Variant Analysis

# Project Overview

This project analyzes ~120,000 annotated variants from a Harvard Personal Genome Project individual, using functional and conservation scores to prioritize potentially impactful SNPs.

# Objective

- Compare genomic feature annotations (functionGVS vs. functionDBSNP)

- Evaluate variant impact using PolyPhen, Grantham, and GERP scores

- Select and analyze high-impact variants in PRDM16, IL37, and PNPLA1

- Link variants to phenotype, population frequency, and clinical relevance

# Functional Annotation Comparison

functionGVS integrates annotations from the Genome Variation Server, offering a broad context from multiple integrated data sources.

functionDBSNP is based on the NCBI dbSNP database, assigning function based on standard SNP identifiers and curated functional consequences.

Differences arise from reference builds and interpretation approaches. Example: one variant is labeled as "non-coding exon" in functionGVS but "non-coding transcript variant" in functionDBSNP.

# Variant Effect Prediction

|  |  |
| --- | --- |
| **Method** | **Purpose** |
| **PolyPhen** | Predicts structural/functional impact of amino acid changes |
| **Grantham Score** | Measures chemical dissimilarity between amino acids |
| **GERP (consScore)** | Indicates evolutionary conservation of genomic regions |

Grantham score quantifies physicochemical differences, while PolyPhen integrates structure and sequence conservation. GERP reflects evolutionary constraint.

# Score Comparison Insight

In the dataset, examples show:

- Variants with high Grantham and high PolyPhen generally reflect damaging changes.

- GERP scores often align with PolyPhen on highly conserved regions.

- Discrepancies exist - e.g., a variant with a high Grantham score but a benign PolyPhen score may indicate biochemical change without functional disruption.

# Variant Selection Criteria

Variants were filtered using PolyPhen predictions. Only variants with high impact scores were considered. Three were prioritized based on gene function relevance and known disease associations: PRDM16, IL37, PNPLA1.

Selected Variants

# Variant 1: PRDM16 (rs2493292)

- Function: Transcription factor in energy metabolism and brown fat development

- Change: Proline to Leucine (missense)

- Impact: Possibly damaging (PolyPhen); Grantham score suggests significant change

- Protein Impact Insight: Proline is rigid and often disrupts secondary structures. Substitution to leucine might disturb helix-loop integrity.

- Clinical: Linked to left ventricular noncompaction (LVNC8), diastolic/systolic blood pressure traits

- Zygosity: Heterozygous - may show partial phenotype in dominant traits or be asymptomatic carrier

- Population: T allele ~11% global; up to 17% in Europeans

# Variant 2: IL37 (rs3811046)

- Function: Anti-inflammatory cytokine in immune regulation

- Change: Missense variant

- Impact: Moderate functional alteration; PolyPhen = possibly damaging; Grantham = moderate

- Protein Impact Insight: Potential effect on cytokine-receptor interaction domain, possibly altering immune response modulation

- Clinical: Associated with IBD, IL-1 levels, and TB-related inflammatory response

- Zygosity: Homozygous - increases phenotypic expression likelihood for recessive traits

- Population: T allele ~60%, 81% in East Asians

# Variant 3: PNPLA1 (rs34598813)

- Function: Lipid metabolism and skin barrier function

- Change: Missense mutation (Aspartate to Glutamate)

- Impact: Conservative substitution, but in a critical lipid-binding region may affect hydrolytic efficiency

- Protein Impact Insight: May reduce triglyceride processing and barrier formation

- Clinical: Linked to autosomal recessive congenital ichthyosis

- Zygosity: Heterozygous - individual likely carrier, low phenotype risk unless in compound heterozygosity

- Population: G allele ~8% globally

# Summary of Genetic Variant Scoring

- Investigated SNPs in PRDM16, IL37, and PNPLA1 using dbSNP, GVS, and PolyPhen-2

- Scored variants using Grantham and GERP for conservation and impact analysis

- Identified potentially deleterious mutations affecting protein stability and function

- Prioritized variants with disease relevance for further experimental validation

# References

1. Allam, G. et al. (2016). Association of IL-37 gene polymorphisms with susceptibility to tuberculosis in Saudi subjects. *Microbiology and Immunology, 60*(11), 778–786. <https://doi.org/10.1111/1348-0421.12444>
2. Ahmad, F. et al. (2021). Variants in the PNPLA1 Gene in Families with Autosomal Recessive Congenital Ichthyosis Reveal Clinical Significance. *Molecular Syndromology, 12*(6), 351–361. <https://doi.org/10.1159/000516943>
3. SeattleSeq, dbSNP, UniProt, OMIM