PanelAppRex aggregates disease gene panels and facilitates sophisticated search

Quant Group[1], Simon Boutry[2], Ali Saadat[2], Jacques Fellay[2], Sinisa Savic[3], Luregn J. Schlapbach[4], and Dylan Lawless * [4]

[4]Department of Intensive Care and Neonatology, University Children's Hospital Zurich, University of Zurich, Zurich, Switzerland.
[2]Global Health Institute, School of Life Sciences, École Polytechnique Fédérale de Lausanne, Switzerland.
[3]Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds, Leeds, UK.
[1]The quantitative omic epidemiology group.

June 25, 2025

Word count: 2560

*Addresses for correspondence: Dylan.Lawless@kispi.uzh.ch

Abstract

Gene panel data is critical for variant interpretation and genomic diagnostics, yet existing resources remain fragmented, inconsistently annotated, and difficult to interrogate programmatically. We present PanelAppRex, a harmonised resource and interactive search platform integrating over 58,000 curated gene-disease panel associations, including NHS-approved diagnostic panels. It supports natural language-style queries by gene, phenotype, disease group, and mode of inheritance (MOI), returning machine-readable results in multiple export formats. The system achieved 100% top-panel accuracy in benchmarked case studies of genetic immune disease diagnosis. The core dataset includes standardised gene identifiers, disease annotations, MOI, and supporting literature, enabling integration into bioinformatic pipelines. Beyond interactive search, PanelAppRex provides a structured foundation for modelling genome-wide diagnostic priors and evidence-aware variant interpretation.

Availability: The platform data and source code is openly available at [ZENODO LINK] and https://github.com/DylanLawless/PanelAppRex. PanelAppRex is available under the MIT licence. The dataset is maintained for a minimum of two years following publication.

Acronyms

CSV comma-separated values . . . 6
GE Genomics England . . . 3
HGNC Human genome organisation Gene Nomenclature Committee . . . 4
iei Inborn Errors of Immunity . . . 6
iem Inborn Errors of Metabolism . . . 6
HTML HyperText Markup Language . . . 4
JACI Journal of Allergy and Clinical Immunology . . . 5
MOI Mode of Inheritance . . . 4
OMIM Online Mendelian Inheritance in Man . . . 4
PDF Portable Document Format . . . 6
PID Primary immunodeficiency . . . 5
RAG Retrieval-augmented generation . . . 3

1 Introduction

Disease-gene panels are widely used in both clinical and research settings to support the diagnosis and interpretation of genetic disorders. These panels provide structured lists of genes known to be associated with specific phenotypes or disease groups, helping clinicians prioritise candidate variants during genomic analysis. Sources like Genomics England (GE)'s PanelApp and PanelApp Australia host comprehensive, expert-curated panels that are actively maintained to reflect new genetic knowledge (1). For instance, these resources are integral to the NHS National Genomic Test Directory and the 100,000 Genomes Project (1). Despite their utility, these panel datasets are distributed across multiple platforms and formats, and are difficult to aggregate programmatically without data loss or inconsistency. Manual selection, interpretation, and cross-referencing of gene panels remain labour intensive, especially when integrating with other variant-level annotations from genomic resources.

Accurate causal variant interpretation requires consistent integration of structured annotations from key resources such as GE's PanelApp, ClinVar, UniProt, and Ensembl (1–4). To address this, we developed PanelAppRex, a tool that automates panel aggregation and prepares a machine-readable dataset with complete coverage of core gene-level fields, including gene identifiers, inheritance modes, disease terms, and supporting evidence. The dataset is suitable for integration into AI-based workflows and downstream variant interpretation pipelines. In parallel, PanelAppRex provides a natural language-style search interface that streamlines discovery by supporting queries based on gene names, phenotypes, disease groups, and other key attributes. This functionality is enhanced by evidence-based Retrieval-augmented generation (RAG), allowing users to explore curated gene panel information efficiently.

2 Materials and methods

2.1 Data

The PanelAppRex core model contained 58,592 entries consisting of 52 sets of annotations, including the gene name, disease-gene panel ID, diseases-related features, confidence measurements. Disease interactions were sourced from GE's PanelApp (1). Data from gnomAD v4 comprised 807,162 individuals, including 730,947 exomes and 76,215 genomes (6). This dataset provided 786,500,648 single nucleotide variants and 122,583,462 InDels, with variant type counts of 9,643,254 synonymous, 16,412,219 missense, 726,924 nonsense, 1,186,588 frameshift and 542,514 canonical splice site variants. ClinVar data were obtained from the variant summary dataset available from the NCBI FTP site, and included 6,845,091 entries, which were processed into 91,319 gene classification groups and a total of 38,983 gene classifications (2). Data from Ensembl was sourced for validation of identifiers such as Online Mendelian Inheritance in Man (OMIM) gene IDs and Human genome organisation Gene Nomenclature Committee (HGNC) symbols (4).

2.2 Implementation

PanelAppRex was implemented in R and integrated data from the sources listed above. It performed credentialed access to APIs to retrieve all approved panels, merging them into two formats: a simplified version (Panel ID, Gene) and a complex version including metadata such as confidence level, Mode of Inheritance (MOI), and disease information, and several metadata summary statistics. In addition, the tool incorporated a search module to execute complex user queries. The search functionality supports queries by gene names, phenotypes, disease names, disease groups, panel names, genomic locations and other identifiers. RAG was used to improve the natural queries in hidden states based on evidence about the disease and gene function by supplementing the data with additional sources including ClinVar, UniProt, etc.

2.3 Usage

We provide both a pre-compiled user interface and an analysis-ready dataset. On a desktop browser, the user interface allows for queries via the integrated search bar in HyperText Markup Language (HTML), where a JavaScript function splits the query into individual terms and progressively filters all entries - retaining only those that match all active terms while ignoring unmatched ones. This enables users to perform complex, partial matching queries (e.g. "paediatric *RAG1* primary immunodeficiency skin disorder") to rapidly identify the panel most closely associated with their hypothesis on primary immunodeficiency and paediatric skin disorders. After querying and selecting a panel, the complete underlying data is returned.

The analysis-ready dataset contains the same dataset in multiple file formats. Our recommended strategy is a bioinformatic approach which can use the union of returned panels, or to apply a scoring system to rank panels in multi-group analyses. Bioinformatically, users can import the provided, ready-for-use, datasets in TSV or Rds formats. A typical use case might involve users merging with their own omic data based on gene or Ensembl ID. The following code snippet, available in `minimal_example.R`, demonstrates how to load the data in R:

```
# TSV format
path_data <- "../data"
core_path <- paste0(path_data, "/PanelAppData_combined_core")
minimal_path <- paste0(path_data, "/PanelAppData_combined_minimal")

df_core <- read.table(
    file= paste0(core_path, ".tsv"),
    sep = "\t", header = TRUE)

df_minimal <- read.table(
    file= paste0(minimal_path, ".tsv"),
    sep = "\t", header = TRUE)

# Rds format
rds_path <- paste0(path_data, "/PanelAppData_combined_Rds")
df_core <- readRDS(file= rds_path)
```

2.4 Validation for completeness in core data

To ensure the reliability and completeness of the core dataset, we systematically assessed whether key gene-level fields were present for each entry after merging the core dataset. Specifically, we checked for non-missing values in gene symbols, associated publications, disease panel names, MOI, and OMIM gene IDs. Where HGNC or Ensembl gene IDs were missing, we used programmatic queries to the Ensembl database via the '`biomaRt`' R package to recover the identifiers using the available HGNC symbol as input (8). This validation step was applied to the full integrated PanelAppRex dataset.

2.5 Benchmarking for manual queries

To mimic a clinician who is trying to diagnose a new disease such as Primary immunodeficiency (PID), we systematically selected genetic diagnosis case studies from the current online catalogue from the Journal of Allergy and Clinical Immunology (JACI), using the first five results (9–13). Although the method presented here is generalisable to genetic diagnosis across disease areas, we chose to validate it using PID case studies, as this is the primary focus of our own research.

The clinical background from each of these studies was used to construct keyword queries from patient features, simulating a naïve starting position for a clinician. We then tested whether our PanelAppRex tool could successfully retrieve panels that included the final causal gene reported in each case study. The sources, queries, and results are shown in **Table 1**.

3 Results

3.1 Core dataset

PanelAppRex successfully aggregated data, currently from 451 panels, and several genomics databases to offer a user-friendly search functionality (**Figure 1**). Users can retrieve results filtered by gene names, phenotypes, disease groups and other criteria. The system returned a table view with panel details and provided options for exporting results in comma-separated values (CSV), Excel, or Portable Document Format (PDF) formats. Bioinformatic uses may include generating virtual panels, constructing prior odds, or supporting formal reporting in qualifying variant protocols. Summary statistics are reported in **Figure S1**. The majority of genes were cited with high confidence. Some genes were present in more panels that others; for instance, RAG1 was present in 7 panels. The majority of panels had fewer than 1000 genes; a examples we annotate panel ID 398 with 572 genes of PID which are well established as consensus in the community, also called Inborn Errors of Immunity (iei) and a second example of panel ID 1220 with 1675 genes which are associated with unexplained death in infancy and sudden unexplained death in childhood and also called Inborn Errors of Metabolism (iem).

3.2 Validation shows completeness in the core dataset

The raw dataset exhibited near-complete coverage in most fields after merging the core dataset. Specifically, 99.9915% of gene entries had HGNC symbols and 99.2315% had Ensembl gene IDs. All entries included at least one publication, a disease panel name, a MOI, and an OMIM gene ID (100% each). After recovery of missing identifiers, completeness reached 100% across all these core fields. This validated dataset ensures that users can confidently build on a consistent foundation, using standardised and stable identifiers to link or enrich the core data with external resources (**Figure 2**).

3.3 Benchmarking for manual queries confirms accurate retrieval of causal disease genes

To evaluate the practical utility of our system on a desktop browser, we applied the benchmarking approach described in methods using five published case studies of PID

(9–13). The tool's query process retrieved gene panels using natural language-style input. For example, in the first case study on Hereditary Angioedema, the authors reported suspecting that the condition was linked to "*SERPING1*, Factor XII, and edema" which we used as the query terms. Although the true causal gene term *F12* (using its official HGNC name) was not mentioned, the inclusion of the alias name and/or related disease terms enabled the system to correctly retrieve the panel containing *F12* based on the hidden search knowledge.

While an expert might readily recognise "Factor XII" as an alias name for *F12*, this correct query result demonstrates that the system can make such connections automatically, supporting users with varying levels of genetic expertise.

In our evaluation, PanelAppRex returned panels in which the causal gene was present in 93% of all returned panels, with the subjective user-selected "best panel" achieving a perfect accuracy of 100% (**Figure 1**). These metrics were derived by comparing the causal gene identified from the case study abstracts to the panels returned by our query, and by applying a subjective relevance measure to mimic intuitive panel selection - acknowledging that certain panels, such as the established PID gene panel, are inherently more reliable than broader, less specific panels. Overall, the results confirm that our approach accurately identifies the most relevant panels and effectively supports clinical decision-making in complex diagnostic scenarios.

3.4 Structured inheritance annotations reveal overlapping gene-MOI relationships

To assess the distribution of inheritance annotations in, we summarised all unique gene-MOI combinations across 58,592 curated panel entries as displayed in **Figure S2**. Among 6,280 distinct genes, a total of 9,237 unique gene-MOI pairs were identified, reflecting instances where a gene appears with multiple inheritance modes across different panels. The most common annotations were **ar!** (**ar!**) and **ar!**, consistent with their prevalence in monogenic disease panels. These structured inheritance annotations underpin key applications enabled by PanelAppRex, including MOI-filtered search and model-based prior estimation.

Table 1: **Summary of case study queries and PanelAppRex results.** *Case study 5 had five individual cases and patient information was significantly longer that other studies. It was therefore converted to keyword query automatically by the OpenAI model o3-mini to remove subjective bias and align with the other queries. "Subjective best panels" are those reasonably preferable to the clinical query and unlikely to be excluded by users. In the benchmark scenario, broad or less relevant panels (e.g. "COVID-19 research") might be deprioritised in favour of more clinically aligned options such as "Fetal anomalies", or "Paediatric disorders". Summarised in **Figure 1**.

Case study (Ref)	**Source title**	**Query**	**Causal gene**	**Result panels**	**Subjective best panels**	**Panels with causal gene**	**Subjective relevance ratio**	**Causal gene in all results**	**Result ID, panelName, geneCount**
1 (9)	Genetic Analysis As a Practical Tool to Diagnose Hereditary Angioedema with Normal C1 Inhibitor: A Case Report	SERPING1 Factor XII edema	F12	3	1	3	0.3	1	64 COVID-19 research 695; 192 Primary immunodeficiency or monogenic inflammatory bowel disease 572; 311 Research panel - Severe Paediatric Disorders 2691
2 (10)	Severe dermatitis, multiple allergies, and metabolic wasting syndrome caused by a novel mutation in the N-terminal plakin domain of desmoplakin	SAM syndrome DSG1 dermatitis metabolic wasting	DSP	3	3	3	1	1	205 Fetal anomalies 2185; 210 DDG2P 2422; 211 Paediatric disorders 3903
3 (11)	Hematopoietic stem cell transplantation in a patient with proteasome-associated autoinflammatory syndrome (PRAAS)	resistant cutaneous vasculitis SH2D1A	PSMB4	3	1	2	0.3	0.7	64 COVID-19 research 695; 192 Primary immunodeficiency or monogenic inflammatory bowel disease 572; 311 Research panel - Severe Paediatric Disorders 2691
4 (12)	Autoimmune lymphoproliferative syndrome caused by a homozygous null FAS ligand (FASLG) mutation	Autoimmune lymphoproliferative syndrome ALPS lymphoproliferation hypergammaglobulinemia autoimmune cytopenia	FASLG	2	1	2	0.5	1	64 COVID-19 research 695; 192 Primary immunodeficiency or monogenic inflammatory bowel disease 572
5* (13)	Fatal combined immunodeficiency associated with heterozygous mutation in STAT1	primary immunodeficiency recurrent pneumonia chronic diarrhea oral thrush bronchiectasis lymphadenopathy hepatosplenomegaly autoimmune hepatitis Addison	STAT1	1	1	1	1	1	192 Primary immunodeficiency or monogenic inflammatory bowel disease 572

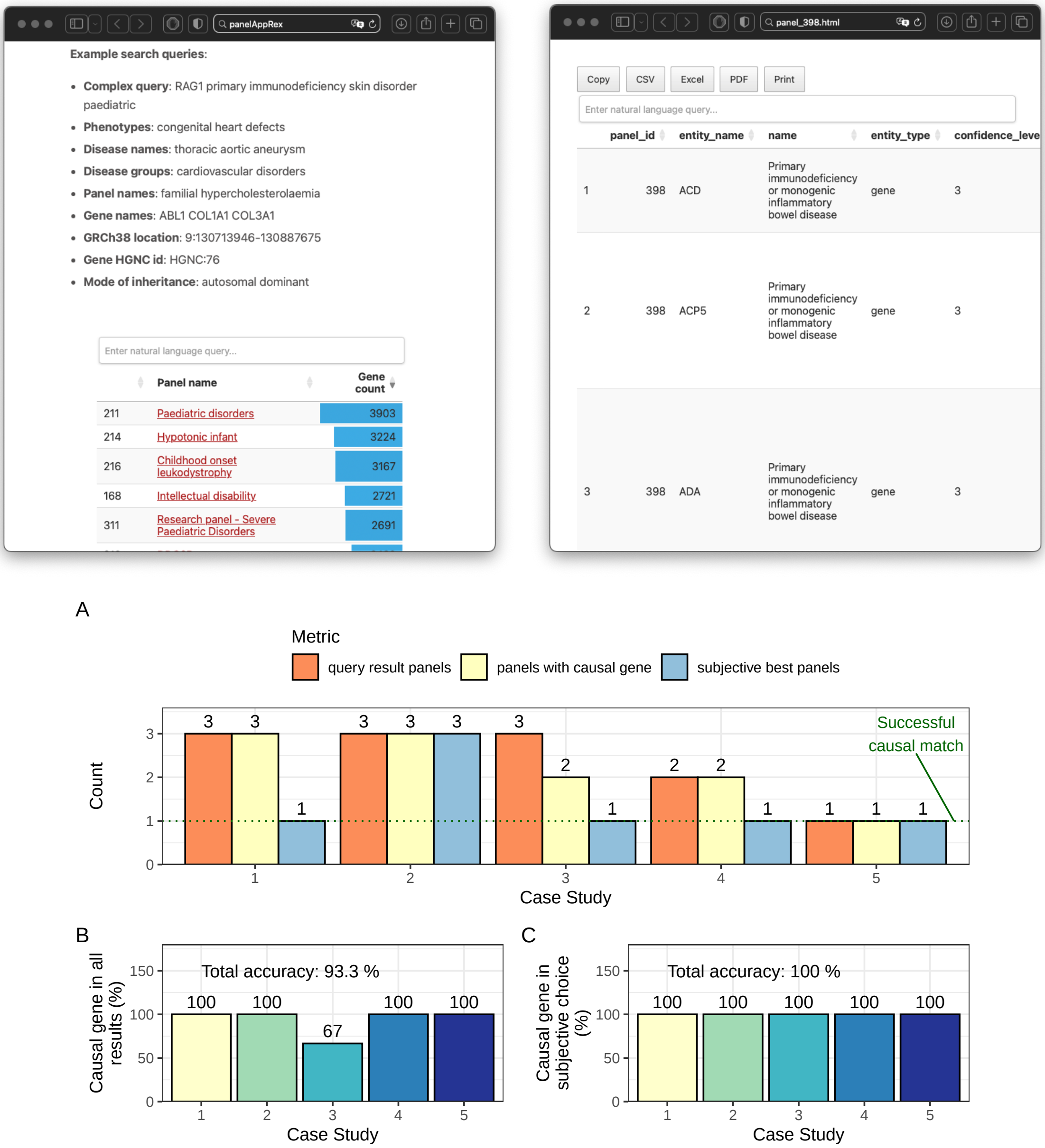

Figure 1: **PanelAppRex interface displaying search results for a complex query.** Top: screenshots showing the search interface, with the left panel displaying the full database before filtering and the right panel showing detailed results for a selected panel. Bottom: benchmark metrics are presented as follows: (A) for each of the 5 case studies, the total number of panels returned, the number of panels that included the causal gene, and the single best panel selected (always 1 by default); (B) the percentage of all returned panels that included the true causal gene; (C) the percentage of the single best panels that contained the true causal gene.

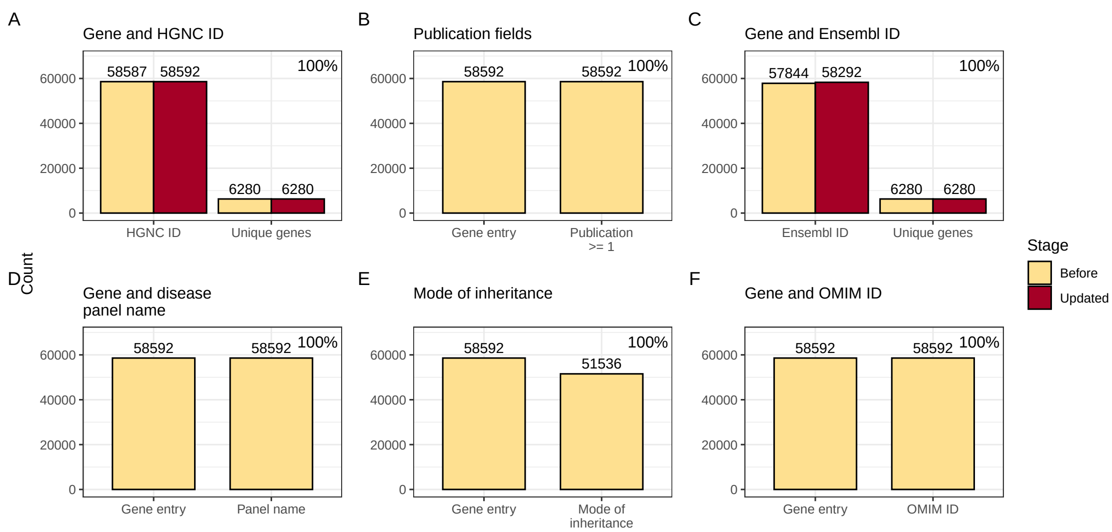

Figure 2: **Validation and recovery of core annotation fields in the PanelAppRex dataset.** (A) Unique genes and HGNC IDs before and after retrieving missing HGNC entries via Ensembl. (B) Availability of publication annotations for genes. (C) Unique genes and Ensembl gene IDs before and after biomart-based recovery of missing Ensembl IDs. (D) Gene entries with associated disease panel names. (E) Gene entries with annotated MOI. (F) Gene entries linked to an OMIM gene ID. Each bar shows the count of entries with non-missing values for the respective field. Updated fields (HGNC and Ensembl ID) reflect values recovered via external lookup using HGNC symbols. Percentage shows the complete recovery of coverage across features.

4 Discussion and conclusion

PanelAppRex provides an accessible and validated platform for querying, aggregating, and exporting curated disease-gene panel data. Designed for both clinical and research use, it simplifies study planning and variant interpretation through a user-friendly interface, while also offering a harmonised dataset for programmatic workflows. The core model integrates over 58,000 panel-gene associations, with full coverage of key fields including HGNC symbols, Ensembl IDs, OMIM annotations, MOI, and supporting publications.

Beyond immediate usability, PanelAppRex lays a foundation for genome-wide statistical modelling of disease risk. With annotations for gene-disease relationships, the dataset can support rigorous estimation of the prior probability of observing variant classifications (e.g. benign, pathogenic) under different MOI. This helps address a persistent challenge in clinical genetics: the lack of principled priors for variant interpretation that account not only for known pathogenic variants (true positives), but also for unobserved variants and negative evidence (false negative causal pathogenic variant and true negative absence of causal variants) (14; 15). By integrating high-resolution allele frequencies from gnomAD (6), curated classifications from ClinVar (2), and structured gene-disease associations (1), PanelAppRex can enable the derivation of calibrated genome-wide priors for Bayesian models of genetic risk.

These capabilities go beyond panel lookup to support evidence-aware diagnostics. Probabilistic models incorporating both observed and unobserved variation can quantify uncertainty, resolve ambiguous findings, and refine variant prioritisation, scaling naturally to support complex, multi-gene disorders. The dataset also provides a substrate for AI-driven approaches to variant interpretation, including probabilistic inference, reinforcement learning, and deep annotation pipelines (16; 17).

Several limitations should be acknowledged. Not all known coding genes are currently linked to a disease panel; we prioritised high-confidence, traceable annotations over broad but less reliable coverage. Some genes are over-represented across multiple panels due to historical research biases, leading to non-uniform panel enrichment. Additionally, not all panels returned by queries may be equally informative. For example, during benchmarking, the COVID-19 research panel frequently appeared alongside the PID panel due to overlapping genes. While technically accurate, such panels may be less relevant to users focused on clinical diagnosis of immunodeficiency. In bioinformatic workflows, these choices can be refined systematically or excluded using filters or downstream logic. In one of the five case studies, a disease-based query returned three panels, one of which did not contain the causal gene. This is expected, as broader classifications may include non-specific panels, hindering sensitivity. However, the query was still considered 100% successful, as the causal gene appeared in the combined result set, consistent with the recommended bioinformatic strategy of using the union of returned panels.

PanelAppRex bridges expert-curated panel knowledge with genome-scale statistical reasoning. It offers a robust tool for interactive search, a validated dataset for programmatic access, and a scalable framework for quantitative genomic modelling. Future work will focus on expanding supported queries, integrating additional variant types and annotations, and enabling more sophisticated applications in variant interpretation and risk estimation.

5 Conclusion

PanelAppRex offers a robust solution for bioinformatically aggregating and querying gene panel data. Its additional user interface with sophisticated search feature simplifies data exploration and enhances variant interpretation.

Acknowledgements

We acknowledge GE for providing public access to the PanelApp data. The use of data from GE panelapp was used under the Apache License 2.0. The use of data from UniProt was licensed under Creative Commons Attribution 4.0 International (CC BY 4.0). ClinVar asks its users who distribute or copy data to provide attribution to them as a data source in publications and websites (2). Ensembl data and code are available without restriction; software is provided under the Apache License 2.0 (4).

Contributions

DL performed main analyses and wrote the manuscript.

Competing interest

The authors declare no competing interest.

Ethics statement

This study only used data which was previously published and publicly available, as cited in the manuscript. This SwissPedHealth study, under which this work was carried out, was approved based on the advice of the ethical committee Northwest and Central Switzerland (EKNZ, AO_2022-00018). The study was conducted in accordance with the Declaration of Helsinki.

Funding

This project was supported through the grant Swiss National Science Foundation (SNF) 320030_201060, and NDS-2021-911 (SwissPedHealth) from the Swiss Personalized Health Network and the Strategic Focal Area 'Personalized Health and Related Technologies' of the ETH Domain (Swiss Federal Institutes of Technology).

References

[1] Antonio Rueda Martin, Eleanor Williams, Rebecca E. Foulger, Sarah Leigh, Louise C. Daugherty, Olivia Niblock, Ivone U. S. Leong, Katherine R. Smith, Oleg Gerasimenko, Eik Haraldsdottir, Ellen Thomas, Richard H. Scott, Emma Baple, Arianna Tucci, Helen Brittain, Anna De Burca, Kristina Ibañez, Dalia Kasperaviciute, Damian Smedley, Mark Caulfield, Augusto Rendon, and Ellen M. McDonagh. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nature Genetics*, 51(11):1560–1565, November 2019. ISSN 1061-4036, 1546-1718. doi: 10.1038/s41588-019-0528-2. URL https://www.nature.com/articles/s41588-019-0528-2.

[2] Melissa J Landrum, Jennifer M Lee, Mark Benson, Garth R Brown, Chen Chao, Shanmuga Chitipiralla, Baoshan Gu, Jennifer Hart, Douglas Hoffman, Wonhee Jang, Karen Karapetyan, Kenneth Katz, Chunlei Liu, Zenith Maddipatla, Adriana Malheiro, Kurt McDaniel, Michael Ovetsky, George Riley, George Zhou, J Bradley Holmes, Brandi L Kattman, and Donna R Maglott. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, 46(D1):D1062–D1067, January 2018. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkx1153. URL http://academic.oup.com/nar/article/46/D1/D1062/4641904.

[3] The UniProt Consortium, Alex Bateman, Maria-Jesus Martin, Sandra Orchard, Michele Magrane, Aduragbemi Adesina, Shadab Ahmad, Bowler-Barnett, and Others. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Research*, 53(D1):D609–D617, January 2025. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkae1010. URL https://academic.oup.com/nar/article/53/D1/D609/7902999.

[4] Sarah C Dyer, Olanrewaju Austine-Orimoloye, Andrey G Azov, Matthieu Barba, If Barnes, Vianey Paola Barrera-Enriquez, Arne Becker, Ruth Bennett, Martin Beracochea, Andrew Berry, Jyothish Bhai, Simarpreet Kaur Bhurji, Sanjay Boddu, Paulo R Branco Lins, Lucy Brooks, Shashank Budhanuru Ramaraju, Lahcen I Campbell, Manuel Carbajo Martinez, Mehrnaz Charkhchi, Lucas A Cortes, Claire Davidson, Sukanya Denni, Kamalkumar Dodiya, Sarah Donaldson, Bilal El Houdaigui, Tamara El Naboulsi, Oluwadamilare Falola, Reham Fatima, Thiago Genez, Jose Gonzalez Martinez, Tatiana Gurbich, Matthew Hardy, Zoe Hollis,

Toby Hunt, Mike Kay, Vinay Kaykala, Diana Lemos, Disha Lodha, Nourhen Mathlouthi, Gabriela Alejandra Merino, Ryan Merritt, Louisse Paola Mirabueno, Aleena Mushtaq, Syed Nakib Hossain, José G Pérez-Silva, Malcolm Perry, Ivana Piližota, Daniel Poppleton, Irina Prosovetskaia, Shriya Raj, Ahamed Imran Abdul Salam, Shradha Saraf, Nuno Saraiva-Agostinho, Swati Sinha, Botond Sipos, Vasily Sitnik, Emily Steed, Marie-Marthe Suner, Likhitha Surapaneni, Kyösti Sutinen, Francesca Floriana Tricomi, Ian Tsang, David Urbina-Gómez, Andres Veidenberg, Thomas A Walsh, Natalie L Willhoft, Jamie Allen, Jorge Alvarez-Jarreta, Marc Chakiachvili, Jitender Cheema, Jorge Batista da Rocha, Nishadi H De Silva, Stefano Giorgetti, Leanne Haggerty, Garth R Ilsley, Jon Keatley, Jane E Loveland, Benjamin Moore, Jonathan M Mudge, Guy Naamati, John Tate, Stephen J Trevanion, Andrea Winterbottom, Bethany Flint, Adam Frankish, Sarah E Hunt, Robert D Finn, Mallory A Freeberg, Peter W Harrison, Fergal J Martin, and Andrew D Yates. Ensembl 2025. *Nucleic Acids Research*, 53(D1):D948–D957, January 2025. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkae1071. URL https://academic.oup.com/nar/article/53/D1/D948/7916352.

[5] Dylan Lawless. PanelAppRex aggregates disease gene panels and facilitates sophisticated search. March 2025. doi: 10.1101/2025.03.20.25324319. URL http://medrxiv.org/lookup/doi/10.1101/2025.03.20.25324319.

[6] Konrad J Karczewski, Laurent C Francioli, Grace Tiao, Beryl B Cummings, Jessica Alföldi, Qingbo Wang, Ryan L Collins, Kristen M Laricchia, Andrea Ganna, Daniel P Birnbaum, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809):434–443, 2020.

[7] Xiaoming Liu, Chang Li, Chengcheng Mou, Yibo Dong, and Yicheng Tu. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome Medicine*, 12(1):103, December 2020. ISSN 1756-994X. doi: 10.1186/s13073-020-00803-9. URL https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-020-00803-9.

[8] Wolfgang Huber Steffen Durinck <Biomartdev@Gmail. Com>. biomaRt, 2017. URL https://bioconductor.org/packages/biomaRt.

[9] Luisa Karla P. Arruda, Luana Delcaro, Priscila B. Botelho Palhas, Marina M. Dias, Valdair F. Muglia, Erick C. Castelli, Konrad Bork, and Adriana S. Moreno. Genetic Analysis As a Practical Tool to Diagnose Hereditary Angioedema with Normal C1 Inhibitor: A Case Report. *Journal of Allergy and Clinical Immunology*, 135 (2):AB197, February 2015. ISSN 00916749. doi: 10.1016/j.jaci.2014.12.1578. URL https://linkinghub.elsevier.com/retrieve/pii/S0091674914033594.

[10] Maeve A. McAleer, Elizabeth Pohler, Frances J.D. Smith, Neil J. Wilson, Christian Cole, Stuart MacGowan, Jennifer L. Koetsier, Lisa M. Godsel, Robert M. Harmon, Robert Gruber, Debra Crumrine, Peter M. Elias, Michael McDermott, Karina Butler, Annemarie Broderick, Ofer Sarig, Eli Sprecher, Kathleen J. Green,

W.H. Irwin McLean, and Alan D. Irvine. Severe dermatitis, multiple allergies, and metabolic wasting syndrome caused by a novel mutation in the N-terminal plakin domain of desmoplakin. *Journal of Allergy and Clinical Immunology*, 136(5): 1268–1276, November 2015. ISSN 00916749. doi: 10.1016/j.jaci.2015.05.002. URL https://linkinghub.elsevier.com/retrieve/pii/S0091674915006533.

[11] Dorit Verhoeven, Dieneke Schonenberg-Meinema, Frédéric Ebstein, Jonas J. Papendorf, Paul A. Baars, Ester M.M. Van Leeuwen, Machiel H. Jansen, Arjan C. Lankester, Mirjam Van Der Burg, Sandrine Florquin, Saskia M. Maas, Silvana Van Koningsbruggen, Elke Krüger, J. Merlijn Van Den Berg, and Taco W. Kuijpers. Hematopoietic stem cell transplantation in a patient with proteasome-associated autoinflammatory syndrome (PRAAS). *Journal of Allergy and Clinical Immunology*, 149(3):1120–1127.e8, March 2022. ISSN 00916749. doi: 10.1016/j.jaci.2021.07.039. URL https://linkinghub.elsevier.com/retrieve/pii/S0091674921012446.

[12] Aude Magerus-Chatinet, Marie-Claude Stolzenberg, Nina Lanzarotti, Bénédicte Neven, Cécile Daussy, Capucine Picard, Nathalie Neveux, Mukesh Desai, Meghana Rao, Kanjaksha Ghosh, Manisha Madkaikar, Alain Fischer, and Frédéric Rieux-Laucat. Autoimmune lymphoproliferative syndrome caused by a homozygous null FAS ligand (FASLG) mutation. *Journal of Allergy and Clinical Immunology*, 131 (2):486–490, February 2013. ISSN 00916749. doi: 10.1016/j.jaci.2012.06.011. URL https://linkinghub.elsevier.com/retrieve/pii/S0091674912009645.

[13] Nigel Sharfe, Amit Nahum, Andrea Newell, Harjit Dadi, Bo Ngan, Sergio L. Pereira, Jo-Anne Herbrick, and Chaim M. Roifman. Fatal combined immunodeficiency associated with heterozygous mutation in STAT1. *Journal of Allergy and Clinical Immunology*, 133(3):807–817, March 2014. ISSN 00916749. doi: 10.1016/j.jaci.2013.09.032. URL https://linkinghub.elsevier.com/retrieve/pii/S0091674913014796.

[14] William B. Hannah, Mitchell L. Drumm, Keith Nykamp, Tiziano Pramparo, Robert D. Steiner, and Steven J. Schrodi. Using genomic databases to determine the frequency and population-based heterogeneity of autosomal recessive conditions. *Genetics in Medicine Open*, 2:101881, 2024. ISSN 29497744. doi: 10.1016/j.gimo.2024.101881. URL https://linkinghub.elsevier.com/retrieve/pii/S2949774424010276.

[15] Johannes Zschocke, Peter H. Byers, and Andrew O. M. Wilkie. Mendelian inheritance revisited: dominance and recessiveness in medical genetics. *Nature Reviews Genetics*, 24(7):442–463, July 2023. ISSN 1471-0056, 1471-0064. doi: 10.1038/s41576-023-00574-0. URL https://www.nature.com/articles/s41576-023-00574-0.

[16] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J.

Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, August 2021. ISSN 0028-0836, 1476-4687. doi: 10.1038/s41586-021-03819-2. URL https://www.nature.com/articles/s41586-021-03819-2.

[17] Jun Cheng, Guido Novati, Joshua Pan, Clare Bycroft, Akvilė Žemgulytė, Taylor Applebaum, Alexander Pritzel, Lai Hong Wong, Michal Zielinski, Tobias Sargeant, Rosalia G. Schneider, Andrew W. Senior, John Jumper, Demis Hassabis, Pushmeet Kohli, and Žiga Avsec. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*, 381(6664):eadg7492, September 2023. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.adg7492. URL https://www.science.org/doi/10.1126/science.adg7492.

6 Supplemental

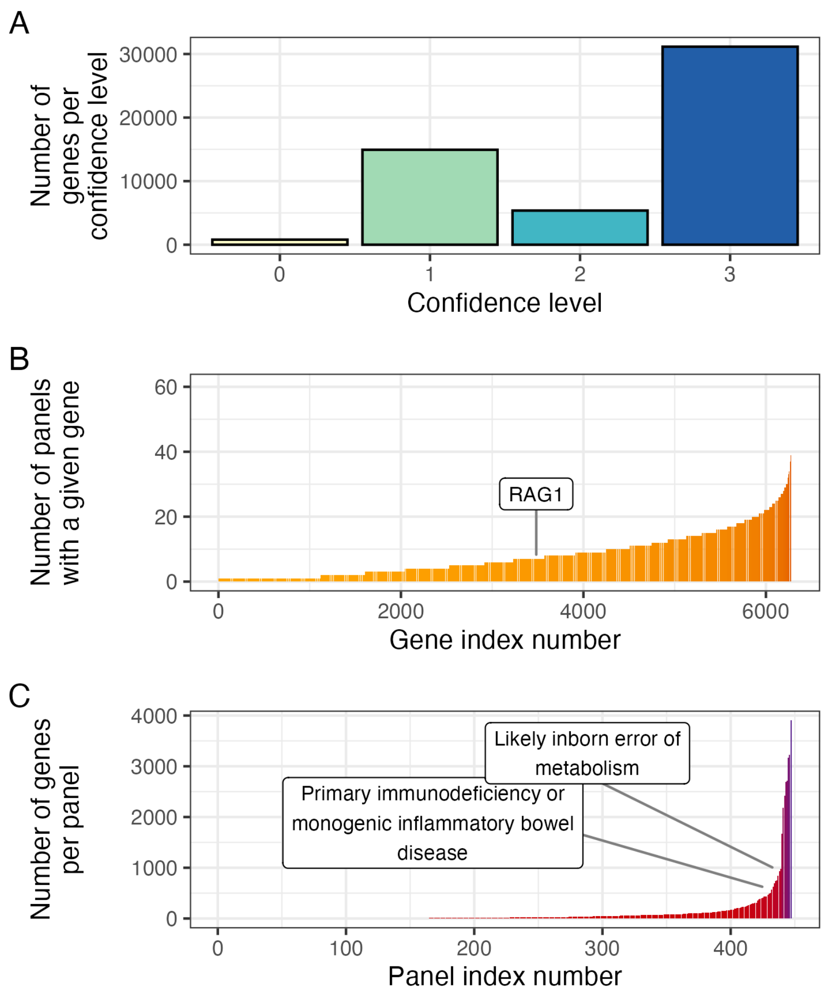

Figure S1: **Summary of gene confidence levels, reuse across panels, and panel sizes in the PanelAppRex dataset.** (A) Number of genes per confidence level. (B) Number of panels in which each gene is included, with the example gene *RAG1* highlighted to demonstrate that it is present in 7 panels. (C) Number of genes per panel, with two representative panels annotated: panel ID 398 (PID/iei) and panel ID 1220 (iem).

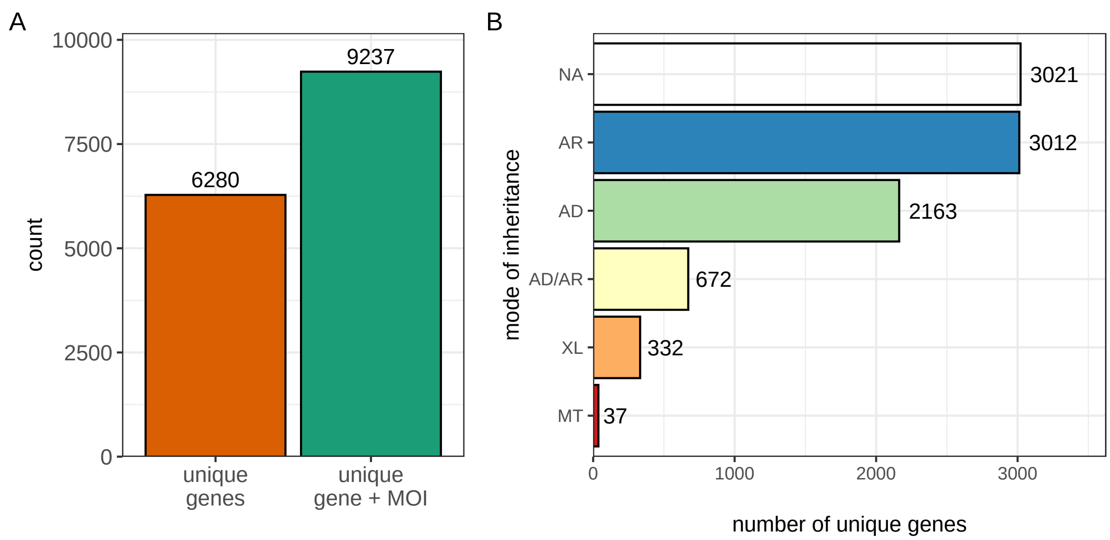

Figure S2: **Summary of mode of inheritance annotations in the PanelAppRex dataset.** (A) Counts of unique genes annotated with each MOI, based on non-redundant gene-MOI combinations. (B) Total number of unique genes and total number of gene-MOI combinations in the harmonised PanelAppRex dataset.