# Genomic and Immunological Characterization of RAG-Deficient SCID: Implications for Personalized Therapeutic Approaches

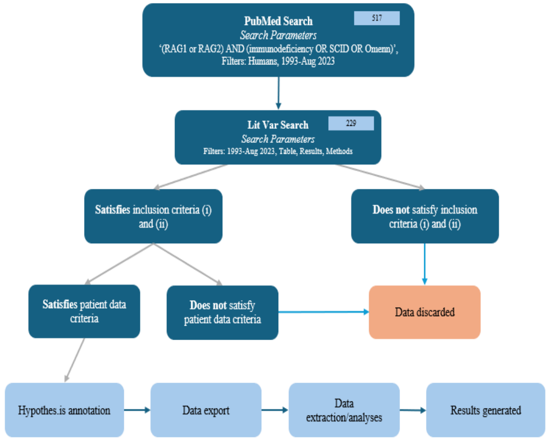
Roshni Arun[1], Justyne Ross[1], Courtney Thaxton[1]
[1]Department of Genetics at the University of North Carolina at Chapel Hill
Poster #3

## Background

- Severe combined immunodeficiency disease (SCID) encompasses a set of disorders with varying phenotypic expression
- *RAG*-deficient SCID is caused by complete or partial reduction in functionality of the RAG1 and RAG2 proteins
  - RAG proteins serve as crucial contributors to the genetic diversity of lymphocyte populations in the immune system, promoting combinatorial genetic rearrangement
- This analysis aims to improve the immunological characterization of *RAG*-deficient SCID by associating phenotype with type of genetic variation through observation of clinical presentation and genotype
- A parsing protocol was utilized to extract HTML data from annotated literature on *RAG*-deficient SCID patients
- Patient data was collected from demographic, clinical, and laboratory data, followed by categorizing types of genetic variants: missense, nonsense, and frameshift
- **Objective:** Derive phenotypic and genotypic patterns in patient data, serving as a foundation for future assessment of common treatments administered to each patient population, allowing for a comprehensive analysis of personalized therapeutic strategies based on genotype
- Utilizing predictive genomic screening can serve as a preventative measure to address potential pathogenicity based on the patient's genome

## Methods

Papers for this review were selected through a literature search on LitVar database using '(RAG1 or RAG2) AND (immunodeficiency OR SCID OR Omenn)'. Below illustrates the workflow of selecting *RAG*-deficient SCID patient data from 1994 to Aug 2023. Depicts literature search strategy, paper selection criteria, and data curation protocol utilized. Discarded data was literature that did not meet the criteria for inclusion.




Literature Inclusion Criteria:

i. Reported patient(s) SCID diagnosis
ii. Unambiguous identification of variant(s) in *RAG1* or *RAG2*

Datapoint inclusion criteria:

- Disease assertion
- Phenotypes that mapped to a HPO identifier OR CD3+, CD4+, CD8+ cell counts, or phenotyping information that does not have an appropriate HPO identifier
- Variant identification [name, ClinVarID, CAID, gnomAD frequency]

## Protocol Development

Annotations from Hypothes.is were converted to a CSV for further analyses

**Code Infrastructure:**

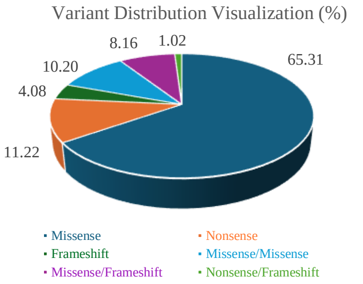
1. Importing the *pandas* library for data manipulation and tabular extraction.
2. Utilizing a loop to filter through columns and eliminate unnecessary characters and strings
3. Defining a function to separate long strings of text via a predefined delimiter
4. Utilizing list comprehension to create a list from parsed data and exporting it to a CSV

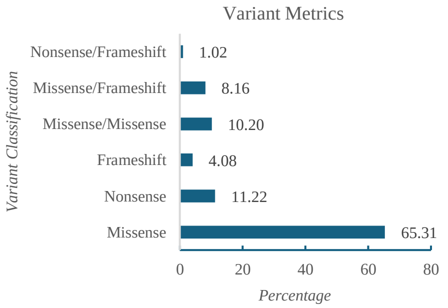
This parsing protocol allowed for data parsing and transformation by extracting and structuring from imported CSV files from Hypothes.is


## Genotype-Phenotype Associations

| Gene | Variant Allele | Count | Recurrent Phenotypes |
|---|---|---|---|
| RAG1 | p.L454Q | 2 | N/A |
| | p.W959X | 2 | N/A |
| | p.R396C | 4 | N/A |
| | p.R561H | 2 | N/A |
| | p.K86Vfs*33 | 8 | Severe viral infection (HP:0031692, n=8), Eosinophilia (HP:0001880, n=3), Lymphopenia (HP:0001888, n=8), Skin rash (HP:0000988, n=3), Pneumonia (HP:0002090, n=3), Autoimmune hemolytic anemia (HP:0001890, n=3), Hepatosplenomegaly (HP:0001433, n=3) |
| | p.H612R | 2 | Severe varicella zoster infection (HP:0032170, n=2), granulomas (HP:0032252, n=2) |
| | p.R559S | 2 | N/A |
| | p.S480G | 2 | N/A |
| | p.I956T | 2 | Lymphopenia (HP:0001888, n=2) |
| | p.R975W | 3 | Pneumonia (HP:0002090, n=2), Infection following live vaccination (HP:0020085, n=2) |
| | p.R394Q | 4 | Lymphopenia (HP:0001888, n=3), Decreased circulating antibody level (HP:0004313, n=3), Early onset of disease (n=3) |
| | p.H249R | 5 | Pneumonia (HP:0002090, n=3), Splenomegaly (HP:0001744, n=4), Diarrhea (HP:0002014, n=3), Lymphadenopathy (HP:0002716, n=3), BCGitis (HP:0020086, n=2) |
| | p.K820R | 2 | Diarrhea (HP:0002014, n=2), Splenomegaly (HP:0001744, n=2) |
| RAG2 | p.G35V | 15 | Pneumonia (HP:0002090, n=4), Diarrhea (HP:0002014, n=3), Oral thrush (HP:0009098, n=3), Skin rash (HP:0000988, n=2), GVHD (n=4), BCGitis (HP:0020086, n=8), Abnormal myocardium morphology (HP:0001637, n=2), RSV (n=2), Persistent CMV viremia (HP:0032247, n=2) |
| | p.G157V | 2 | Persistent CMV viremia (HP:0032247, n=2) |
| | p.E480X | 2 | Sepsis (HP:0100806, n=3) |
| | p.R73H | 2 | N/A |

## Results




Visualizations illustrating distribution of most common variants across *RAG1/RAG2* populations.

## Conclusions

**Value of Study**

- Understanding pathogenicity of variants is critical for patient diagnosis, treatment, and/or disease management
- Annotation of literature assists in identification of variants mentions for genes associated with disease
- Annotation enables increased efficiency in review and application of curation framework

**Discussions**

- Variability in NMD machinery in *RAG2* complicates determining if mutations cause phenotypic effects via protein absence or incorrect protein structure
- A need for biocuration validation: incorporating a review process helps to build trust in curation framework
- Variation in phenotypic descriptions in published literature and the lack of standard use of identifiers like Human Phenotype Ontology (HPO) present challenges in assessing genotype-phenotype correlations and phenotypic overlap among patients

**Future Experimentation**

- Utilize search workflow as a template for further curation on *RAG1* and *RAG2* or other target genes
- **Establish and iterating on a standardized pipeline** for data analysis facilitates consistency and comparability across different genetic studies
- Longitudinal studies involving larger patient cohorts and comprehensive genomic analyses are needed to further inform the development of novel therapeutic interventions

By analyzing patient data, derived patterns correlate genetic variants with the severity and nature of SCID symptoms, providing a foundation for clinicians to predict the clinical course of the disease based on the patient's genotype.

## Acknowledgements

**A Thank You**

Special thanks to the Berg Lab at the Department of Genetics in the University of North Carolina at Chapel Hill for their continued support and guidance on this project. This work was supported by the National Human Genome Research Institute (NHGRI) of the National Institutes of Health (NIH) under the award number U24HG009650. This content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Health.