GLB1 Variant Catalog — Key Insights Report

Version 1.0 · May 2026 · Cure GM1 Foundation

The following observations emerge from the first release of the comprehensive, multi-source *GLB1* variant catalog (1,295 variants; ClinVar · LOVD · gnomAD v4 · NTSAD · OMIM · published literature · 2025 WORLD Symposium) and its comparison against real-world evidence (RWE) from the 2025 WORLD Symposium cohort — 52 alleles from approximately 26 participants.

Catalog Summary: 1,295 variants · 122 Pathogenic (9.4%) · 128 Likely Pathogenic (9.9%) · 79 P/LP (6.1%) · 54 Conflicting (4.2%) · 248 VUS (19.2%) · 525 Likely Benign (40.5%) · 61 Benign (4.7%) · 78 Not Classified (6.0%)

24 variants observed in the 2025 WORLD Symposium cohort are highlighted in amber across all three output formats (PDF, Excel, CSV).

A note on ClinVar data: ClinVar records represent distinct variant submission entries across contributing laboratories and institutions. A variant present once in ClinVar may reflect one clinical observation or many — ClinVar does not report observation counts per submission, and no phase or genotype data (e.g., compound heterozygous vs. homozygous) is available from ClinVar records alone. All frequency-based observations in this document derive exclusively from the WORLD Symposium cohort, where allele counts were directly recorded.

1. Cross-Database Corroboration Is Currently Unavailable for 39% of the Catalog

503 variants appear in ClinVar only — and are not yet observed with additional population, database, or literature evidence. Further publication of these variants with phenotype and biochemical data will strengthen their interpretation.

Note: ClinVar entries may involve one or multiple submitters. What this insight identifies is specifically the absence of corroboration from any *other database* — not an assumption about the number of ClinVar submitters for any given variant.

2. LOVD-Only Variants Are Largely Unclassified

33 variants in the catalog come exclusively from LOVD and do not appear in ClinVar. The majority are Not Classified — no germline pathogenicity assessment. These are real variants known in academic and European laboratory databases.

Insights from the 2025 WORLD Symposium GM1 Cohort

3. The Arg201 Residue Is the Frequent Hotspot in the WORLD Symposium Cohort

Two adjacent variants at a single codon — **c.601C>T (p.Arg201Cys)** and **c.602G>A (p.Arg201His)** — together account for **16 of 52 WORLD Symposium alleles (31%)** at a single amino acid position, making Arg201 the highest-frequency position observed in this cohort.

These are two independent substitutions at the same codon (C→T producing Cys; G→A producing His), each appearing **8 times** in the WORLD Symposium data. Both are classified P/LP in the catalog. Arg201 is located within the enzyme's catalytic site; Hofer 2009 analysed it mechanistically in the context of a different substitution at this position.

Note on generalisability: The WORLD Symposium cohort comprises approximately 26 self-selected participants, predominantly from North American rare disease networks. Allele frequencies derived from this dataset describe the observed composition of this specific cohort and should not be extrapolated as population-representative estimates.

4. Significant Genetic Heterogeneity Is Observed

The top 5 WORLD Symposium variants account for **48% of all 52 alleles**. The top 3 alone account for 37%. Yet there are 27 unique variants across those 52 alleles — the remaining ~52% is distributed across 22 lower-frequency variants, most observed only once.

cDNA	Protein	Pathogenicity	WORLD Obs. (n=52)
c.601C>T	p.Arg201Cys	P/LP	8 (15%)
c.602G>A	p.Arg201His	P/LP	8 (15%)
c.622C>T	p.Arg208Cys	Pathogenic	3 (6%)
c.464T>G	p.Leu155Arg	Pathogenic	3 (6%)
c.931G>A	p.Gly311Arg	P/LP	3 (6%)

This pattern — a small number of frequently observed alleles atop a highly heterogeneous tail — has direct implications for clinical trial enrolment, gene therapy targeting, and natural history study design. A study enrolled primarily around high-frequency variants would underrepresent the majority of the cohort.

5. Five Conflicting-Classification Variants Appear in Symptomatic WORLD Symposium Patients

Of 54 Conflicting variants in the catalog, **5 appear in WORLD Symposium patients**. These may be reclassified with additional evidence.

cDNA	Protein	WORLD Obs.	Priority Note
c.481T>G	p.Trp161Gly	2	At least one benign/uncertain ClinVar submitter
c.191A>G	p.Tyr64Cys	1	Conflicting ClinVar submissions
c.574T>C	p.Tyr192His	1	Conflicting ClinVar submissions
c.1479G>T	p.Lys493Asn	1	Conflicting ClinVar submissions
c.1442G>A	p.Gly481Glu	1	Conflicting ClinVar submissions; also an OMIM allelic variant

The WORLD observation counts reflect patient-level allele data. Additional clinical data such as functional and segregation evidence, may strengthen the case for reclassification of these five variants.

6. The Richest Literature Source and the Richest RWE Source Have Almost No Allelic Overlap

Santamaria 2006 tags **42 variants** in the catalog — more than any other reference. Bidchol 2015 tags 30, Hofer 2009 tags 26. Yet the Santamaria Gypsy/Roma variants are absent from the WORLD Symposium data, and the Bidchol Indian splice variants are entirely absent. The richest published literature source and the richest RWE source are documenting substantially different patient populations.

For Cure GM1: the WORLD Symposium data is informative for the North American community it serves, but international collaborations (particularly European and Asian centres) would be needed to build a representative global frequency map.

Discussion

The aggregation of clinical, biochemical, functional, and segregation data in publicly available databases is a key component of ongoing efforts to improve variant classification and diagnostic outcomes for GM1 and many rare diseases. We strongly support the continued publication of genotype/phenotype data for all patients with GM1.

Data: ClinVar · LOVD · gnomAD v4 · NTSAD · OMIM · published literature · 2025 WORLD Symposium. RWE: Hastings C, Cogell A, Miller VR, Waggoner C. Genotypic Heterogeneity in GM1. 2025 WORLD Symposium Poster, AllStripes Research, San Francisco, CA. curegm1.org/world2025