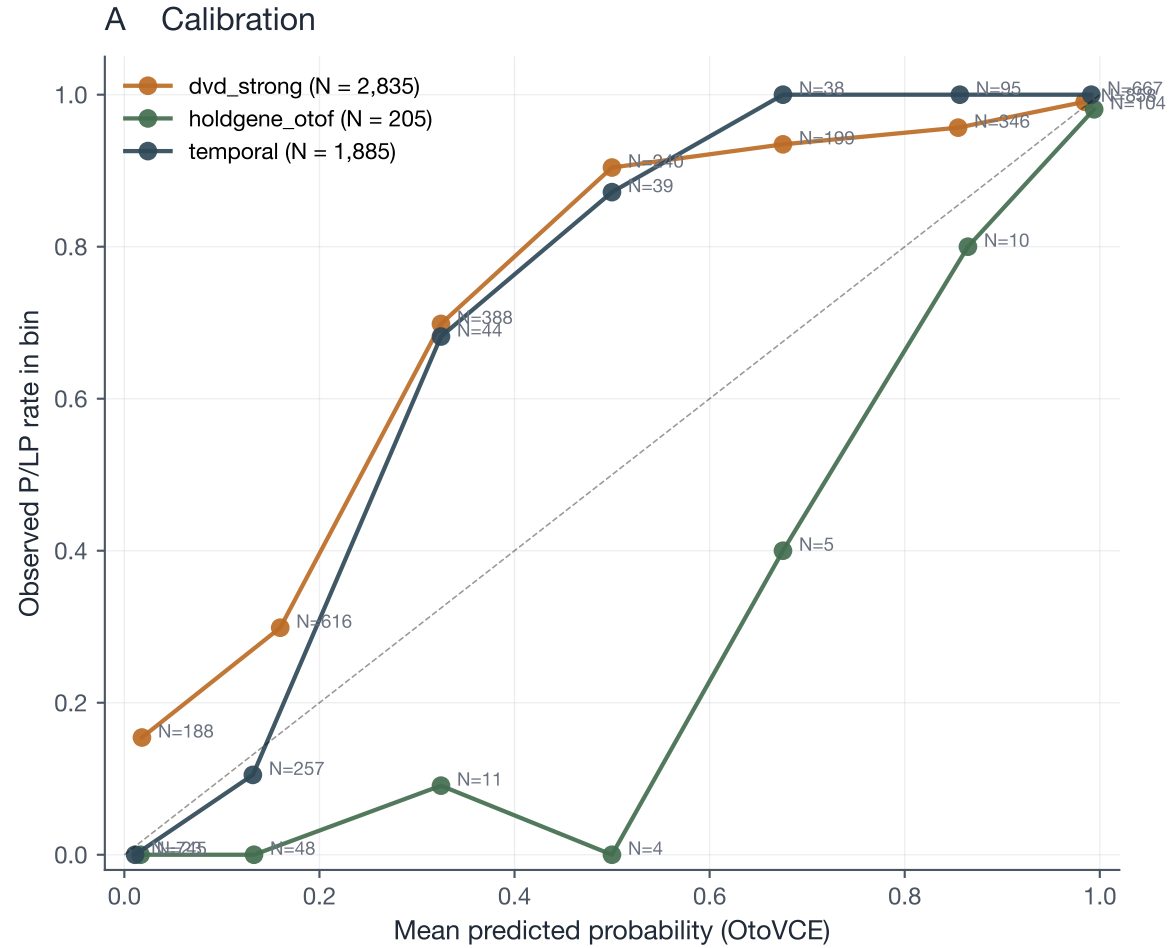
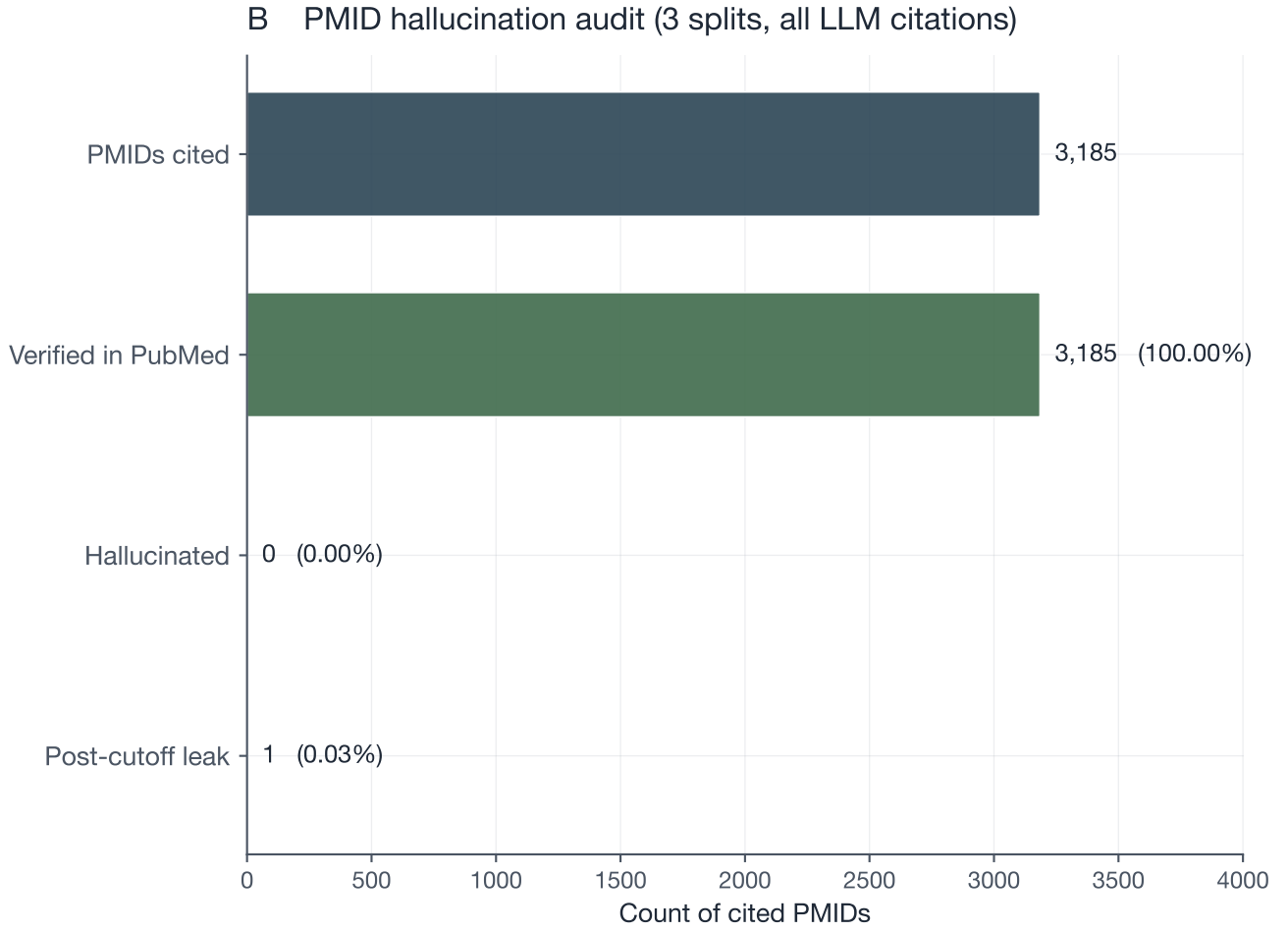
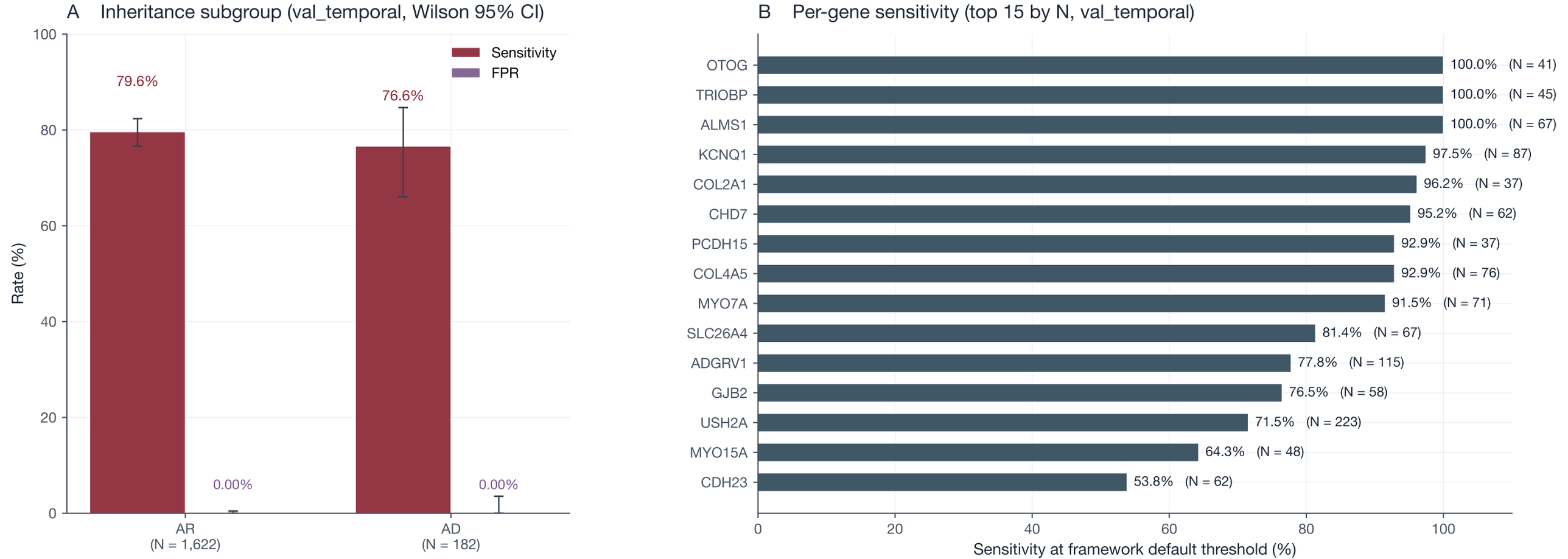
Supplementary Figure S1. Calibration of OtoVCE posterior probabilities and audit of LLM-cited PMIDs

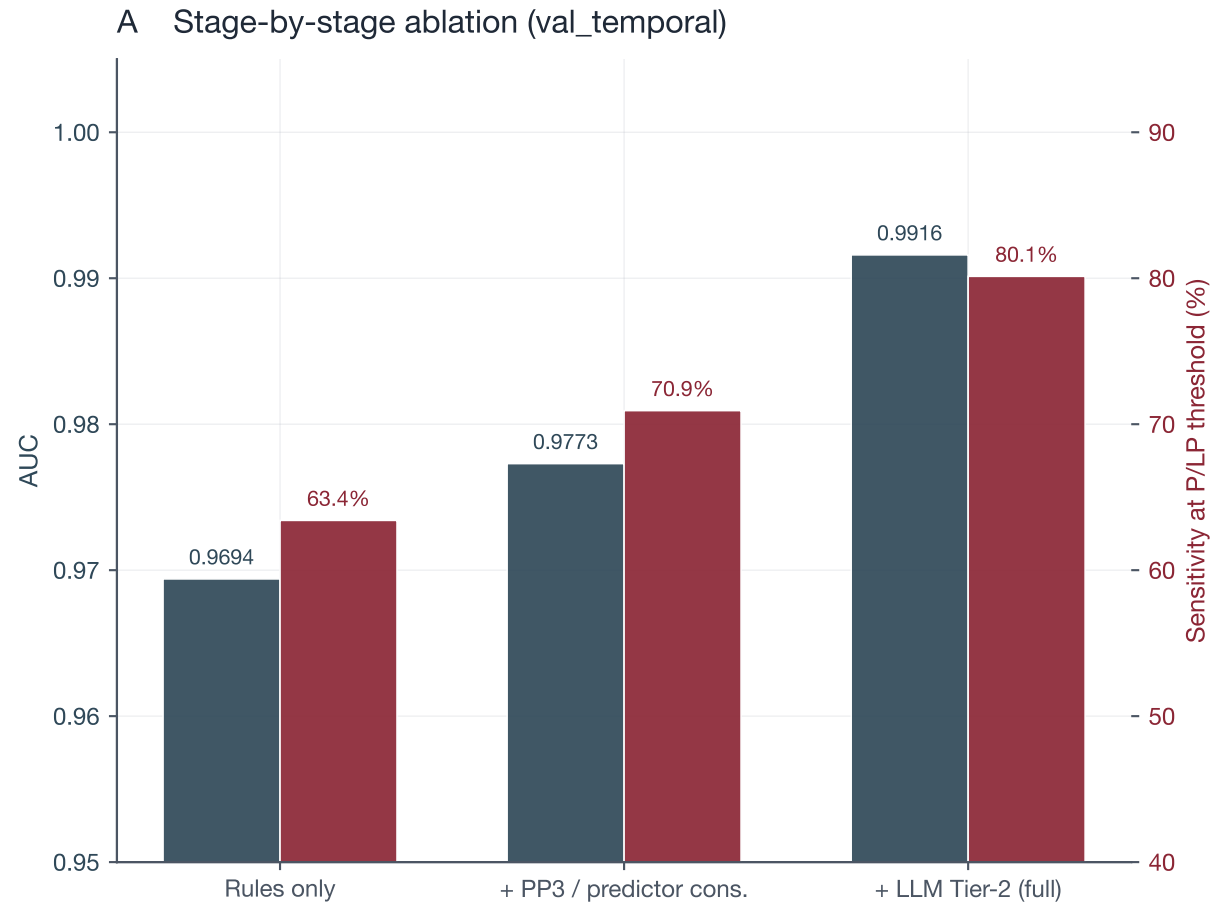


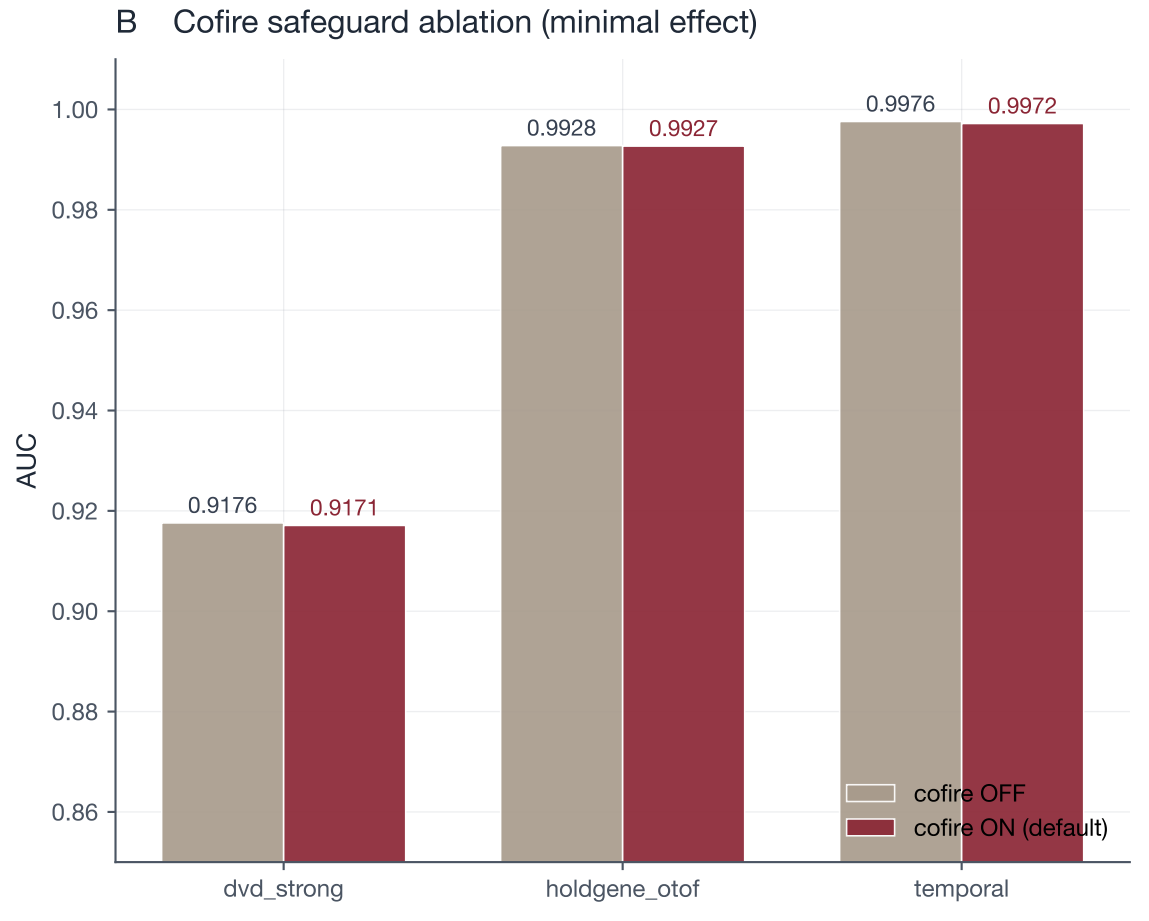


Supplementary Figure S2. Inheritance subgroup and per-gene robustness on val_temporal_extended

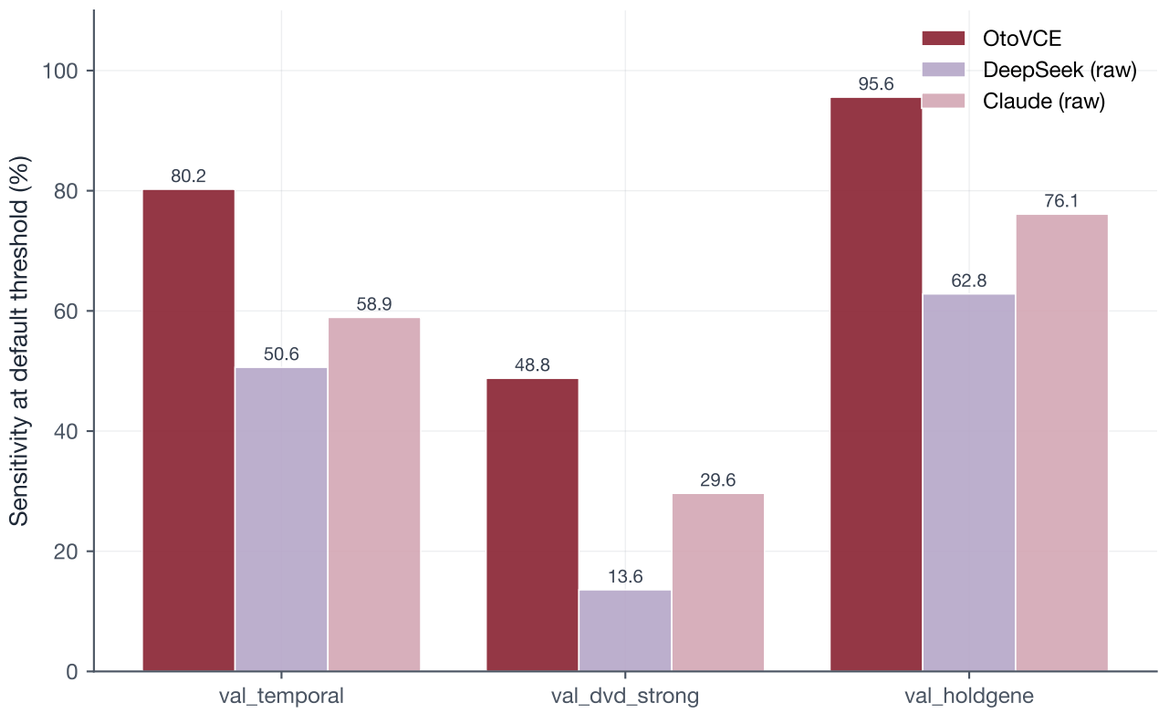


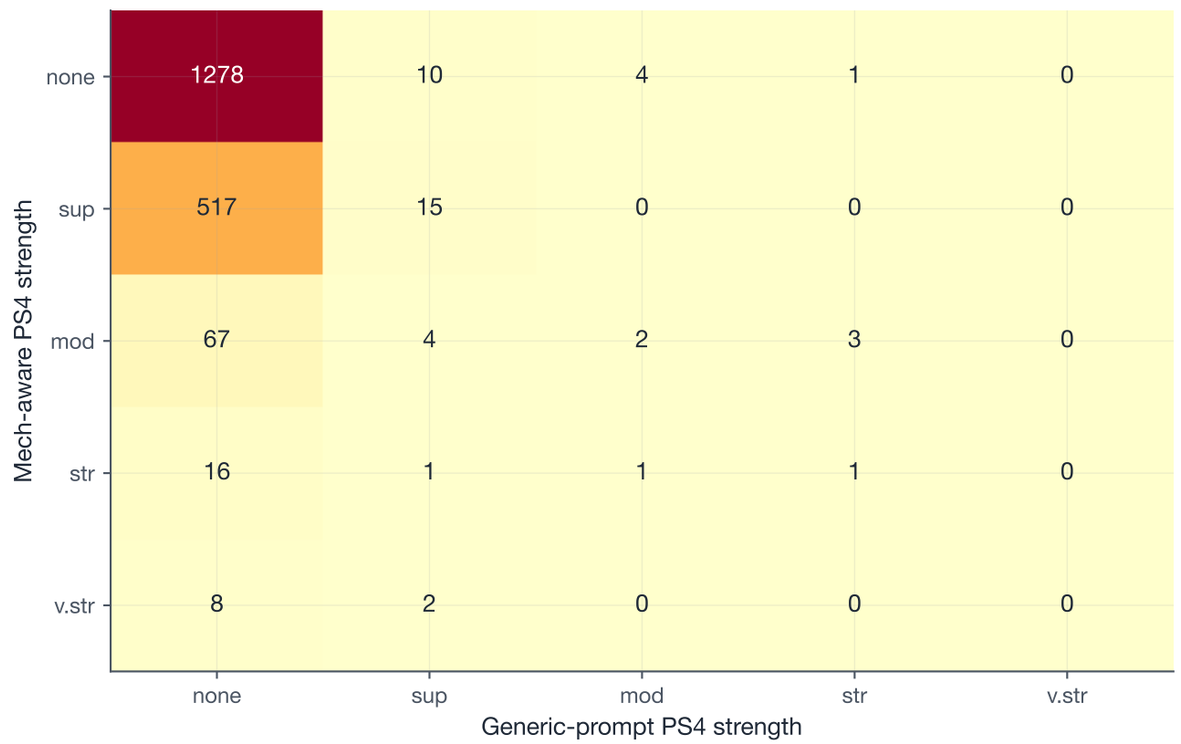
Supplementary Figure S3. Stage-by-stage and cofire-safeguard ablation

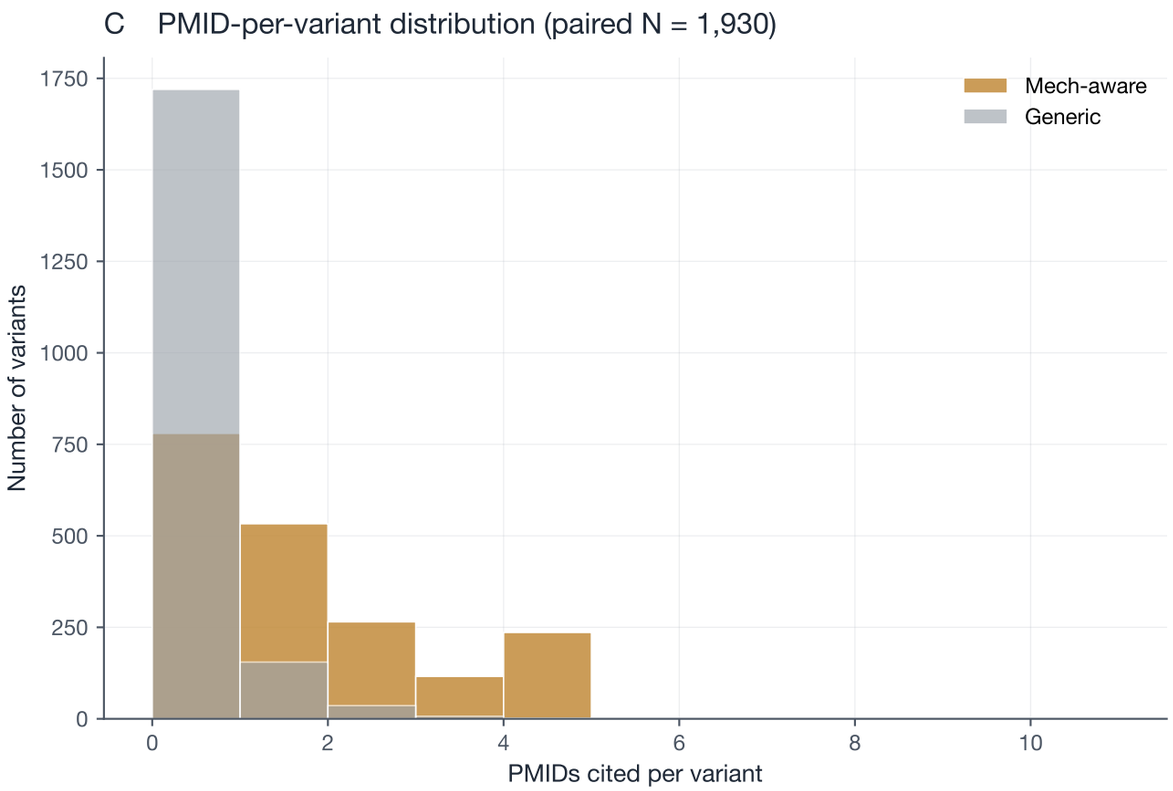




Supplementary Figure S4. Comparison with single-shot LLM baselines and detailed PS4 strength agreement







D Worked-example rationale comparison (COL2A1 p.Arg565Cys)

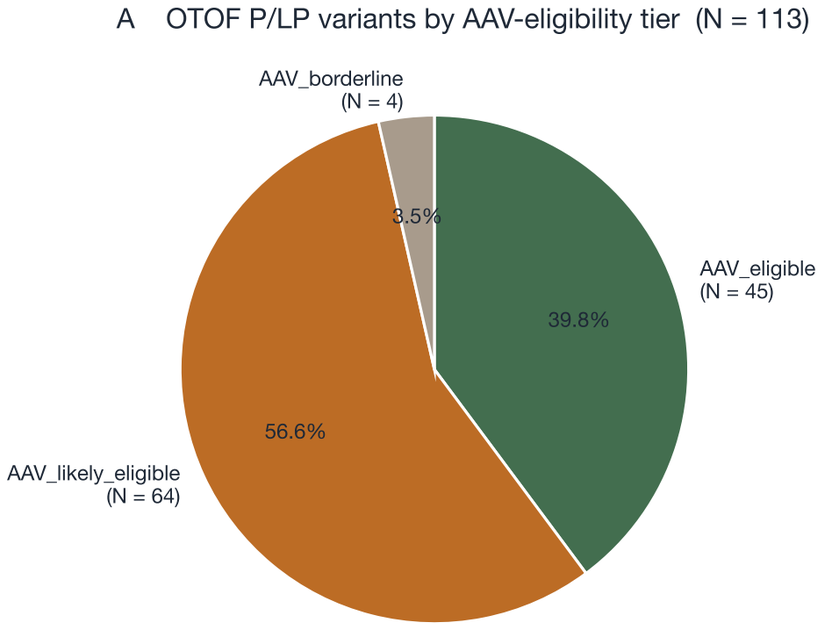
Mech-aware (PS3 = moderate, PS4 = strong; 4 PMIDs):

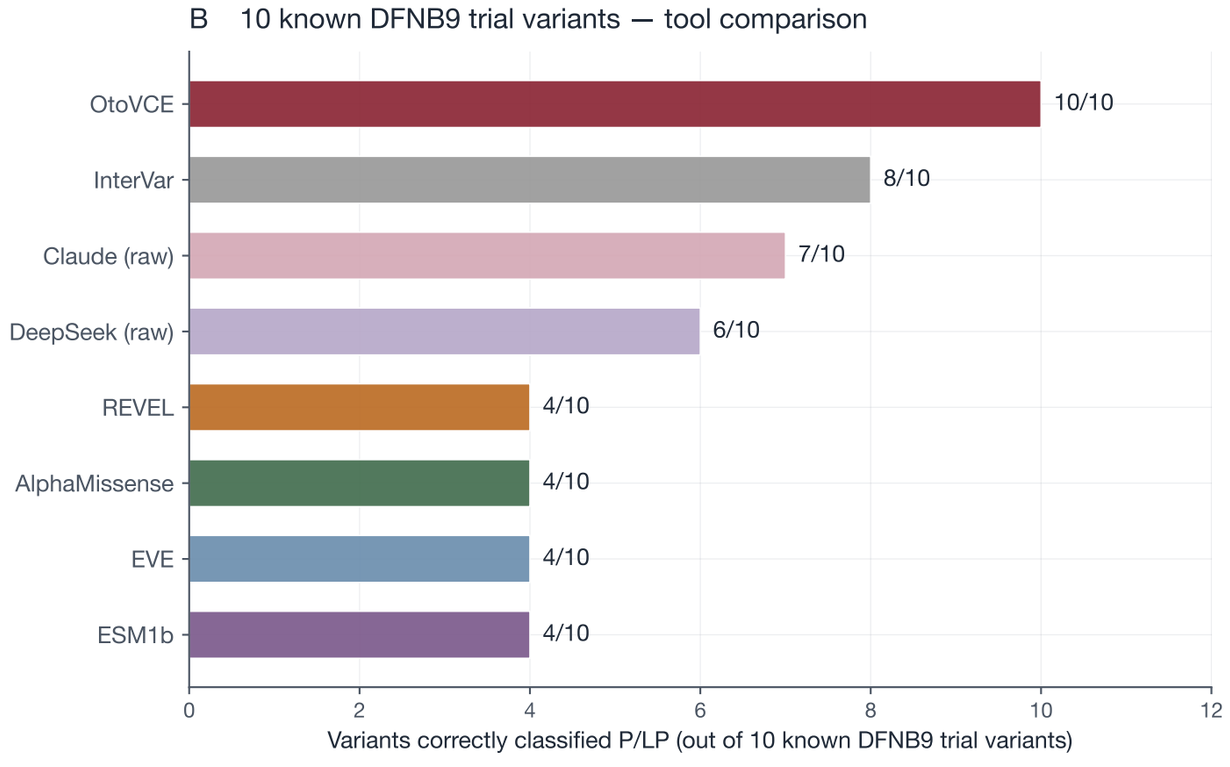
PMID 38073514 demonstrates that COL2A1 mutants (including p.Arg565Cys) form large aggregates in the ER and elevate ER stress, with wild-type negative control — qualitative loss-of-function in a validated assay (PS3 moderate). For PS4, the variant is reported across 4 PMIDs as a recurrent allele in 11 unrelated Stickler probands.

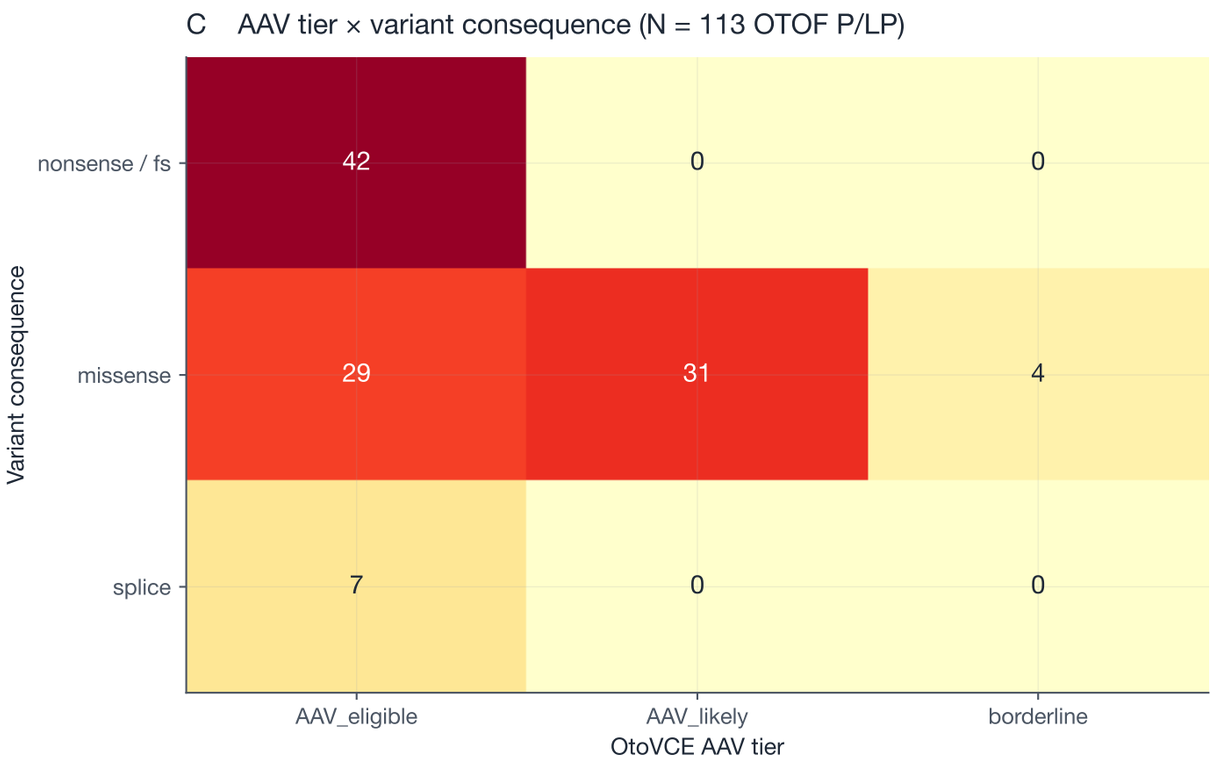
Generic (PS3 = none, PS4 = none; 0 PMIDs):

No retrieved literature provides functional data or proband counts for the specific variant COL2A1 p.Arg565Cys. The papers focus on Stickler syndrome diagnosis and other COL2A1 mutations.

Supplementary Figure S5. AAV-OTOF deployment case study (Chapter D)







D Illustrative AAV-OTOF case (OTOF p.Pro490Arg, recurrent in clinical trials)

Step 2 — rules engine: VUS (net 3 pts)

Step 3 — multi-agent debate: Likely pathogenic (moderate confidence)

Rationale (excerpt):
C2-3 calcium-binding fold; Pro-to-Arg substitution introduces charge and disrupts the beta-sandwich, mechanistically aligned with otoferlin loss-of-function. Computationally convergent (AlphaMissense 0.979, EVE 0.842, ESM1b strong, HL-PLM 0.979 with LoF as top mechanism head), AF functionally absent in gnomAD.

AAV-OTOF eligibility:
In trans with a second pathogenic allele, this variant supports trial enrollment as biallelic pathogenic genotype.

Supplementary Figure S6. val_dvd_full expansion, label-noise growth, precision-recall curves, and classification confusion