

PanelAppRex: A Sophisticated Search Function for Gene Panel Data

Dylan Lawless^{*1}

¹Department of Intensive Care and Neonatology, University Children’s Hospital Zürich, University of Zürich, Switzerland.

March 20, 2025

Abstract

Motivation: Gene panel data provides critical insights into disease-gene correlations. However, aggregating and interrogating this diverse dataset can be challenging. PanelAppRex addresses this by first preparing a machine-readable aggregate and second by offering a sophisticated natural language search interface that streamlines data exploration for both clinical and research applications.

Results: PanelAppRex aggregates gene panel data from source including ClinVar, Uniprot, and Genomics England’s PanelApp, including the approved panels used in the NHS National Genomic Test Directory and the 100,000 Genomes Project. It enables users to execute complex queries by gene names, phenotypes, disease groups and more, returning integrated datasets in multiple downloadable formats. Benchmarking demonstrates that the system effectively simplifies variant discovery and interpretation to enhance workflow efficiency. The greatest benefit is the analysis ready format for bioinformatic integration.

Availability and implementation: PanelAppRex is available under the MIT licence. The source code and data are accessible at <https://github.com/DylanLawless/PanelAppRex>. The dataset is maintained for a minimum of two years following publication.

1 Introduction

Gene panels are pivotal for the diagnosis and interpretation of genetic disorders. Sources like Genomics England’s PanelApp and PanelApp Australia host comprehensive panels that support genomic testing (1). For instance, these are integral in the

^{*}Addresses for correspondence: Dylan.Lawless@uzh.ch

NHS and research projects such as the 100,000 Genomes Project. Despite its utility, manual panel selection and data aggregation remain labour intensive. PanelAppRex was developed to simplify this process by automating data retrieval from the PanelApp API and integrating the data into machine and user-friendly formats. The novel natural language search capability further streamlines the discovery of disease gene panels by allowing queries based on gene names, phenotypes, disease groups and additional key attributes which were generated with evidence-based Retrieval-augmented generation (RAG).

2 Materials and methods

2.1 Implementation

PanelAppRex is implemented in R and integrates data from Genomics England’s PanelApp. It performed credentialed access to the API 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, and disease information), and several metadata summary statistics. In addition, the tool incorporates a natural language processing module to interpret and 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.

2.2 Usage

Online, queries can be executed via the integrated search bar in our HTML version, where a JavaScript function splits the query into individual terms and progressively filters rows - 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.

Bioinformatically, users can import the provided, ready-for-use, datasets in TSV or Rds formats. Users are most likely to merge with their own omic data based on merging with gene/protein 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)

```

3 Results

PanelAppRex successfully aggregates data, currently from 451 panels, and several genomics databases to offer a user-friendly natural language search functionality (**Figure 1**). Users can retrieve results filtered by gene names, phenotypes, disease groups and other criteria. The system returns a table view with panel details and provides options for exporting results in CSV, Excel, or PDF formats.

3.1 Validation benchmarking

To mimic a clinician diagnosing a new disease such as primary immunodeficiency (PID), we began by systematically selecting from the current online catalogue from the Journal of Allergy and Clinical Immunology (JACI), using the first five results (2? ? -4). The clinical background from these studies was used to construct keyword queries from patient features, simulating a naïve starting position for a clinician (full sources, queries, and results in supplemental table 1). We then tested whether our PanelAppRex tool could successfully retrieve panels that included the final causal gene reported in each case study.

In our evaluation, PanelAppRex returned panels in which the causal gene was present in 93% of all cases, with the 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.

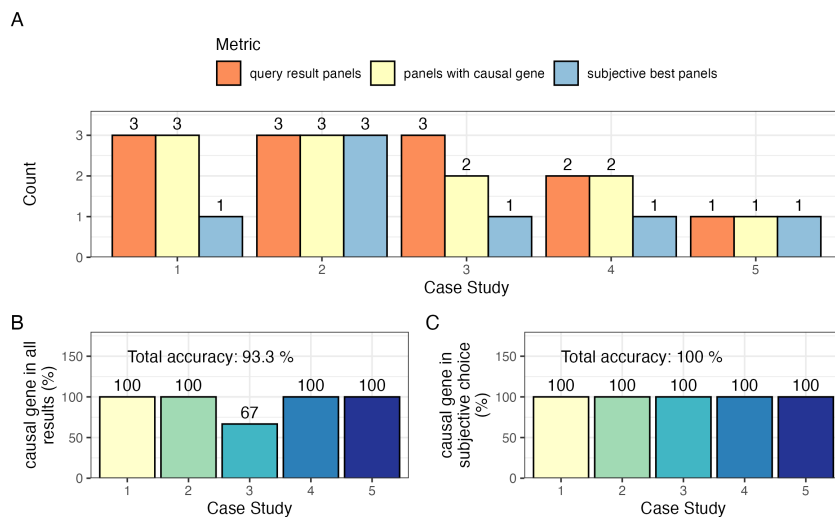
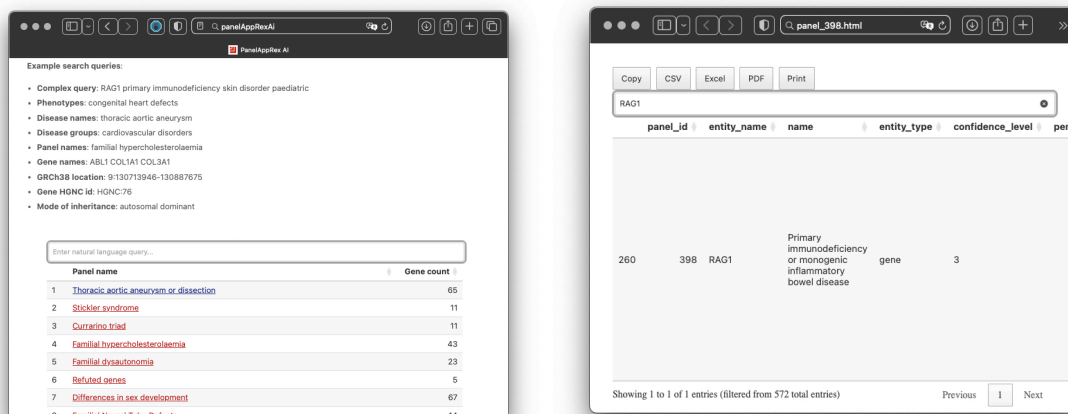


Figure 1: Screenshot of the PanelApp interface displaying search results for a complex natural language query. Key metrics such as panel name and gene count are displayed in a user-friendly table format. The left window shows the whole database before searching for a panel. The right window shows the result of selecting a panel, revealing the complete detailed data table.

4 Summary

PanelAppRex offers a robust solution for aggregating and querying gene panel data. Its sophisticated natural language search feature simplifies data exploration and enhances variant interpretation. Future work will focus on expanding the range of supported queries and integrating additional genomic data sources, further supporting the needs of clinicians and researchers.

Acknowledgements

We acknowledge Genomics England for providing public access to the PanelApp data. The use of data from Genomics England panelapp was licensed under the Apache License 2.0 A permissive license whose main conditions require preservation of copyright and license notices. Contributors provide an express grant of patent rights. Licensed works, modifications, and larger works may be distributed under different terms and without source code. These permissions consist of commercial use, modification, distribution, patent use, and private use.

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] 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>.
- [3] 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>.

- [4] 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>.