



# Systematic analysis of genetic variants with conflicting interpretations of clinical significance

Tatyana E. Lazareva, Yury A. Barbitoff

FSBSI "The Research Institute of Obstetrics, Gynecology and Reproductology named after D.O.Ott", St. Petersburg, Russia

**Motivation and Aim:** The study aimed to systematically analyze properties of genetic variants and their associated genes for which discordance of variant interpretation have been reported in the ClinVar database. It will allow us to develop strategies for improving the consistency of NGS data interpretation.

5-10<sup>th</sup> August 2024, Novosibirsk

# NGS has revolutionized medical genetics and genomics

---

> 7 000

rare diseases

4 731

genes with associated  
monogenic  
disorder(s)

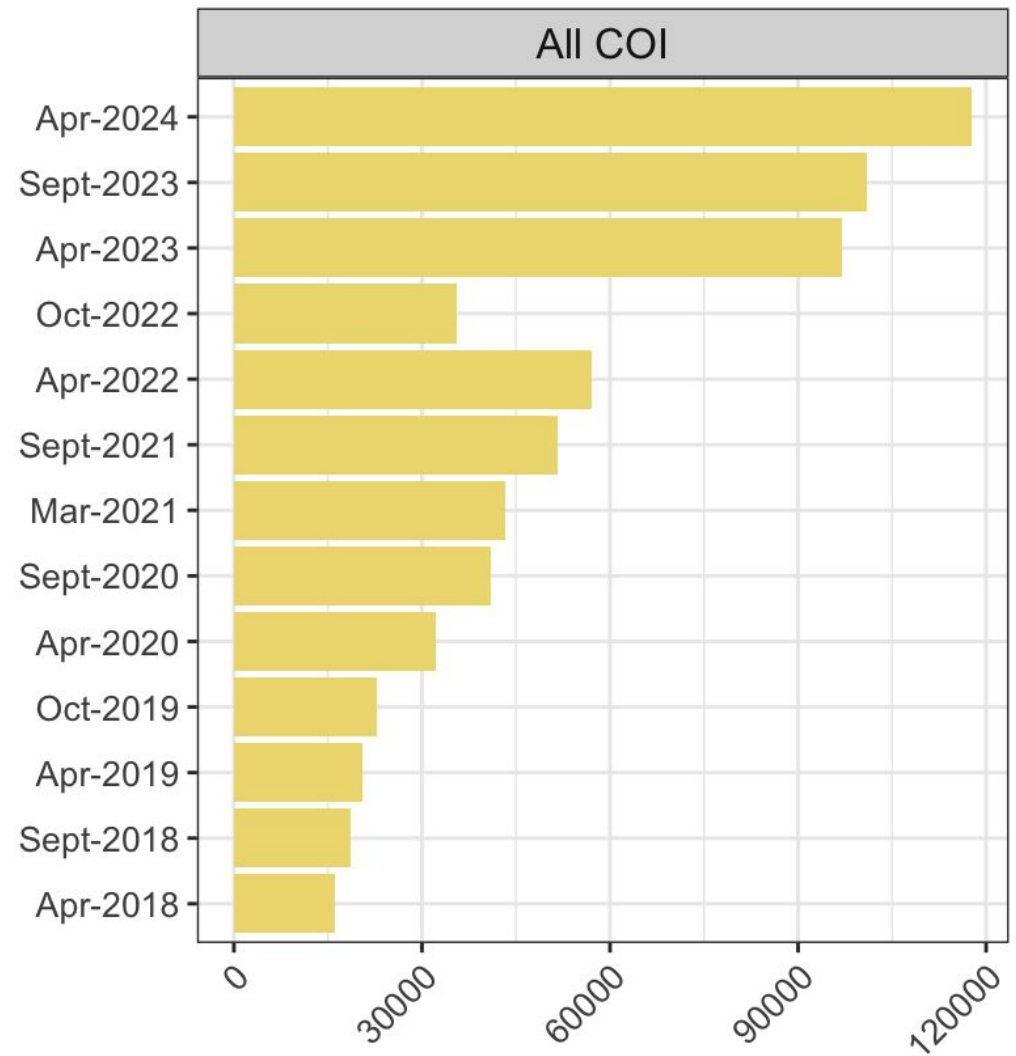
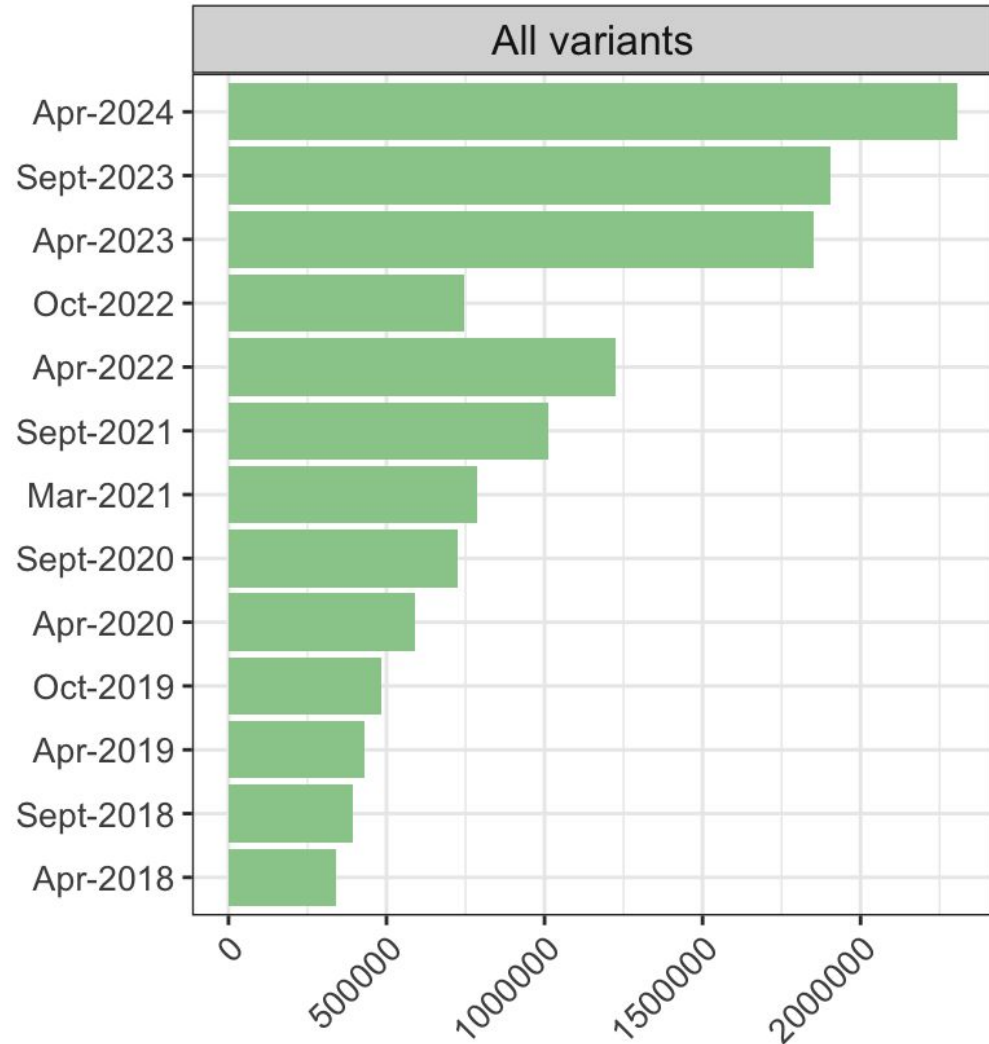
2 296 245

variants in these genes  
reported in ClinVar

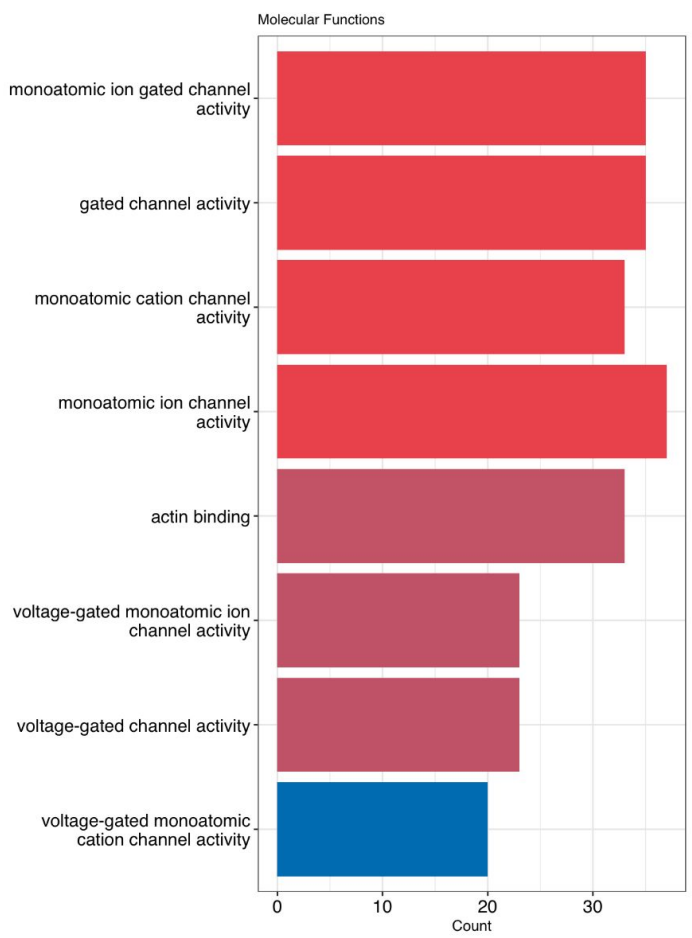
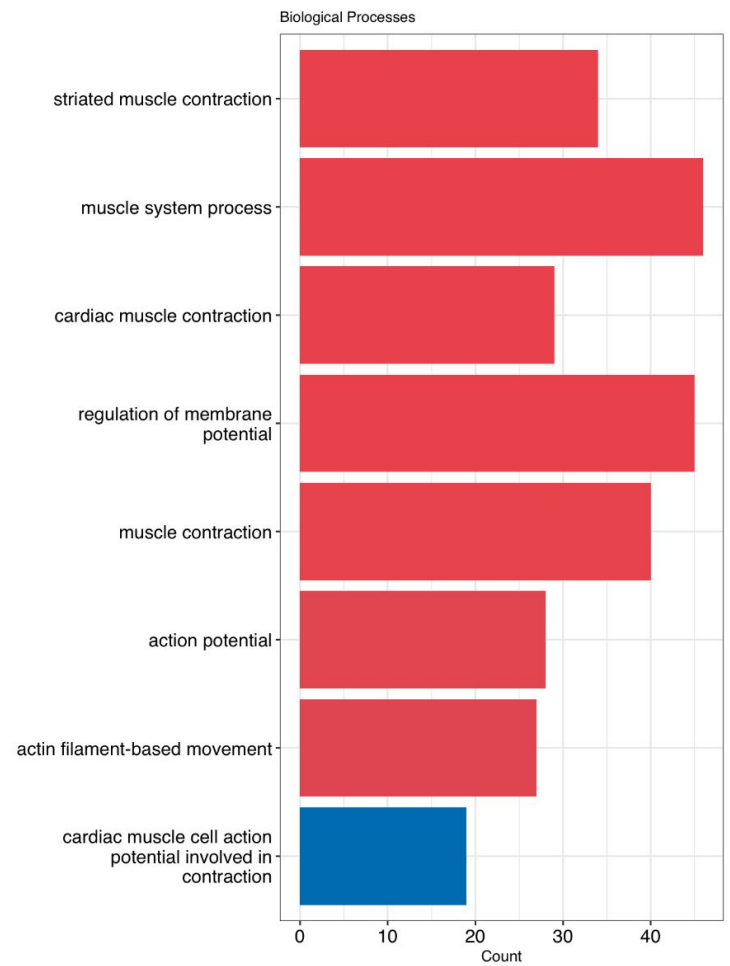
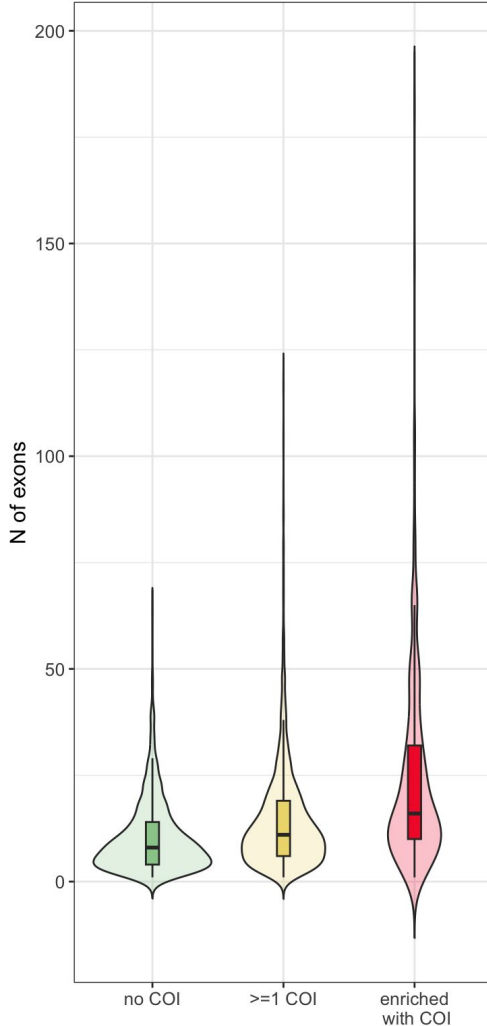
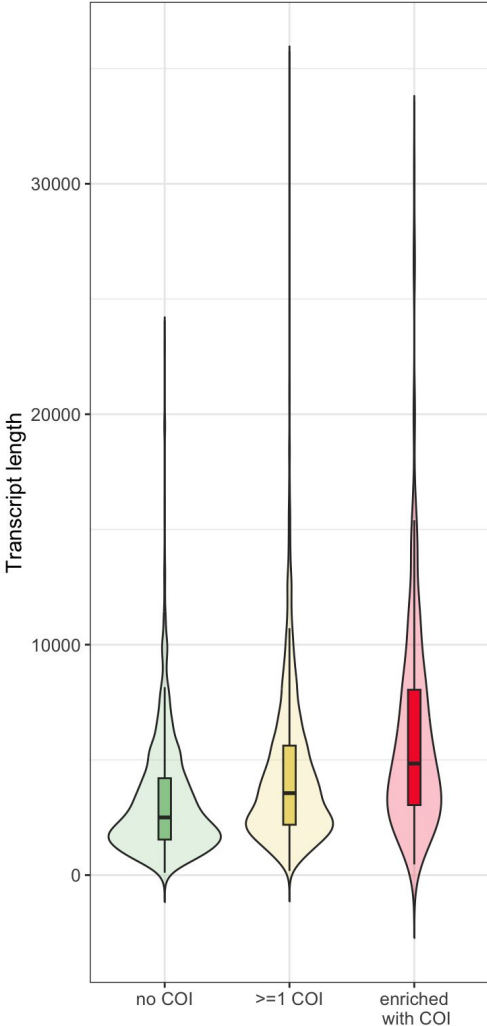
For 78% of genes  
variants with  
conflicting  
interpretations of  
pathogenicity  
have been  
reported

**Methods and algorithms:** We collected ClinVar data (April 2018 – April 2024) on variants in genes linked to rare diseases from OMIM and submitted to Ensembl BioMart. Variants with conflicting interpretations (COI) reported at least once during this period were defined as discordant. We employed the hypergeometric test to identify genes enriched for COI variants. Allele frequencies (gnomAD v2.1 genomes) and variant effects on gene products were analyzed. Additionally, we evaluated evolutionary conservation, isoform expression levels, transcript length, exon number, involvement in biological processes, and the number and type of inheritance of diseases associated with genes harboring (or lacking) COI variants.

Our analysis of **2,296,245** variants from 4,731 OMIM-linked disease genes in ClinVar revealed that **78% contained at least one COI variant.**



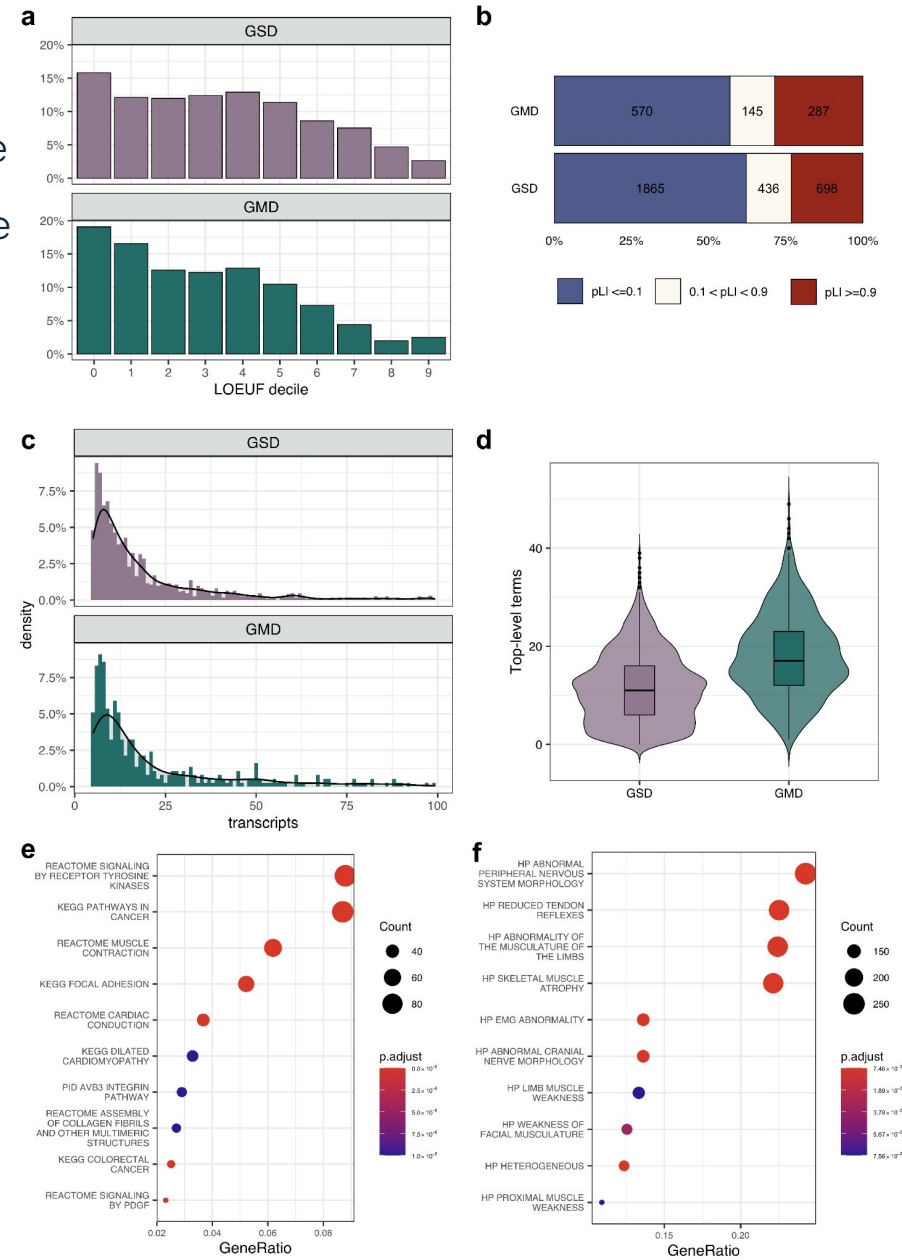
We identified **285 genes with statistically significant enrichment of COI variants**. These genes exhibited increased structural complexity, evidenced by a higher number of exons and longer transcripts ( $p\text{-value} < 2.2 \times 10^{-16}$ ). Notably, these genes were also associated with a **higher number of diseases per gene** ( $p\text{-value} < 2.2 \times 10^{-16}$ ). Moreover, these genes were **enriched for terms related to muscle tissue structure/function and terms related to neuromuscular and cardiovascular pathologies**, often exhibiting **autosomal dominant inheritance**.



Our findings align with previous studies on "phenotypic heterogeneity," where individuals with pathogenic variants in the same gene can exhibit variable disease severity or even develop different diseases

This phenomenon might explain the accumulation of COI variants in genes with multiple associated diseases, particularly those related to neuromuscular and cardiovascular pathologies

The observed associations between gene characteristics and COI variant detection could be utilized for establishing additional rules for annotation and prioritization of variants in these genes



General description of genes linked to multiple genetic disorders. **(a)** A histogram showing the number of genes in each LOEUF decile among genes with 1 (GSD) or with 2 and more associated (GMD) rare diseases. **(b)** Proportion of genes that are tolerant or not to LoF-mutations. **(c)** Histogram showing the distribution of the number of expressed transcripts GTEx data. **(d)** Number HPO terms annotated for the two groups of genes. **(e,f)** Dot plots showing the results of enrichment analysis against genes from MSigDB-derived canonical pathways **(e)** or HPO terms **(f)**.