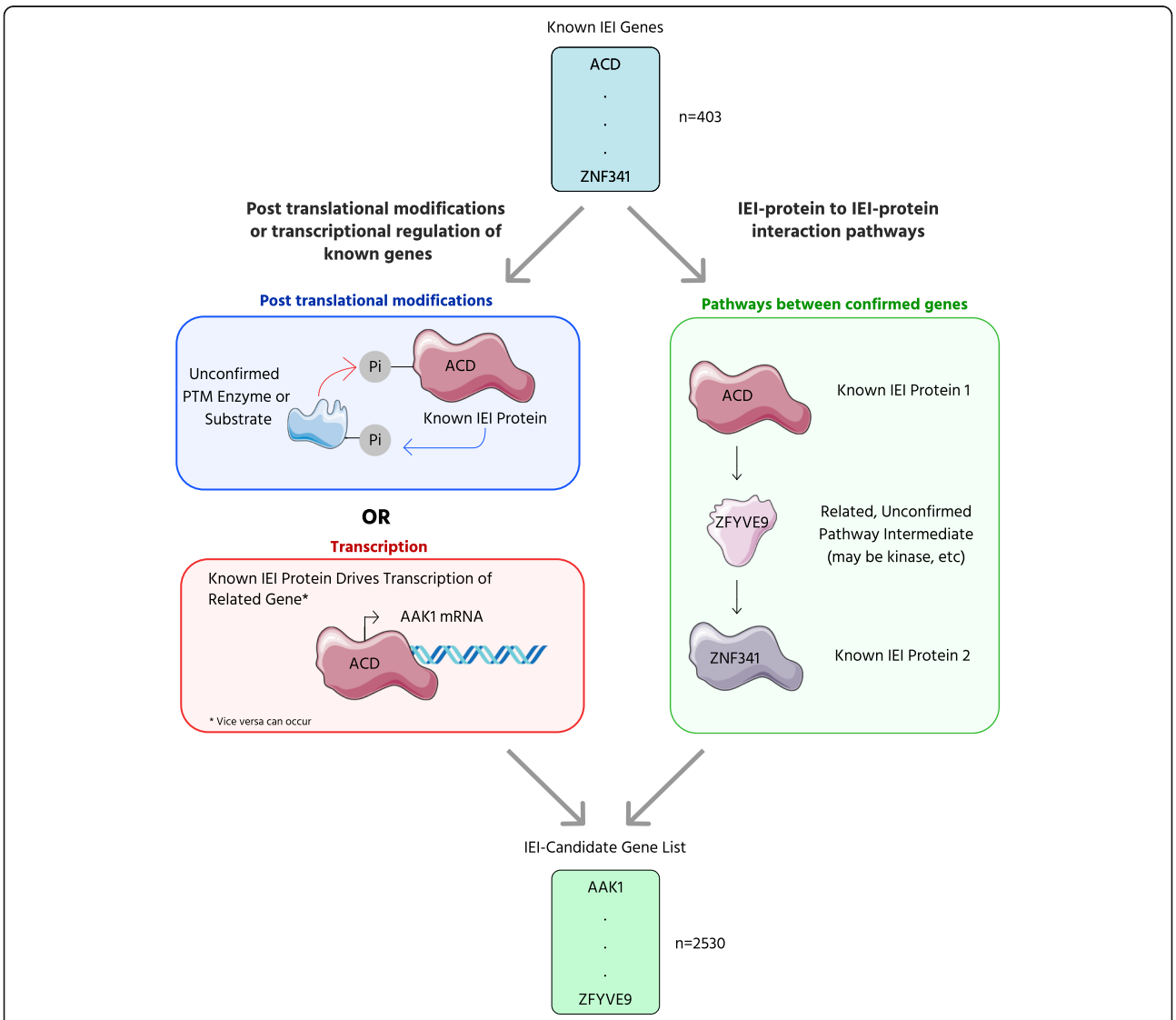

Fig. 1 Overview of our approach. We used the OmniPath set of databases to create a list of candidate IEI genes using two approaches: pathway 1 (left), which uses post-translational modification and transcription factor data, and pathway 2 (right), which uses protein-protein interaction pathways

of the genes ("training") and setting 20 % aside for "testing." We ran the training set of genes through our computational pipeline and assessed how much of the testing 20 % of genes could be rediscovered. We found that, on average, this approach recaptured ~ 50 % of the omitted IEI genes (that is, 10 % of the total set) (Fig. S1A). Additionally, known IEI genes have already been split into nine categories of immune function ("tables" in the IUIS paper [1]). Therefore, we pursued a cross-validation methodology where we iteratively left one whole category out and use the genes in the other categories to rebuild the genes in the left-out category (Fig. S1B). On average, about 40 % of a left-out category was rediscovered. There was large variance in this methodology, where most SCID and CID genes (IUIS table 1) were found, while not many complement genes (IUIS table 8) were. This finding may be due to categorical differences in interactions between pathways; for example, complement genes are known to be exclusive in their interactions [16]. Our validation approach gives us confidence that our methodology is useful in actually discovering IEI genes.

In our discovery pathway 1 approach, each gene was analyzed for its action as a transcription factor, its induction/repression due to a transcription factor, its activity as a post-translational modifier, or as a recipient post-translational modification. We first explore the distribution of known IEI genes by category (Fig. 2A). We then analyzed the relationship our candidate genes to known IEI categories in discovery pathway 1, finding that our list of pathway 1 genes associated more with IUIS table 1 genes that cause cellular and humoral defects, and almost none with genes that cause complement deficiencies (Fig. 2B). Of the relationships described in pathway 1, we found that IEI gene products are post-translationally modified more often than they post-translationally modify other proteins (Fig. 2C-D). In particular, they are likely to be phosphorylated along a signaling pathway that results in cellular response such as transcription. Some IEI genes also encode kinases/phosphatases themselves (Fig. S2A). While IEI gene products also ubiquitinate and cleave other proteins, it was found that modifications to IEI genes are more diverse than the effects of known IEI genes (Fig. S2B). Associated genes may (de)methylate, neddylate, (de)acetylate, and sumoylate IEI genes. Some of these interactions have been described in literature; for example, SUMO modification of STAT1 has been documented [17], but the sumoylation enzymes are not known IEI genes. As the discovery of the *TNFAIP3*, *OTULIN*, *HOIL-1*, and *HOIP*

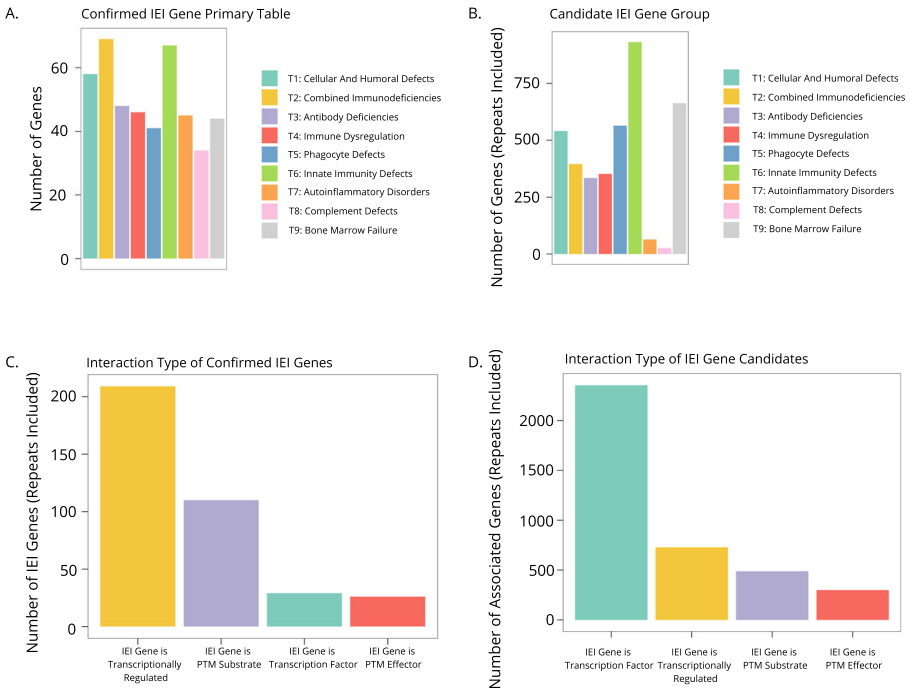

Fig. 2 Discovery Pathway 1 reveals IEI candidates by analyzing transcriptional regulation and post-translational modifications of known IEI genes. **A** Number of confirmed IEI genes in respective IUIS tables as defined by the phenotype that their variants cause. **B** Number of candidate IEI genes in tables classified by their interactions with known IEI genes. For **A** and **B**, only primary tables were utilized. **C** Number of known IEI genes and their respective modifications. **D** Number of candidate IEI genes that transcriptionally regulate IEI genes, are regulated by IEI genes, and PTMs to/from IEI genes

have shown, PTMs are important to regulate proper immunity activity and quiescence; conceivably, a deleterious variant in the genes that modify or are modified by known IEI proteins could alter immune function and cause a novel IEI.

In discovery pathway 1, we also analyzed transcriptional regulation related to IEI genes. Notably, we found that IEI proteins more often are involved as effectors of this regulation as opposed to recipients of it. The STAT family is a set of transcription factors of many downstream pathways relevant to immune function; defects in STAT action are known to cause susceptibility to infection, autoimmunity, and immune dysregulation [18]. Defects in the downstream products of IEI genes that act as transcription factors (TFs), such as STAT4/6, and the upstream TFs that induce transcription of IEI genes may alter immune function and be found to cause immunodeficiency.

In discovery pathway 2, we constructed a set of putative "pathways," the starts and ends of which are every known IEI-gene. Along the pathway lie validated protein-protein interactions, validated by literature. For example, two proteins that lie along the pathways between IFNAR1 and STAT3 include TYK2 or JAK1. Alternatively, longer, less biologically probable routes exist, such as WRAP53 -> TP53 -> RPS6KA4 -> MAPK14 -> CSNK2A2 -> DKC1. In discovery pathway 2, we limited the gene-gene pathway distance to five to reduce the biological irrelevance of the putative pathways, knowing that six or seven degrees of separation link virtually every protein with every other. Our analyses of protein-protein interaction pathways from known IEI gene to known IEI gene revealed many previously known IEI genes (Fig. 3A), a result supports that as-yet undiscovered IEI genes may lie along these same routes. We found many new genes using this approach (Fig. 3B). We again found complement-related genes underrepresented, similar to the functional annotations in discovery pathway 1. This result suggests that not many pathways exist ending in a complement protein, that the complement system is perhaps independently regulated and is less intertwined with other immune defenses [16].

To assess whether a particular gene might play a role in multiple pathways, and thus potentially be important to multiple mechanisms of disease, we analyzed the frequency of genes along the protein-protein association pathways. However, many known IEI genes occurred at high frequency within these pathways, while others only occurred one or a few times (Fig. 3C). For example, CTLA-4, a well-known inhibitory receptor whose deficiency causes antibody deficiency and autoimmunity, was only found once as a pathway intermediate in between two IEI genes. Therefore, we decided against culling low-frequency genes from our list of candidates.

Our list of 2,530 related genes is smaller than the list of over 3,000 candidate genes from the Itan group [3]. Furthermore, since we utilized a compendium of many databases, our associations are likely to be biologically relevant and supported by literature. About half of the genes we identified appeared in the Itan group's list, and the other half were novel. Furthermore, to assess how "close" our gene list is to the list of known IEI genes, we performed an analysis using the HGC. We found that our gene list has a median distance of 10.4 between the candidate genes and known IEI genes, which can be described as a "small" to "small-medium" distance (Fig. S3). This result indicates that our candidate genes are biologically close to the known set of IEI genes. 49.3 % of genes were found exclusively in discovery pathway 1, 29.7 % were found exclusively in pathway 2, and 21 % were found via both. As expected, genes found in discovery pathway 1 included kinases, TFs, and other PTM effectors. Pathway 2 recapitulated many of these genes, but also included genes that interact generally with known IEI genes. Genes appearing in both lists should not necessarily be prioritized, as this finding may simply

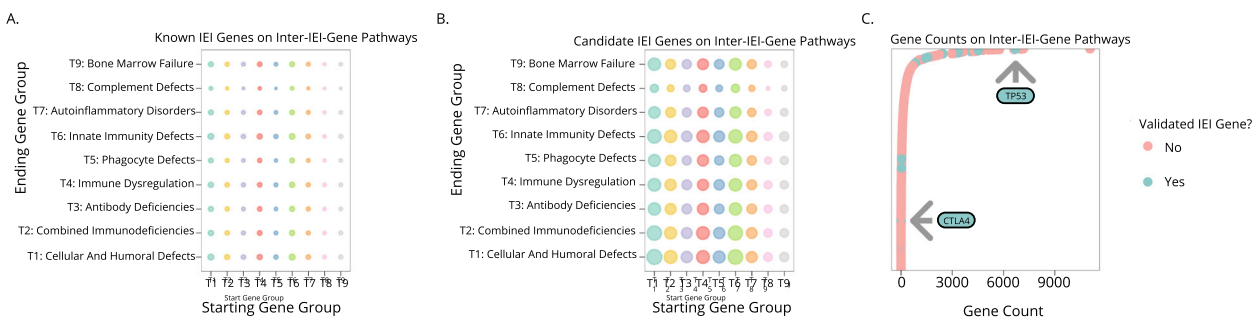

Fig. 3 Discovery Pathway 2 uses protein interaction data to describe pathways between genes. **A** Bubble plot showing the relative numbers of *known* IEI genes that lie on protein interaction pathways between known IEI genes and known IEI genes. **B** Bubble plot showing the relative numbers of *candidate* IEI genes that lie on protein interaction pathways between known IEI genes and known IEI genes. **C** Number of times that a particular gene appears along different pathways between known IEI genes and known IEI genes. Two previously known IEI-causative genes are indicated

be because the same gene function is present in both datasets that were analyzed. The presence of a gene in either pathway is indicated in our list.

Ranking genes by their probability of loss intolerance (pLI) has been suggested as a way to predict the importance of unknown genes to human disease, especially for genetically dominant conditions. This metric is often used in the clinic to filter genes that bear potentially deleterious mutations. High pLI values imply purifying selection in the population when genomic loss-of-function variants (e.g., frameshift, truncation) appear. pLI values are calculated empirically from large databases of healthy individuals by comparing observed variants in the population to the number of variants expected to arise by chance [19]. These databases and pLI values are of course subject to change as more and diverse populations are sampled. To filter our putative list of IEI genes, we considered pLI scores for all known and putative IEI genes. Notably, we found that known IEI genes did not have unilaterally high pLI (Fig. S4A). Disaggregating by inheritance type, high pLI values (> 0.9) were shown to be associated with autosomal dominant inheritance and well associated with X-linked dominant conditions. The pLI for known IEI genes is hardly constrained for autosomal recessive disorders (Fig. S4A). Thus, we have provided a shorter list of 810 high-pLI genes (including genes without available pLIs) for consideration of X-linked dominant or autosomal dominant conditions (Table S2). This subset of genes may be less important for recessive clinical traits, where the entire list of putative genes should be considered.

Since not all IEI genes have high pLI scores, we decided to explore the previously described Gene Damage Index (GDI), which uses the presence of population-level mutational damage to filter out genes with low potential for pathogenicity [12]. That is, higher GDI indicates that healthy humans can tolerate some damage in those genes (i.e., opposite to pLI). However, we found that many known IEI genes also had variable scoring of predicted GDI (Fig. S4B). Predictions of damage from known IEI genes using the GDI results in mostly a "medium" characterization (Fig. S4C). Therefore, qualitative and quantitative GDI predictions have limited utility for filtering putative IEI genes. We found that GDI and pLI do not negatively correlate with each other in known IEI genes (Fig. S4D). Taken together, our results show that filtering on only high pLI or low GDI genes would serve only a limited role and may unnecessarily constrain lists of putative genes.

Known IEI genes are generally expressed in immune cells; we used RNA sequencing data from the Human Protein Atlas to liberally filter for variants expressed in any immune cell type. The Human Protein Atlas database includes RNA sequencing data of 18 immune cell types from healthy donors which comprised five general groups: B cells, dendritic cells, myeloid cells (basophils, neutrophils, eosinophils, monocytes), NK cells, and T cells [15]. For initial inquiry, we analyzed the expression of all genes in each cell group and found most genes not expressed at all in immune cell types (Fig. S5A). However, when only IEI genes were plotted, the distribution was more right-skewed, indicating the expected trend that IEI genes are transcriptionally well expressed in immune cells (Fig. S5B). Expectedly, among the IEI genes that have zero RNA transcripts among immune cells are *AIRE* and complement genes. Therefore, we decided to filter a list for expression above 0 TPM in any immune cell type (Table S3).

Furthermore, since many IEIs manifest with cell-type specific defects, we decided to use RNA sequencing data to also create IEI-related gene lists based on cell-type specific gene expression. Such a list may offer utility when investigators are focused on disordered with focused immunophenotypes. For example, one might begin a search for new SCID genes where only T cells are affected by looking at transcripts that are expressed in T cells or T cell precursors.

To determine a cutoff for what is considered "expressed," we plotted gene expression of IUIS table 1 genes, which affect humoral and cellular immunity, and IUIS table 3 genes, which predominantly affect antibody production, and disaggregated by cell type. The former would be more likely to be expressed in all lymphocytes and the latter in only B cells (Fig. 4A-B). As disorders in IUIS table 1 can manifest due to defects in either T or B cells, we found that many of the genes that were expressed at a low level in T cells were expressed in B cells and vice versa. For example, among the low-expressed genes in B cells, for example, was *CD3D*. *CD3D* is a major component of the T-cell receptor complex and was highly expressed in T cells but not in B cells. CD3D deficiency causes a T-/B + SCID (an IUIS table 1 condition) and is thought to be irrelevant in B cell development, which is supported by its negligible gene expression. Thus, if one were considering a list of IEI genes to be used for patients with B-cell focused defects, we would filter out such T-cell-specific genes as likely irrelevant.

We set our cutoff point for expression as greater than 1 transcript per million (TPM). We picked this limit based on the transcriptional expression of known IEI genes across cell types (Fig. 4A-B). There was only one known IEI gene that causes a primary antibody deficiency (PAD) expressed below 1 TPM in B cells, the gene called *SLC39A7*, which codes for ZIP7. Ostensibly, this finding was unexpected since PADs reflect defects in B cell numbers or function, but an obvious explanation would be that this gene is expressed only early in

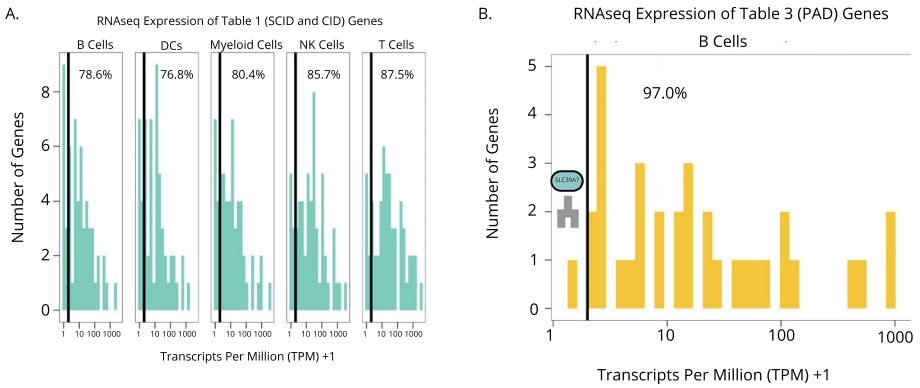

Fig. 4 RNA sequencing data provide a method to filter for relevant IEI genes depending on patient phenotype. **A** Expression levels of SCID and CID-causing genes in the cell types found in the Human Blood Atlas RNAseq expression matrix. **B** Expression levels of antibody deficiency-related genes in B cells. Transformations of TPM + 1 done to visualize 0 values. Percentage of genes expressed above 1 TPM (at 2 when transformed to TPM + 1) cutoff presented in each plot

B-cell development and not in peripheral B cells (Fig. 4B). Indeed, ZIP7 is an essential zinc transporter without which B-cell development is abrogated in the transition from pro- to pre-B cell. Thus, we anticipate that filtering gene lists by expression levels would preserve the majority of interesting genes. The genes that were erroneously filtered out in cell-type specific lists would still be present in the broader, unfiltered list. Our cell-type specific lists for the cell type groups in the Human Blood Atlas are found in the supplement (Tables S4, S5, S6, S7 and S8).

A limitation of our work is that some genes are not described in the protein-interaction databases queried. For example, the complement protein C2 with its known interaction with CD19 was not represented in the protein-protein databases, and this interaction was thus not recognized in our searches using the OmniPath databases. Furthermore, the Human Blood Atlas database holds RNAseq data on major categories of immune cells; gene expression data from clinically relevant subsets such as plasmablasts, effector memory T cells, and others would be valuable. As access to open-source datasets increases and as functional evidence for protein function is released, we will refresh our lists.

Additionally, our lists do not address the problem of isolating non-redundant genes as cause for immunodeficiencies. Better databases of immunological networks and tissue annotation may address this problem. However, given the brevity of our lists, we are less likely to include redundant genes in our datasets. By also further paring our lists for cell-type expression, we increase the likelihood that genes from redundant, non-specific immune pathways are culled. Lastly, there are IEI genes with no or only nominal expression in circulating immune cells (e.g., complement, some cytokines, developmental genes). In these cases, our unfiltered list is likely to be more useful than our filtered lists.

In summary, by using a combination of both 1:1 annotated protein interactions and larger, un-annotated protein-interaction pathways, our approach allows for both a global and local view of proteins that may be relevant to query in future immunodeficiency studies. Notably, we found that high pLI or GDI were not particularly good criteria for determining the pathogenicity of a putatively novel IEI gene, especially those with recessive patterns of inheritance. Our work further advances on previous studies by merging transcriptional expression data with our list of IEI candidate genes derived from protein interactions to ensure that queries based on clinical presentations (i.e., T-cell lymphopenia) or diagnostic hypotheses can be made.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-07909-3.

Additional file 1: Table S1. IEI Candidate Genes without Filter.

Additional file 2: Table S2. IEI Candidate Genes with high (>.9) pLI Filter.

Additional file 3: Table S3. IEI Candidate Genes with Average in all Immune Cells > 0 TPM Filter.

Additional file 4: Table S4. IEI Candidate Genes with Average in B Cells > 1 TPM Filter.

Additional file 5: Table S5. IEI Candidate Genes with Average in DCs > 1 TPM Filter.

Additional file 6: Table S6. IEI Candidate Genes with Average in Myeloid Cells > 1 TPM Filter.

Additional file 7: Table S7. IEI Candidate Genes with Average in NK Cells > 1 TPM Filter.

Additional file 8: Table S8. IEI Candidate Genes with Average in T Cells > 1 TPM Filter.

Additional file 9: Figure S1. Algorithm verification using out-of-sample testing. (A) Percentage of validation set genes rediscovered by algorithm. (B) Percentage of genes from left-out category rediscovered by algorithm. **Figure S2.** Post-translational modifications induced by and received by IEI gene products. (A) The number of targets or substrates for each IEI gene when they are either a kinase, phosphatase, protease, or ubiquitinase. (B) The number of proteins that are known to target IEI proteins for post-translational modification by various mechanisms. **Figure S3.** Human Gene Connectome (HGC) Distance between known and candidate IEI genes. **Figure S4.** Known IEI-causative genes have varying levels of predicted mutational harm. (A) Confirmed IEI gene pLI disaggregated by inheritance type (AD: autosomal dominant, AR: autosomal recessive). (B) Confirmed IEI gene Gene Damage Indices (GDIs) disaggregated by inheritance type. (C) Damage prediction of confirmed IEI genes as classified by the GDI Server. (D) Known IEI gene pLIs plotted against GDIs and disaggregated by inheritance type. **Figure S5.** Known IEI transcripts skew to higher expression in immune cell types. (A) RNAseq expression of all Human Blood Atlas-recorded genes disaggregated by cell types present. (B) RNAseq expression of all IEI genes disaggregated by cell type. Line drawn at 1 TPM (at 2 when transformed to TPM+1). Percentage above 1 TPM (at 2 when transformed to TPM+1) cutoff presented.

Acknowledgements
not applicable.

Authors' contributions
Conceptualization, Analysis, Writing: H.A.K. and M.J.B.; Funding Acquisition, Project administration, Supervision: M.J.B. The author(s) read and approved the final manuscript.

Funding
This work was supported by the NIH/NIAID R01 AI153827, by the E. Richard Stiehm Endowment, and the Jeffrey Modell Foundation.

Availability of data and materials
All data for this study are included in the manuscript, supplementary files, and our github site: https://github.com/humzalikhan/omnipath_IEI.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no conflicts of interest.

Received: 24 February 2021 Accepted: 22 July 2021
Published online: 15 August 2021

References

1. Tangye SG, Al-Herz W, Bousfiha A, Chatila T, Cunningham-Rundles C, Etzioni A, et al. Human Inborn Errors of Immunity: 2019 Update on the Classification from the International Union of Immunological Societies Expert Committee. J Clin Immunol. 2020;40:24–64.
2. Casanova J-L, Conley ME, Seligman SJ, Abel L, Notarangelo LD. Guidelines for genetic studies in single patients: lessons from primary immunodeficiencies. J Exp Med. 2014;211:2137–49.
3. Itan Y, Casanova J. Novel Primary Immunodeficiency Candidate Genes Predicted by the Human Gene Connectome. Front Immunol. 2015;6:1–8. https://doi.org/10.3389/fimmu.2015.00142.
4. Itan Y, Mazel M, Mazel B, Abhyankar A, Nitschke P, Quintana-Murci L, et al. HGCS: An online tool for prioritizing disease-causing gene variants by biological distance. BMC Genomics. 2014;15:1-8.
5. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Res. 2017;45:D362-8. https://doi.org/10.1093/nar/gkw937.
6. Zhang P, Itan Y. Biological network approaches and applications in rare disease studies. Genes (Basel). 2019;10:1–15.
7. Türei D, Korcsmáros T, Saez-Rodriguez J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nat Methods. 2016;13:966–7. https://doi.org/10.1038/nmeth.4077.
8. Valdeolivas A, Turei D, Gabor A. Omnipathr: Import omnipath network. Bioconductor Package. 2019.
9. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. 2019;29:1363–75. https://doi.org/10.1101/gr.240663.118.
10. Licata L, Lo Surdo P, Iannuccelli M, Palma A, Micarelli E, Perfetto L, et al. SIGNOR 2.0, the SIGnaling Network Open Resource 2.0: 2019 update. Nucleic Acids Res. 2020;48:D504–10.
11. Waldron L, Riester M. HGNChelper: Identify and Correct Invalid HGNC Human Gene Symbols and MGI Mouse Gene Symbols. Bioconductor Package. 2019.
12. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–43. https://doi.org/10.1038/s41586-020-2308-7.
13. Human Gene Damage Index server. https://lab.rockefeller.edu/casanova/GDI.
14. Itan Y, Shang L, Boisson B, Patin E, Bolze A, Moncada-Vélez M, et al. The human gene damage index as a gene-level approach to prioritizing exome variants. Proc Natl Acad Sci. 2015;112:13615–20. https://doi.org/10.1073/pnas.1518646112.
15. Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, et al. A subcellular map of the human proteome. Science (80-). 2017;356:eaal3321. https://doi.org/10.1126/science.aal3321.
16. Arkwright PD. Primary Immunodeficiencies of Complement. In: Primary and Secondary Immunodeficiency. Springer International Publishing; 2021. p. 313–30.
17. Rogers RS, Horvath CM, Matunis MJ. SUMO Modification of STAT1 and Its Role in PIAS-mediated Inhibition of Gene Activation. J Biol Chem. 2003;278:30091–7. https://doi.org/10.1074/jbc.M301344200.
18. Milner JD, Vogel TP, Forbes L, Ma CA, Stray-Pedersen A, Niemela JE, et al. Early-onset lymphoproliferation and autoimmunity caused by germline STAT3 gain-of-function mutations. Blood. 2015;125:591–9. https://doi.org/10.1182/blood-2014-09-602763.
19. Fuller ZL, Berg JJ, Mostafavi H, Sella G, Przeworski M. Measuring intolerance to mutation in human genetics. Nature Genetics. 2019;51:772-6.

Publisher's Note