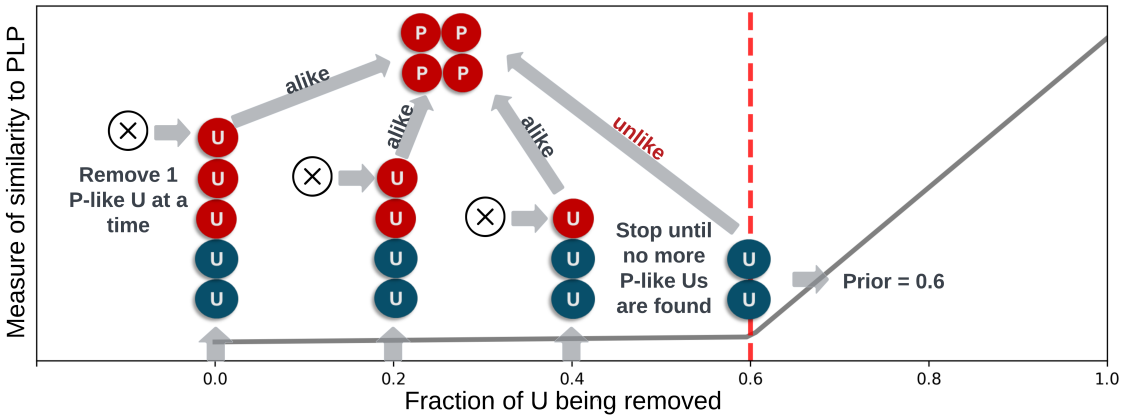
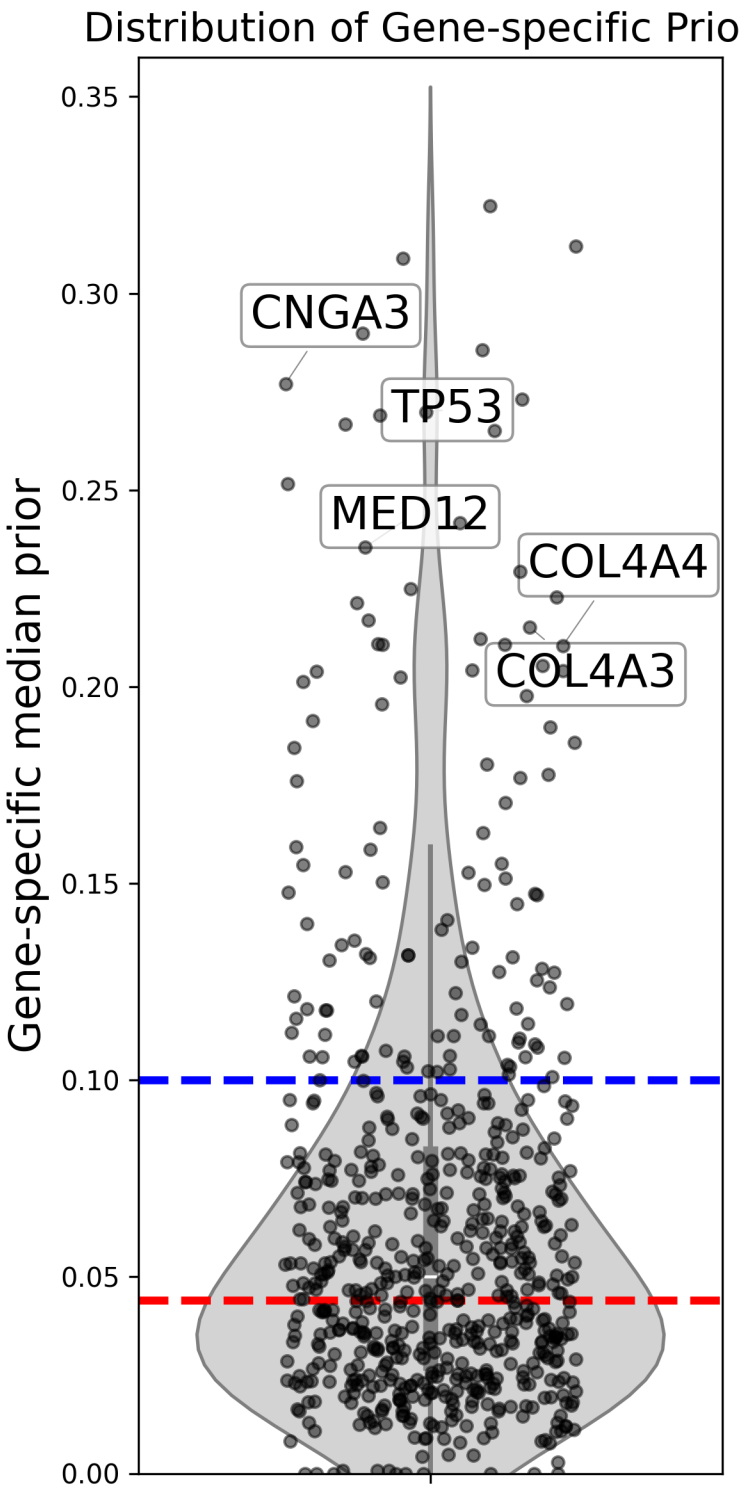
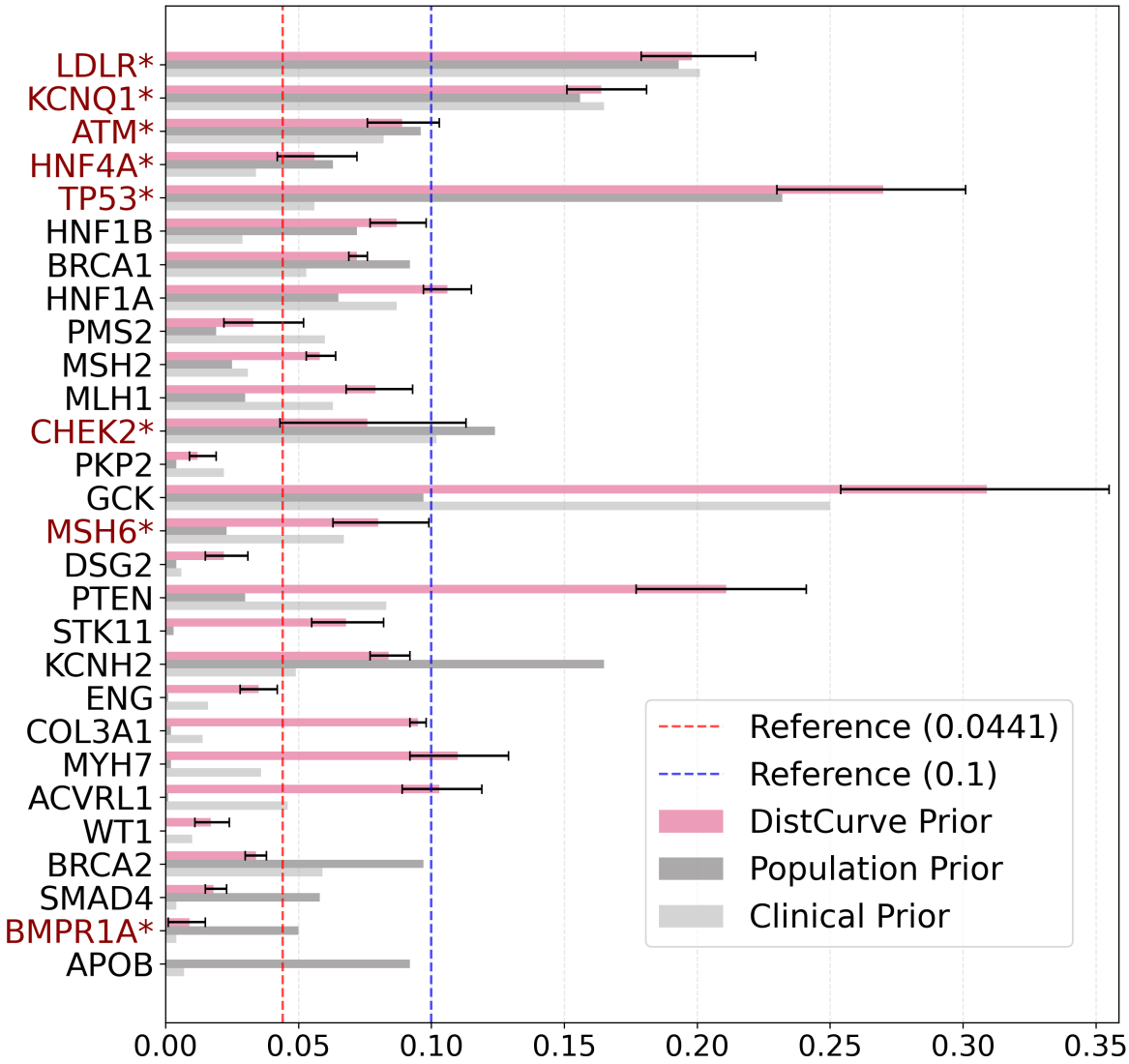
List of Extended Data Figures



Extended Data Fig. 1: DistCurve algorithm schematic
Schematic of the DistCurve algorithm for estimating gene-specific pathogenicity priors. The algorithm iteratively identifies ClinVar pathogenic-like (P-like) variants from gnomAD unlabeled variants (U), removing them at each step. Distance between remaining P (ClinVar pathogenic, red circles) and U (gnomAD, blue/red circles) sets is computed iteratively. Red U circles represent P-like samples identified during the process. The algorithm terminates when no more P-like samples can be identified, indicated by the vertical dashed red line, corresponding to the turning point in the distance curve which estimates the prior.

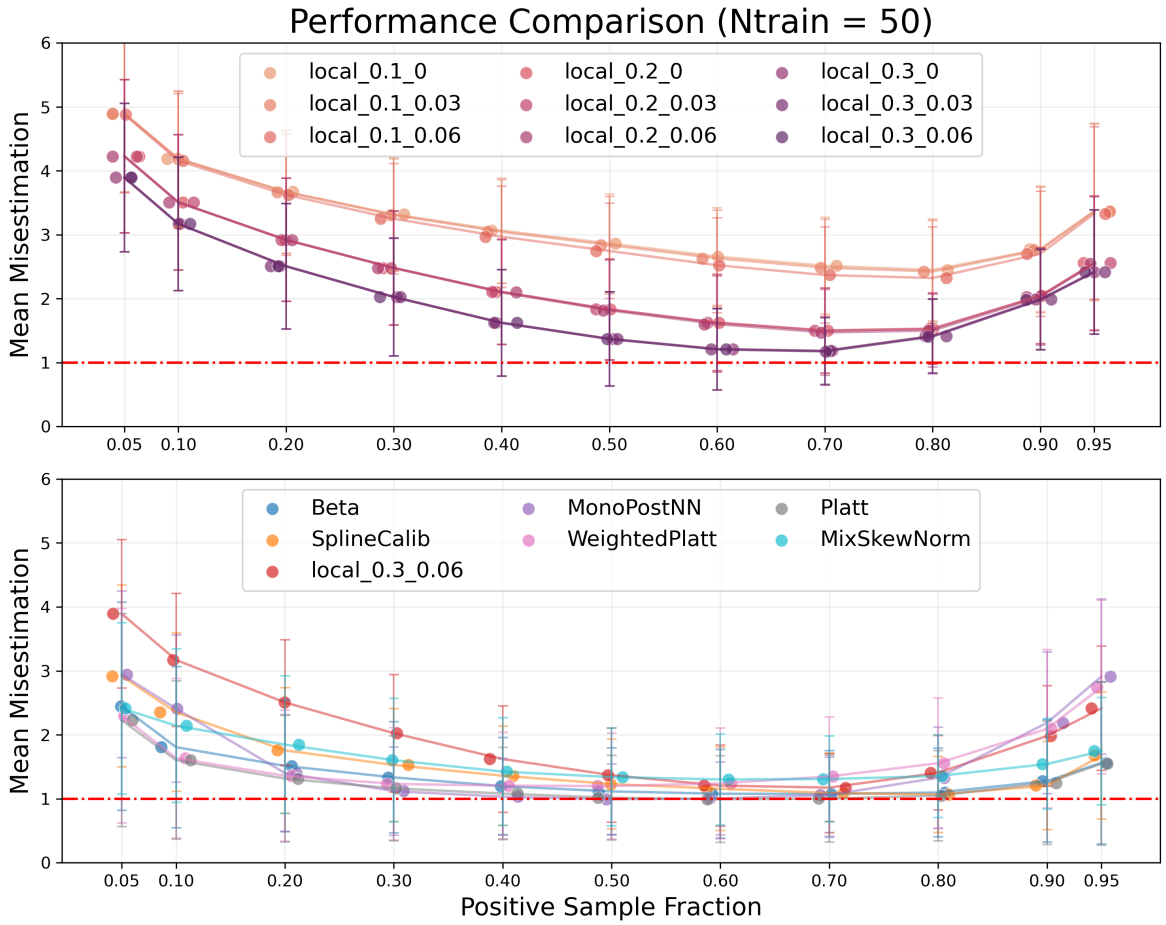


Extended Data Fig. 2: DistCurve-estimated gene-specific pathogenicity priors Distribution of gene-specific prior probabilities of pathogenicity estimated using the DistCurve algorithm across analyzed genes. Each point represents the estimated prior for a single gene, demonstrating substantial heterogeneity relative to commonly used fixed genome-wide priors. The red dashed line indicates the 4.41% prior reported by Pejaver *et al.*, whereas the blue dashed line indicates the 10% prior proposed by Tavtigian *et al.* The five genes with the highest estimated priors are labeled.

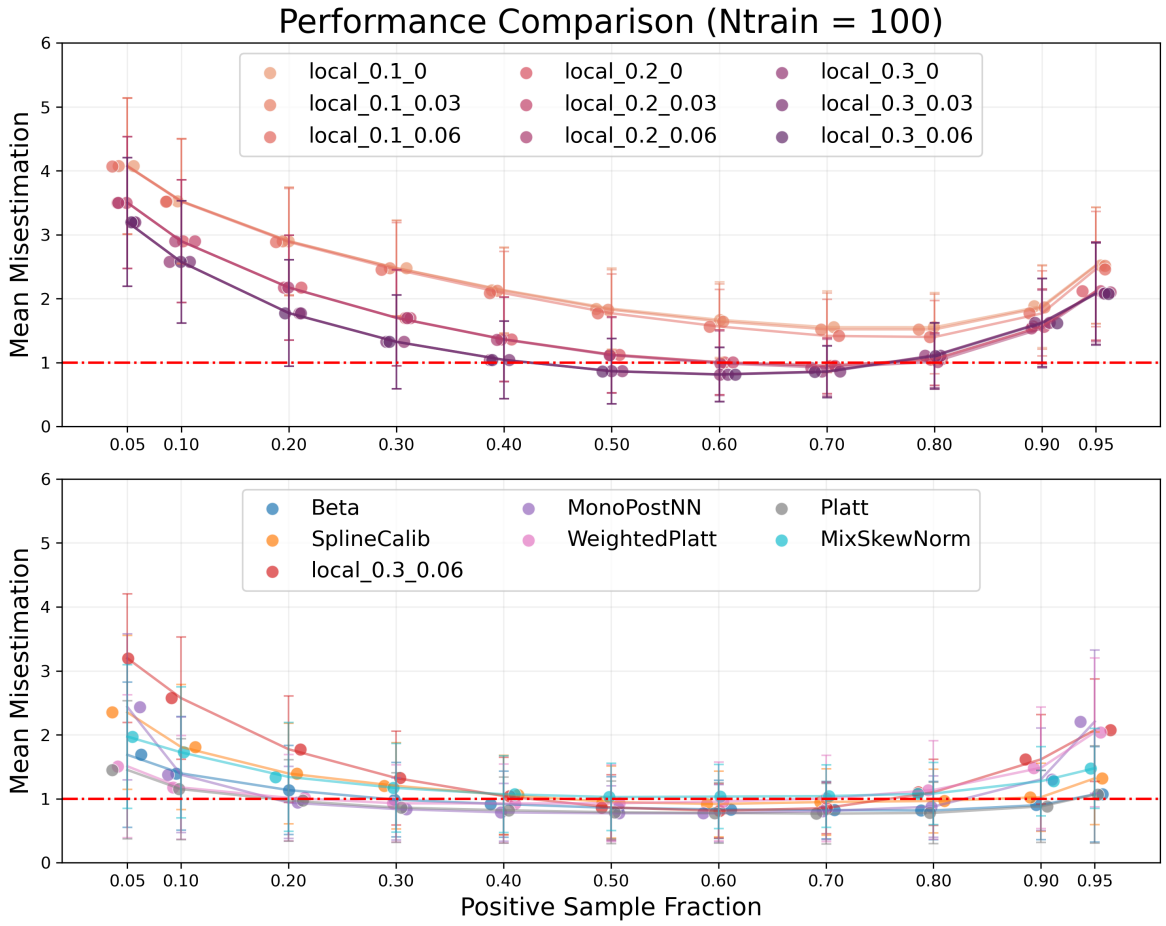


Extended Data Fig. 3: Comparison of gene-specific prior probabilities across data sources

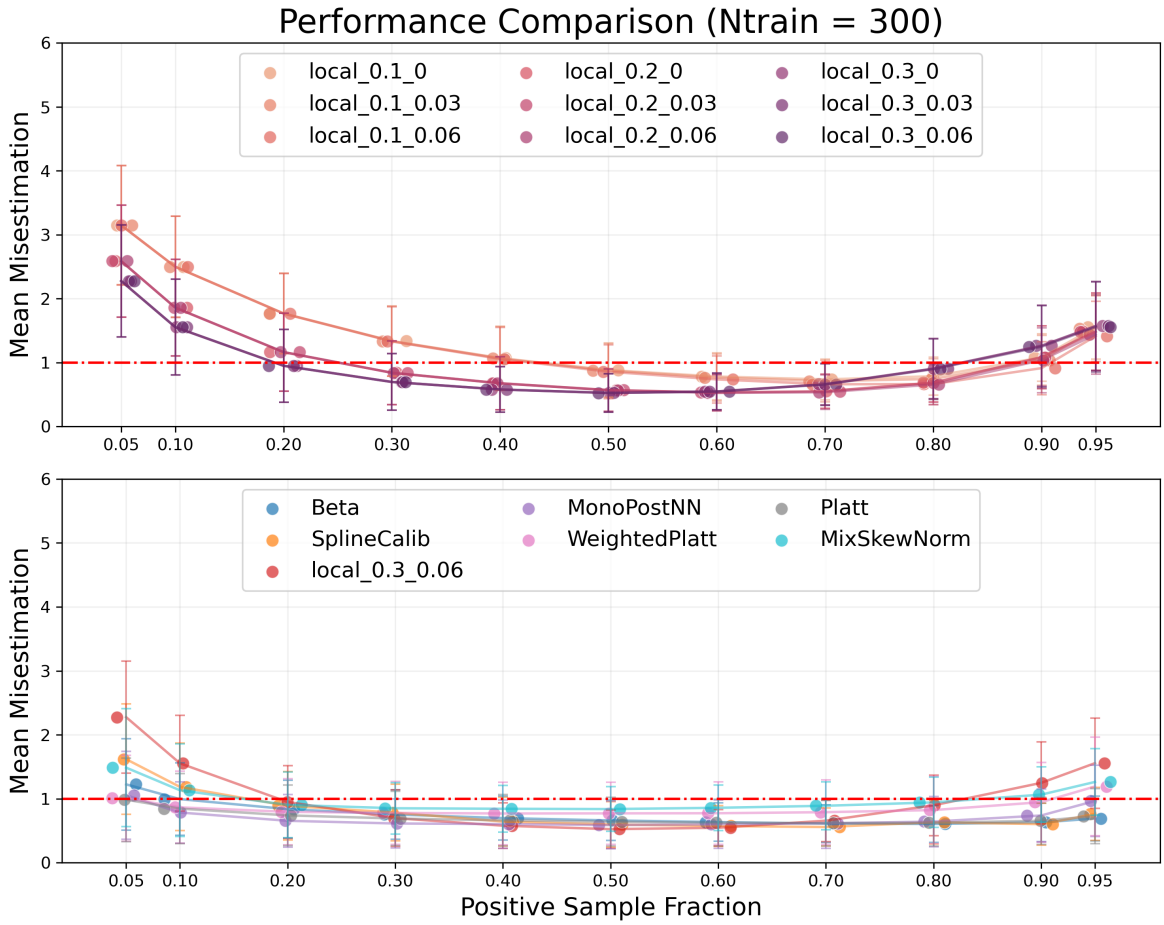
Comparison of DistCurve-derived gene-specific prior probabilities of pathogenicity (pink) with population-based priors derived from UK Biobank (Population prior; dark gray) and clinically derived priors from ClinVar (Clinical prior; light gray). Each set of bars represents one gene. Asterisks indicate genes for which the DistCurve estimate falls within the interquartile range (Q1–Q3) of either the population or clinical prior distribution. Horizontal dashed lines denote previously reported reference priors: 4.41% (red dashed line; Pejaver *et al.*) and 10% (blue dashed line; Tavtigian *et al.*).



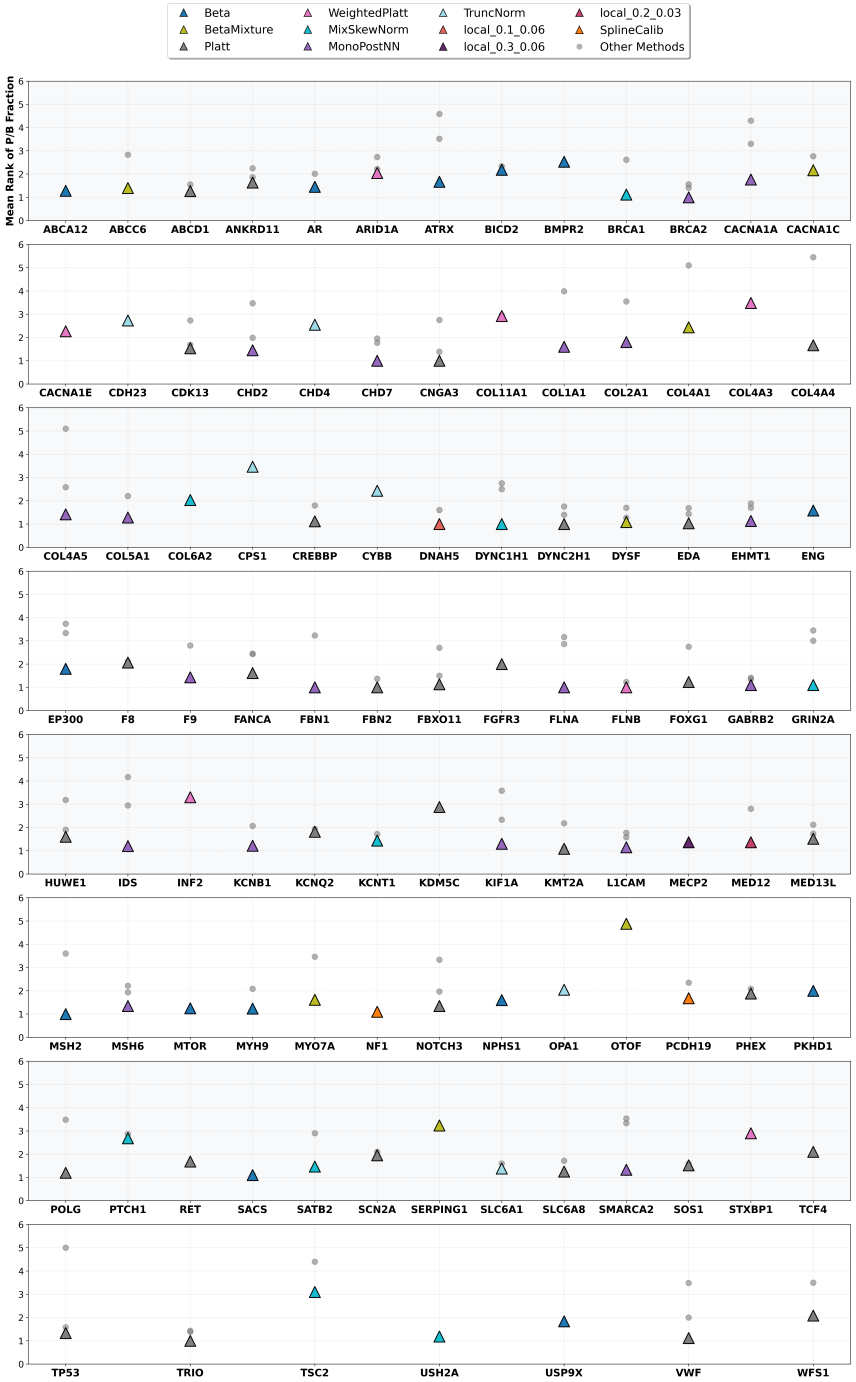
Extended Data Fig. 4: Calibration performance comparison with 50 variants
Performance comparison of calibration methods with calibration set size = 50 variants. Results are evaluated on test sets of 1,000 variants under varying class balance conditions. Error bars denote standard deviations across simulated true prior probabilities. Only methods achieving average misestimation below 1 in at least one scenario are shown.



Extended Data Fig. 5: Calibration performance comparison with 100 variants
Performance comparison of calibration methods with calibration set size = 100 variants. Results are evaluated on test sets of 1,000 variants under varying class balance conditions. Error bars denote standard deviations across simulated true prior probabilities. Only methods achieving average misestimation below 1 in at least one scenario are shown.

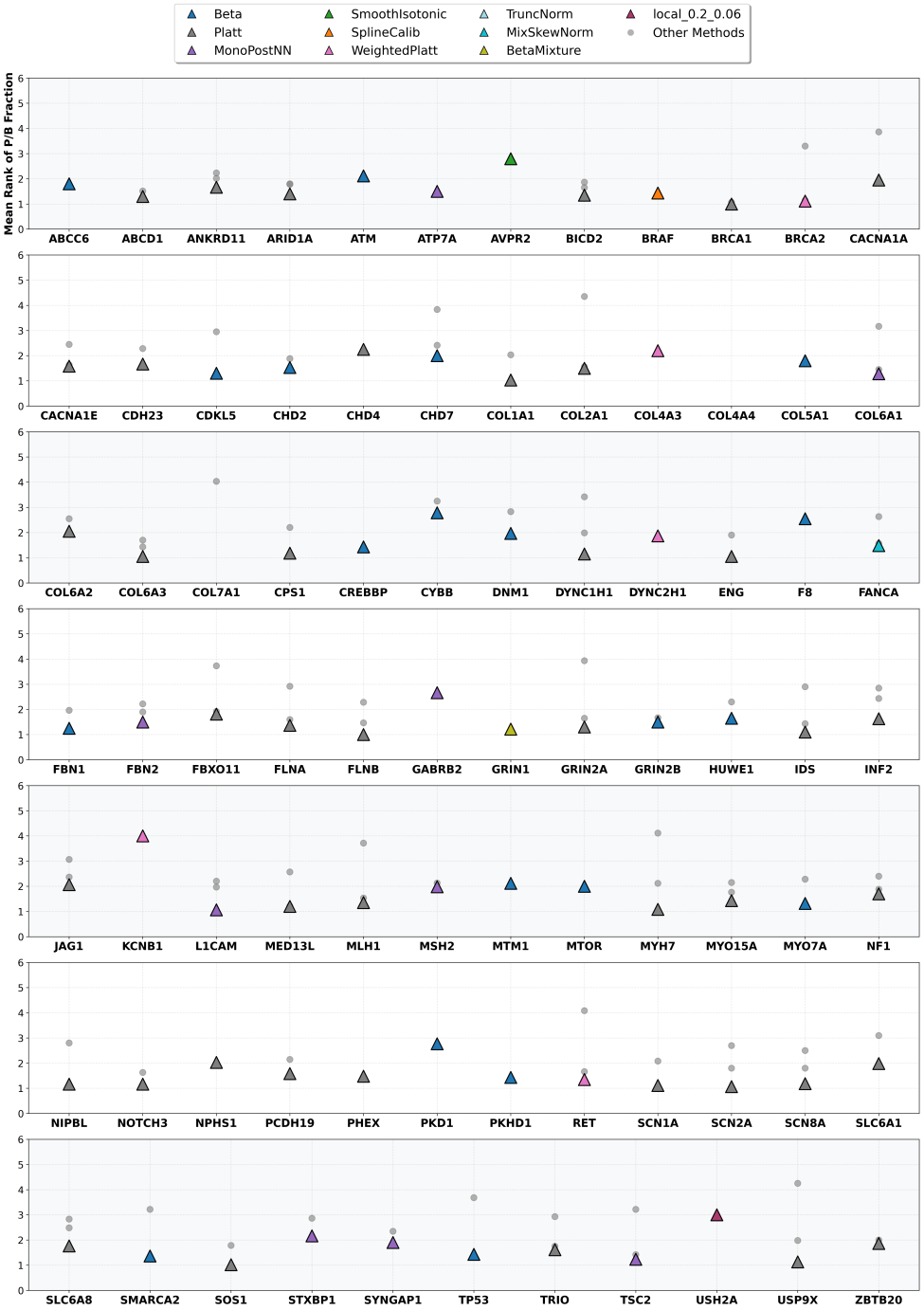


Extended Data Fig. 6: Calibration performance comparison with 300 variants
Performance comparison of calibration methods with calibration set size = 300 variants. Results are evaluated on test sets of 1,000 variants under varying class balance conditions. Error bars denote standard deviations across simulated true prior probabilities. Only methods achieving average misestimation below 1 in at least one scenario are shown.



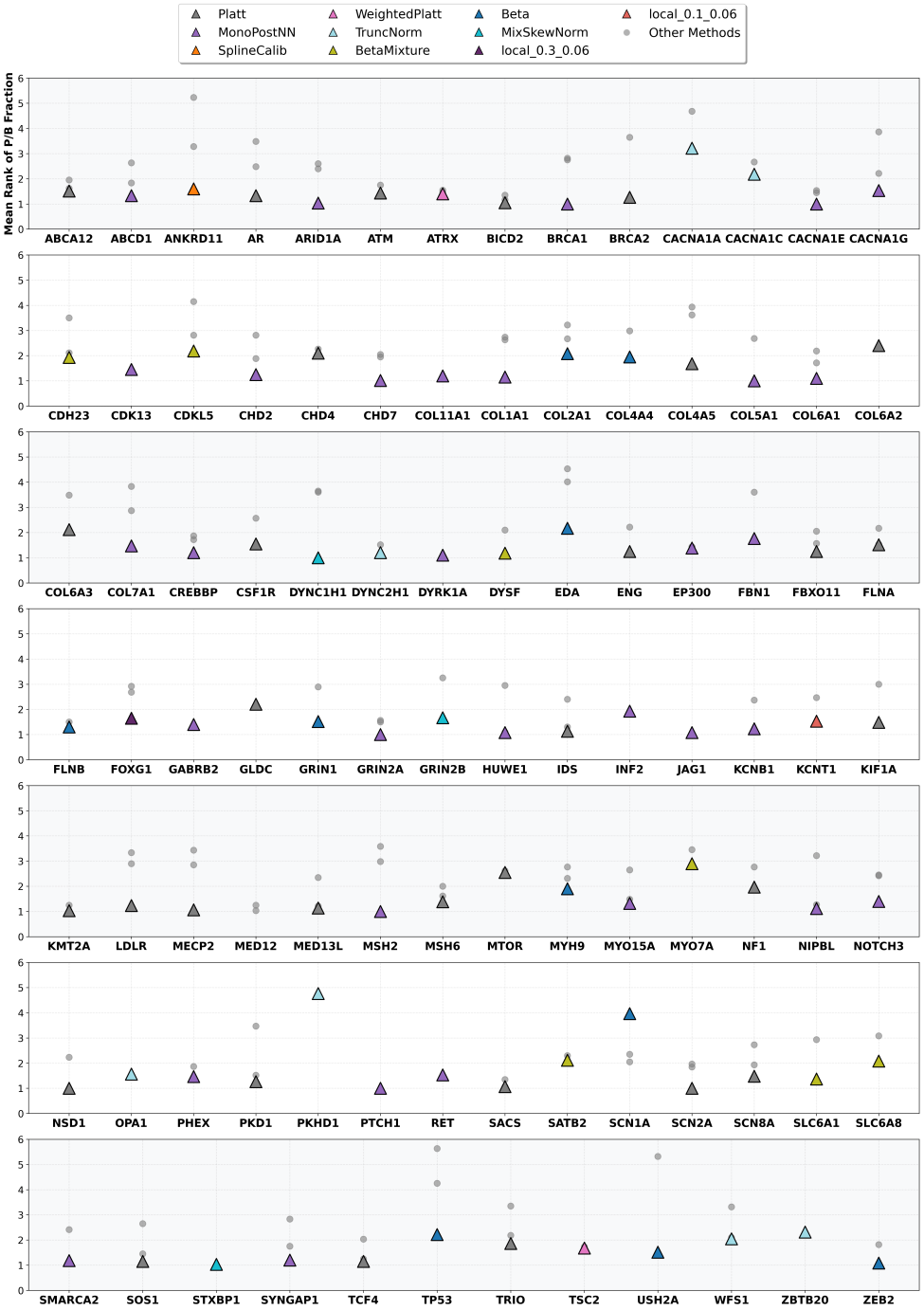
Extended Data Fig. 7: Mean rank by $P/B_{Fraction}$ across calibration methods per gene (REVEL)

Mean rank of the $P_{Fraction}$ and $B_{Fraction}$ metrics across calibration methods filtered by the first 2 steps of the Multi-stage method selection (Methods). Using the simulated score sets, a single calibration method was selected for each gene through a three-stage filtering procedure across 30 simulation datasets; only valid methods are retained in the figure. This is using REVEL score. For each gene, the final selected optimal calibration method is indicated by a colored triangle (with color corresponding to the method), while all other methods are shown as grey circles. Points represent 101 genes evaluated for REVEL.



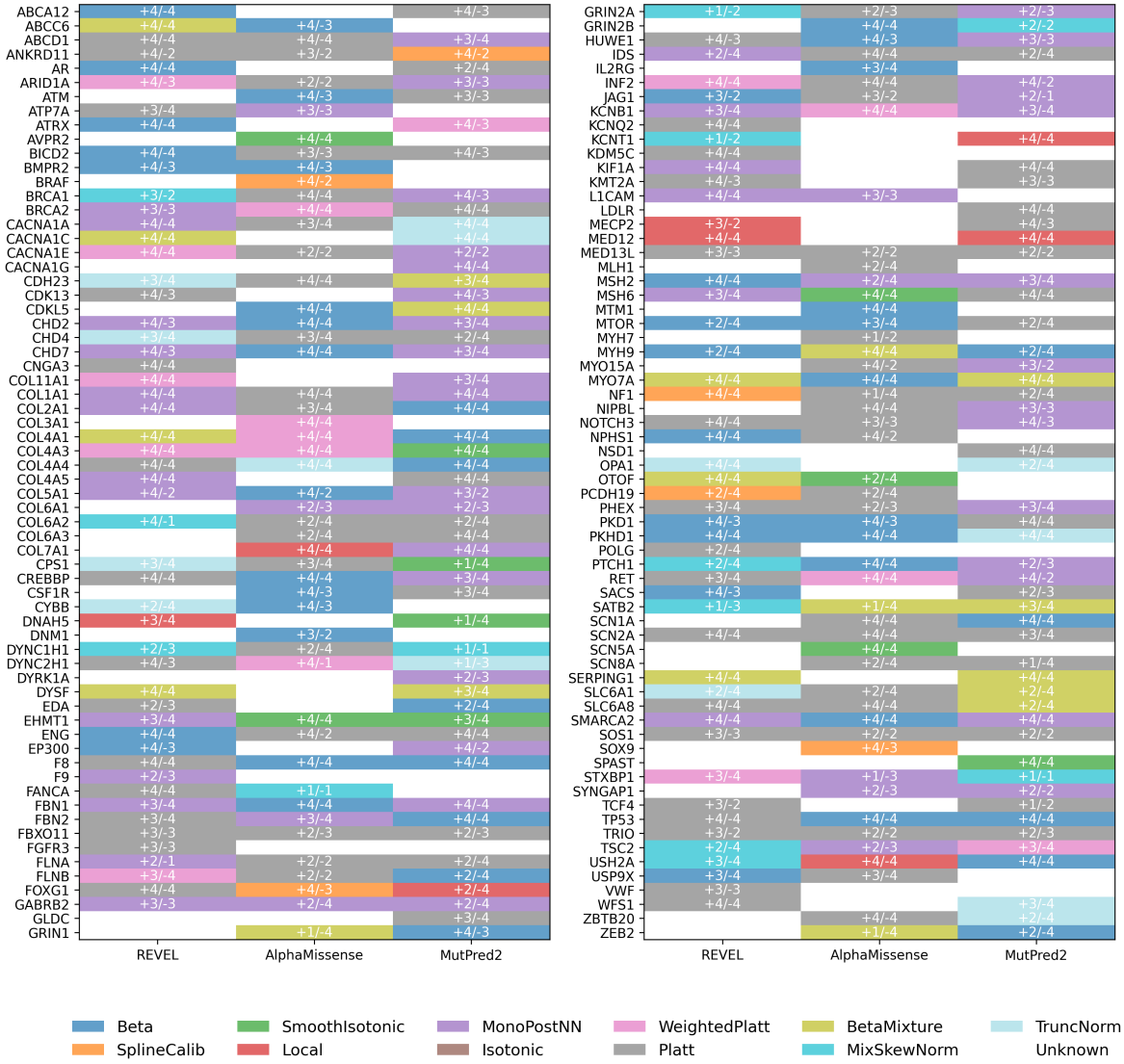
Extended Data Fig. 8: Mean rank by $P/B_{Fraction}$ across calibration methods per gene (AlphaMissense)

Mean rank of the $P_{Fraction}$ and $B_{Fraction}$ metrics across calibration methods filtered by the first 2 steps of the Multi-stage method selection (Methods). Using the simulated score sets, a single calibration method was selected for each gene through a three-stage filtering procedure across 30 simulation datasets; only valid methods are retained in the figure. This is using AlphaMissense score. For each gene, the final selected optimal calibration method is indicated by a colored triangle (with color corresponding to the method), while all other methods are shown as grey circles. Points represent 98 genes evaluated for AlphaMissense.



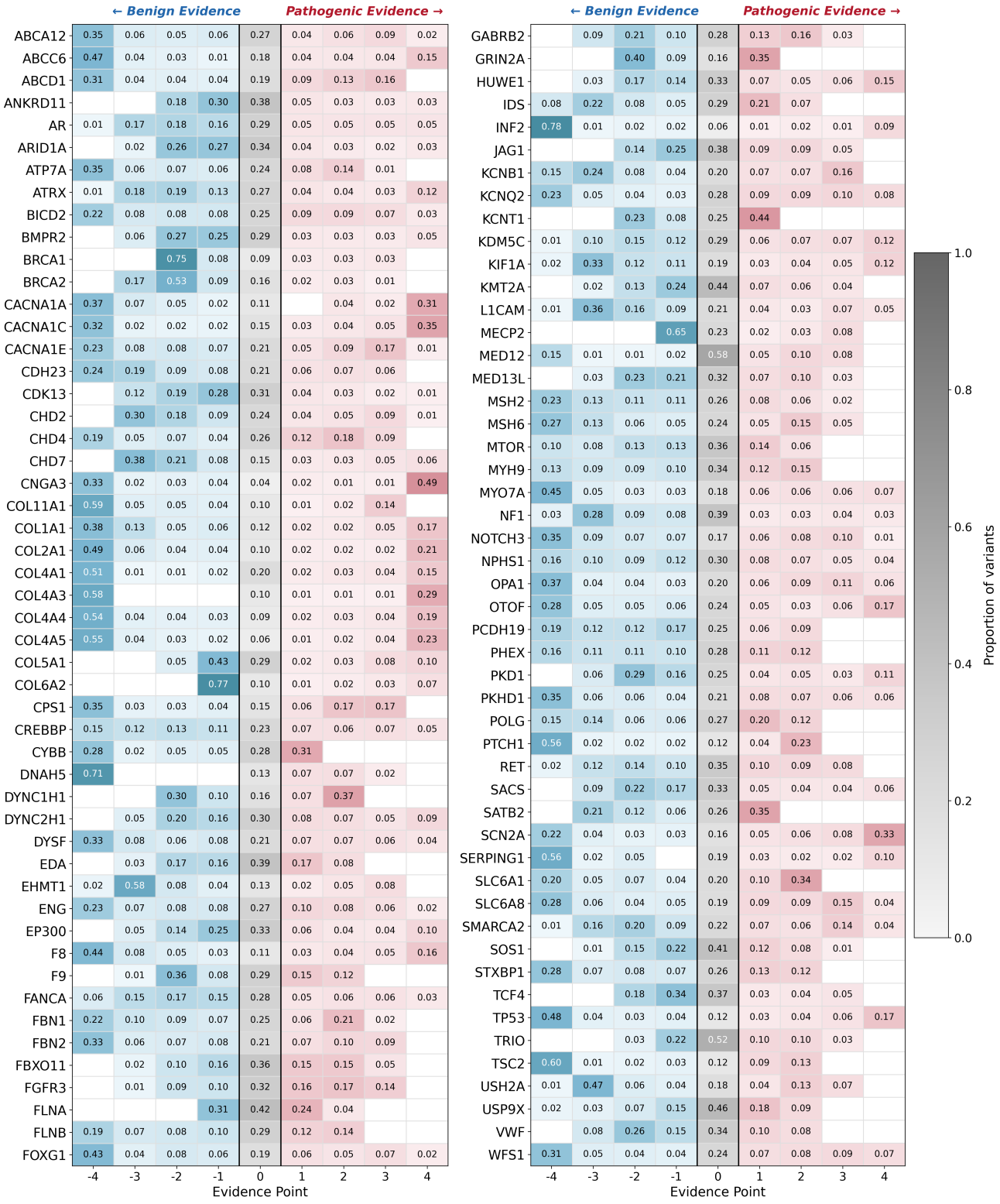
Extended Data Fig. 9: Mean rank by $P/B_{Fraction}$ across calibration methods per gene (MutPred2)

Mean rank of the $P_{Fraction}$ and $B_{Fraction}$ metrics across calibration methods filtered by the first 2 steps of the Multi-stage method selection (Methods). Using the simulated score sets, a single calibration method was selected for each gene through a three-stage filtering procedure across 30 simulation datasets; only valid methods are retained in the figure. This is using MutPred2 score. For each gene, the final selected optimal calibration method is indicated by a colored triangle (with color corresponding to the method), while all other methods are shown as grey circles. Points represent 105 genes evaluated for MutPred2.

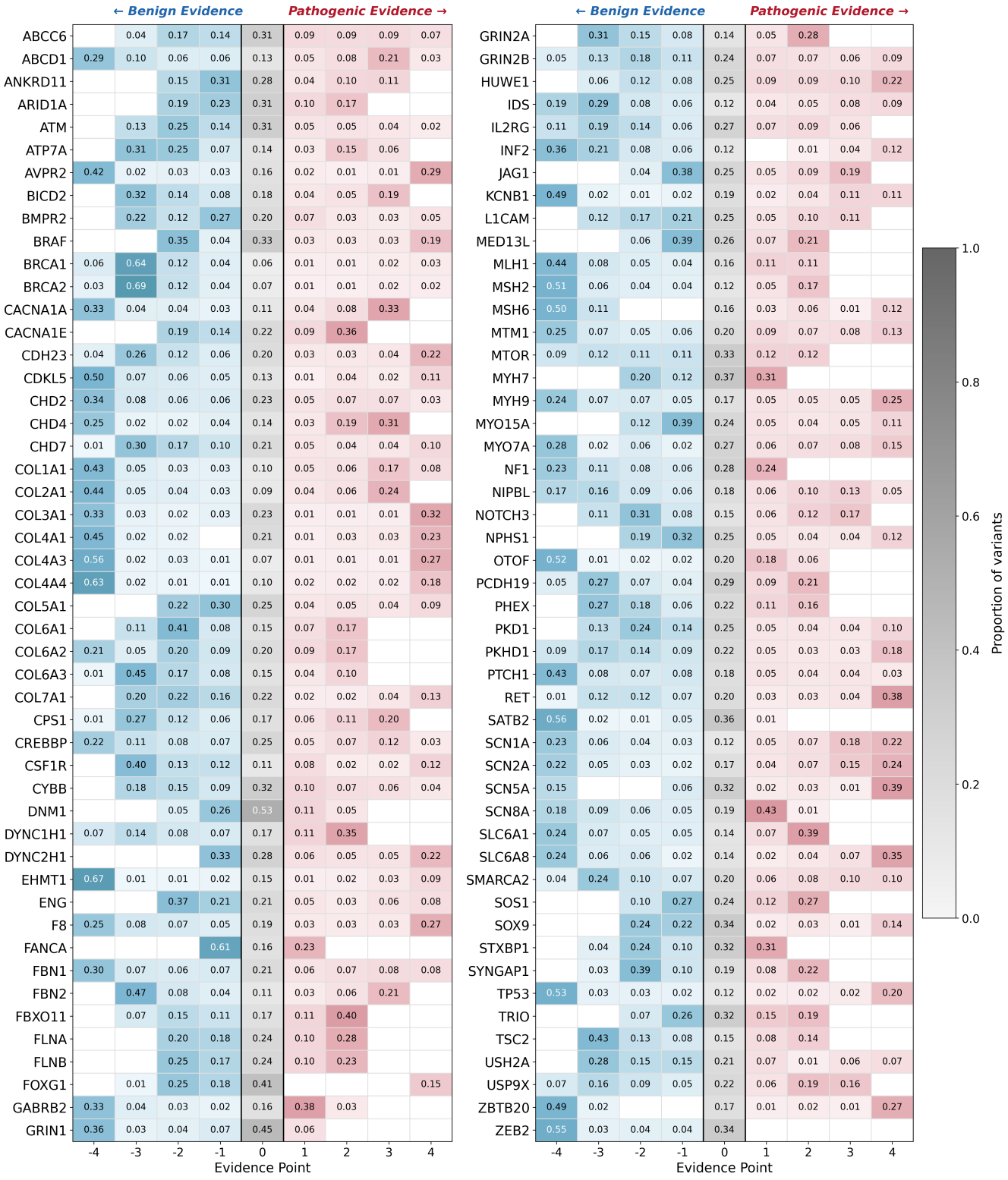


Extended Data Fig. 10: Best calibration method selected per gene across REVEL, AlphaMissense, and MutPred2

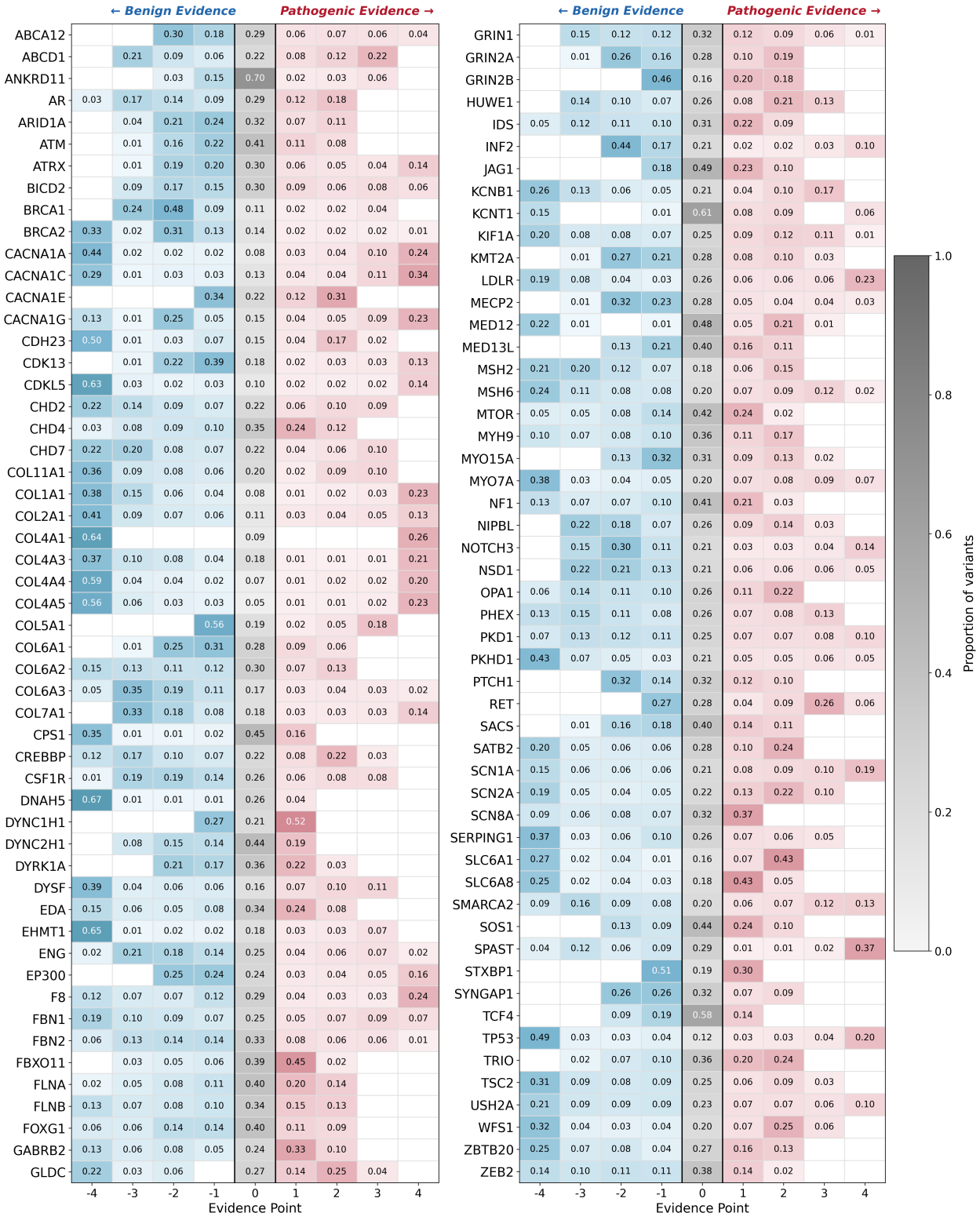
Best calibration method selected for each predictor across genes. Heatmap showing the optimal calibration method selected for each gene and predictor (REVEL, AlphaMissense and MutPred2). Rows represent genes (n = 132), sorted by gene name, and columns represent predictors. Colors denote the selected calibration method, and overlaid text indicates the corresponding benign (negative)/pathogenic (positive) evidence point assigned after calibration. Genes without a selected method are shown in white (Unknown). Across genes, method selection varied by predictor. For REVEL, Platt (32 genes), MonoPostNN (19 genes) and Beta (16 genes) were most frequently selected. For AlphaMissense, Platt was most common (42 genes), followed by Beta (24 genes), with 34 genes labeled Unknown. For MutPred2, MonoPostNN (32 genes) and Platt (31 genes) were most frequently selected, with 27 genes labeled Unknown.



Extended Data Fig. 11: Evidence point assignment heatmap (REVEL)
Gene-specific evidence point assignment proportions using REVEL scores. Each row represents a gene; each column represents an evidence level. Color intensity indicates variant proportion at each level: blue shades (-1 to -4) represent benign evidence; red shades (+1 to +4) represent pathogenic evidence; white (0) indicates indeterminate evidence. Blank cells denote proportions rounded to 0.00.

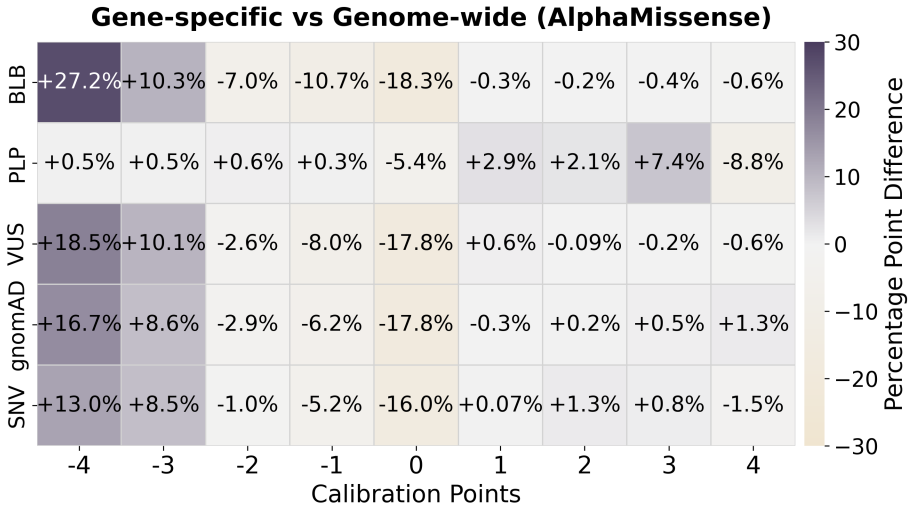


Extended Data Fig. 12: Evidence point assignment heatmap (AlphaMissense) Gene-specific evidence point assignment proportions using AlphaMissense scores. Each row represents a gene; each column represents an evidence level. Color intensity indicates variant proportion at each level: blue shades (-1 to -4) represent benign evidence; red shades (+1 to +4) represent pathogenic evidence; white (0) indicates indeterminate evidence. Blank cells denote proportions rounded to 0.00.

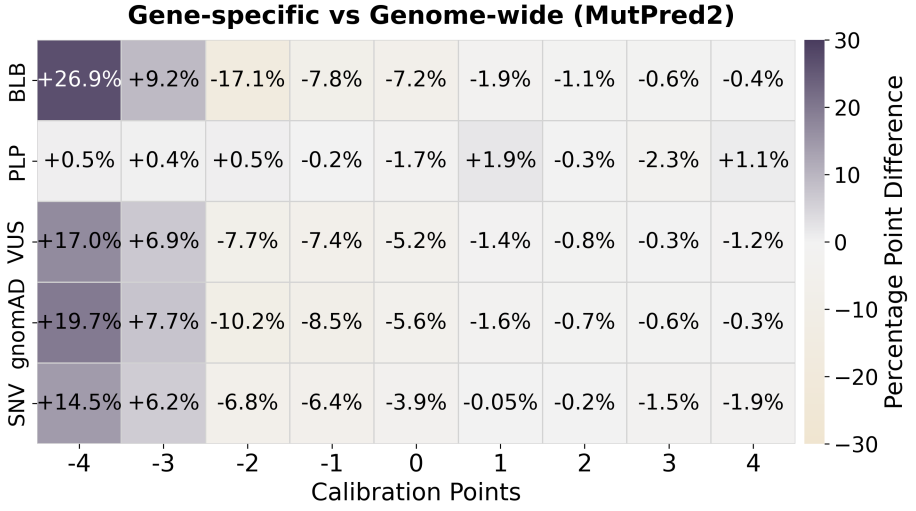


Extended Data Fig. 13: Evidence point assignment heatmap (MutPred2)
Gene-specific evidence point assignment proportions using MutPred2 scores. Each row represents a gene; each column represents an evidence level. Color intensity indicates variant proportion at each level: blue shades (-1 to -4) represent benign evidence; red shades (+1 to +4) represent pathogenic evidence; white (0) indicates indeterminate evidence. Blank cells denote proportions rounded to 0.00.

a AlphaMissense



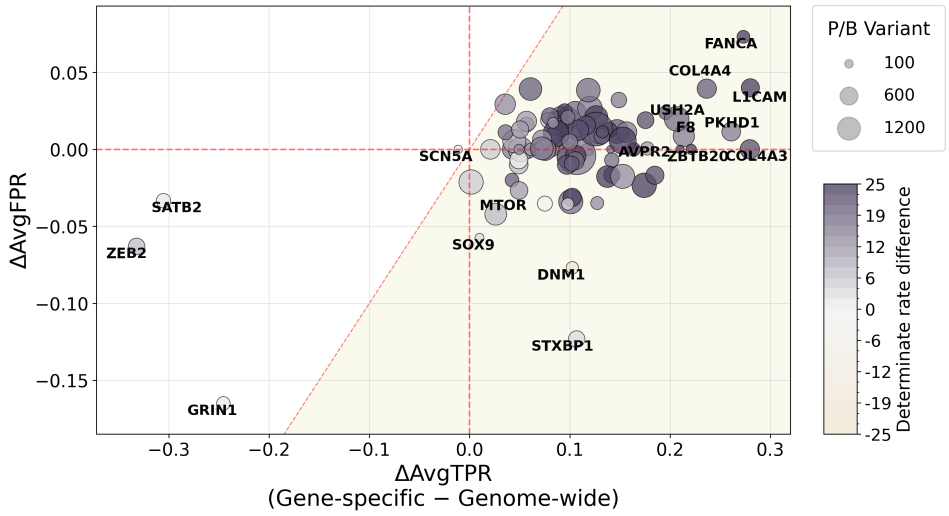
b MutPred2



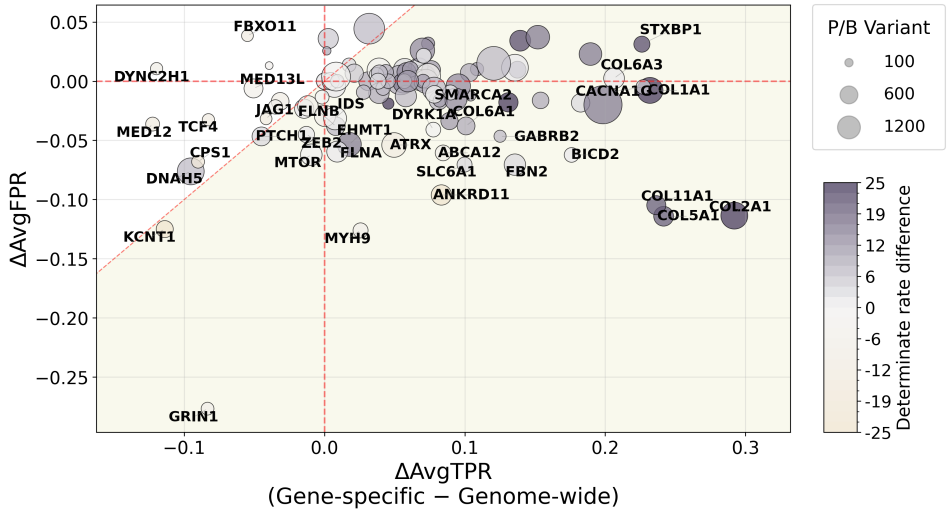
Extended Data Fig. 14: Evidence point assignment difference (gene-specific vs. genome-wide): AlphaMissense and MutPred2

Comparison of evidence point assignments between gene-specific and genome-wide calibration methods. (a) Results using AlphaMissense scores. (b) Results using MutPred2 scores. Heatmaps display the differences in the percentage of evidence point assignments stratified by variant classification or source. Grey indicates increased assignment rates by the gene-specific calibration method compared to genome-wide calibration.

a AlphaMissense

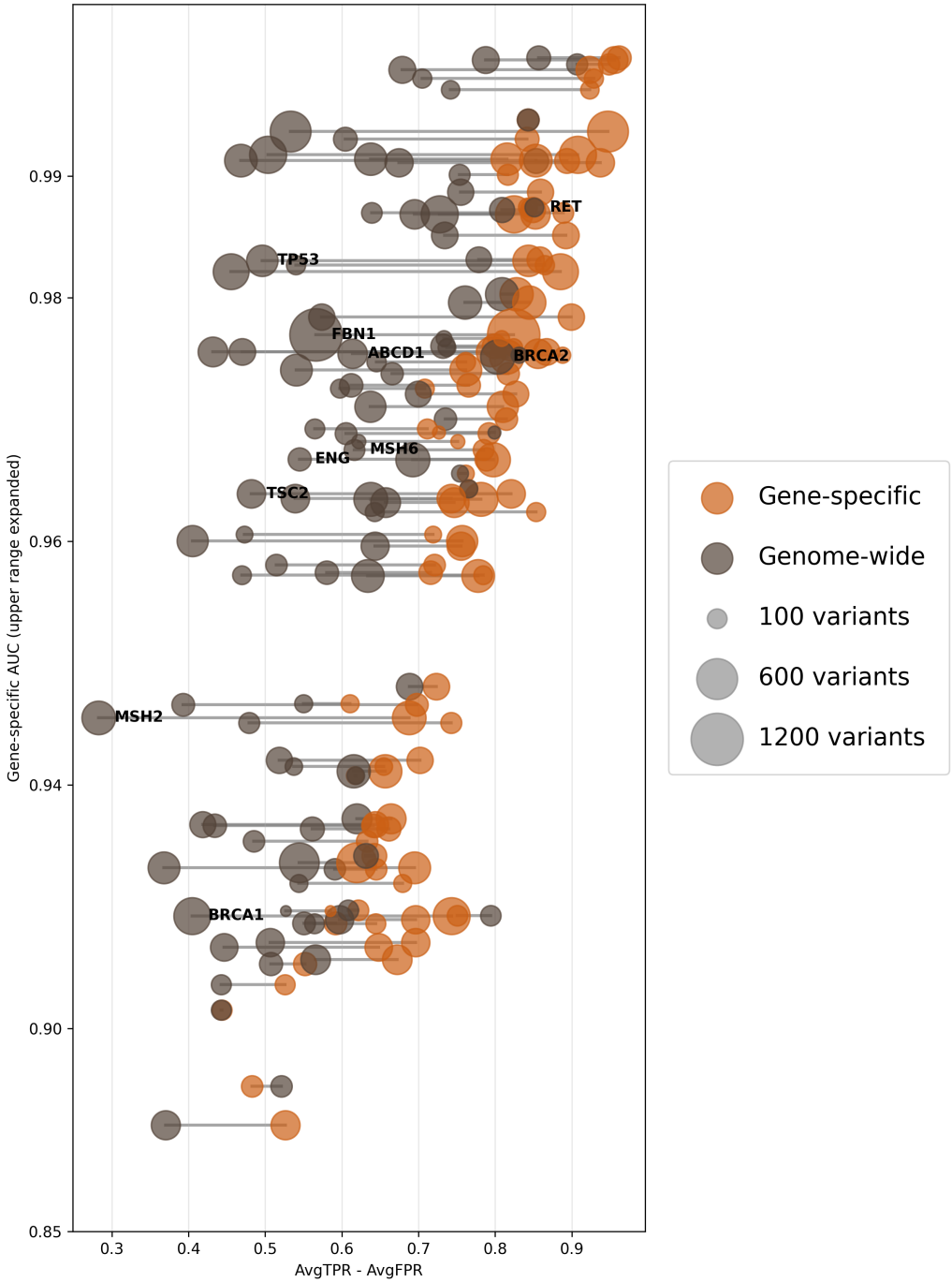


b MutPred2

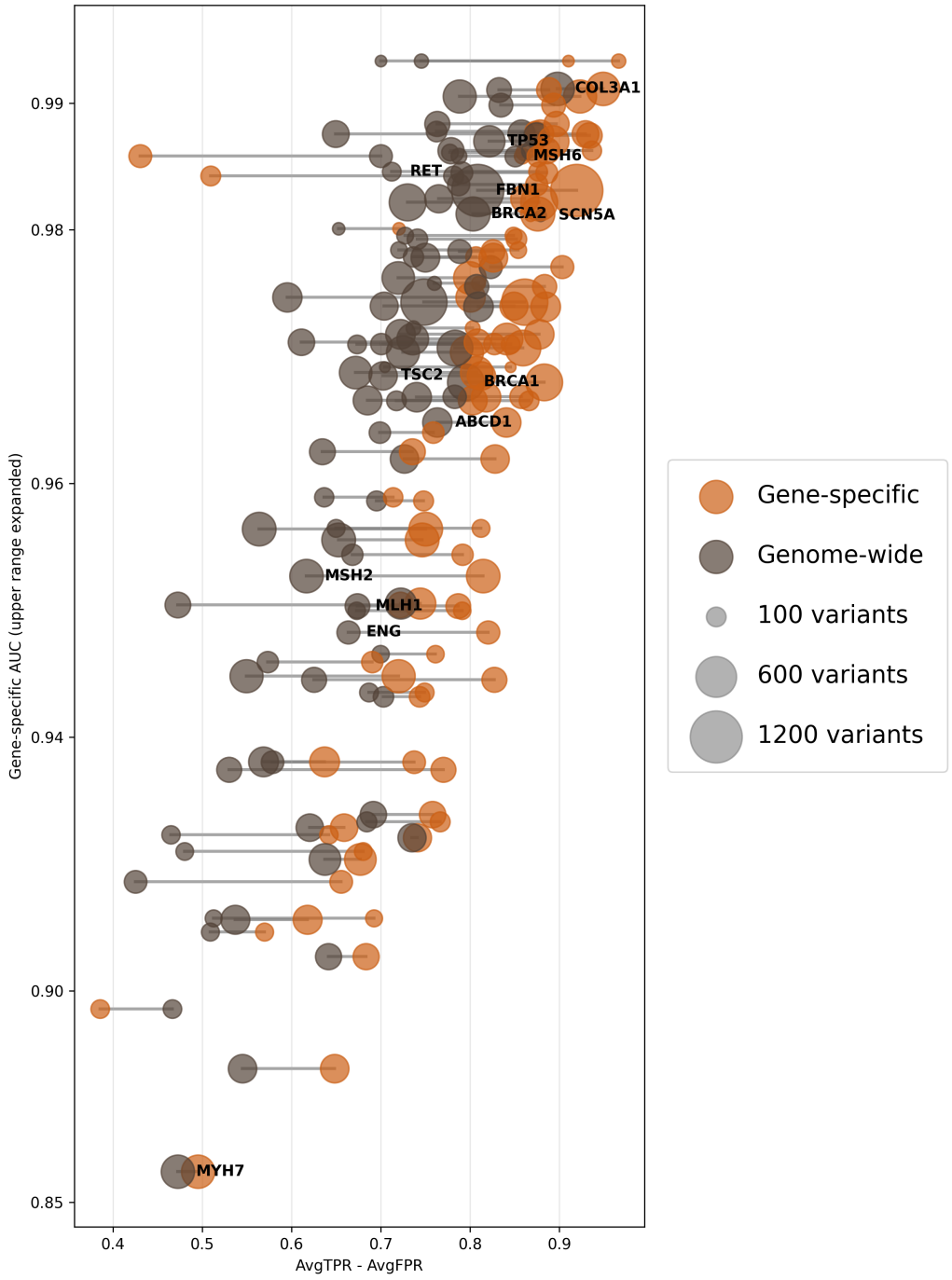


Extended Data Fig. 15: Calibration method comparison: AlphaMissense and MutPred2 (per-gene)

Gene-level changes in sensitivity and false positive rate following gene-specific calibration. (a) Results using AlphaMissense scores. (b) Results using MutPred2 scores. Each point represents a gene. The x-axis shows the average change in true positive rate (ΔAvgTPR) between gene-specific and genome-wide calibration (gene-specific minus aggregated), and the y-axis shows the corresponding change in false positive rate (ΔAvgFPR). Point color encodes the change in evidence coverage (fraction of variants receiving at least -/+1 point of evidence), with deeper grey indicating higher coverage under the gene-specific method. Point size is proportional to the number of pathogenic and benign variants (n_{PB}; P/LP and B/LB). The shaded area indicates increased performance (below the dashed red diagonal $x = y$).

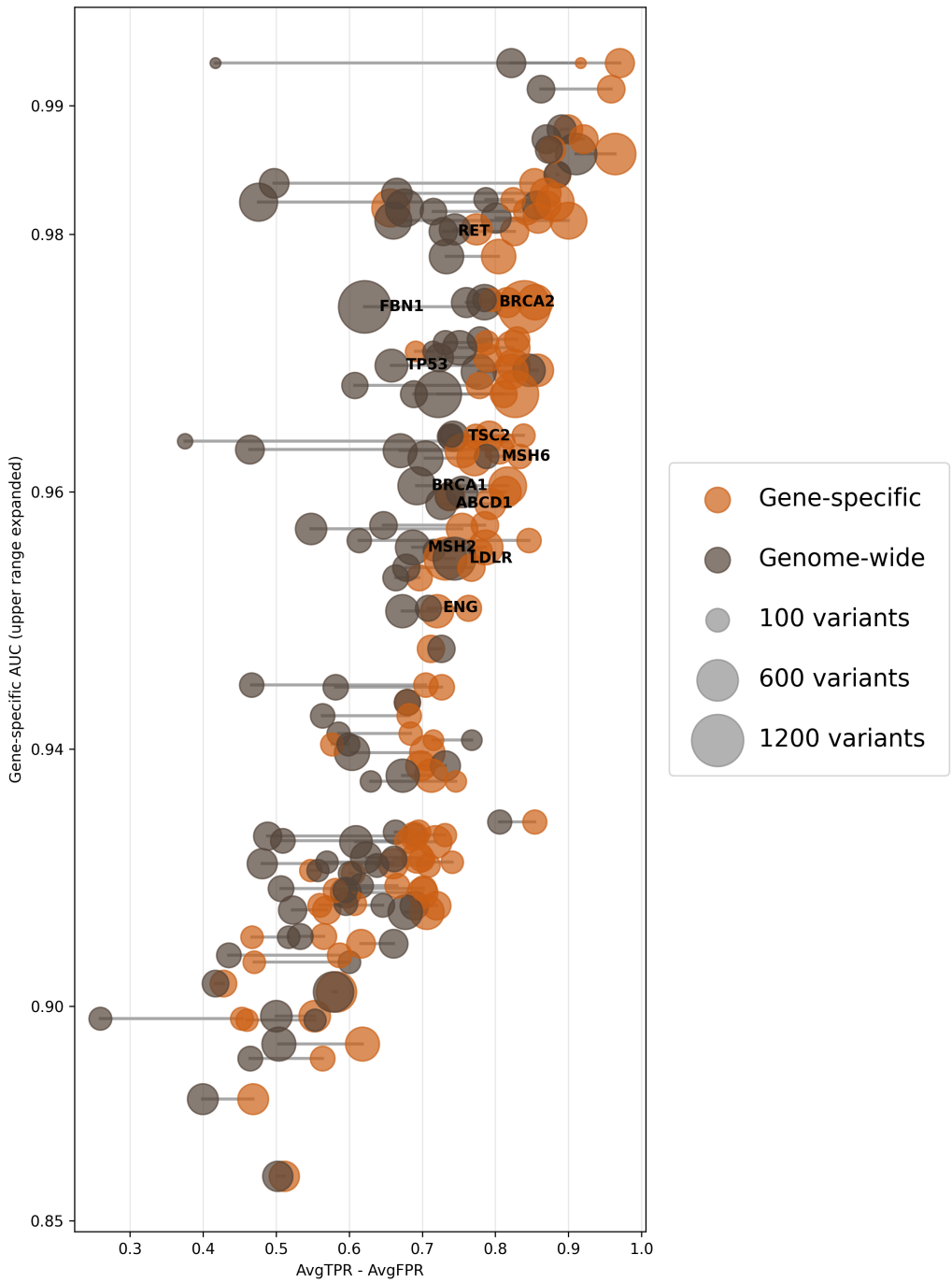


Extended Data Fig. 16: Gene-specific calibration performance and REVEL AUC Gene-specific performance (AvgTPR − AvgFPR) versus gene-specific REVEL AUC shown as a dumbbell plot. Each gene is represented by a pair of connected points (gene-specific calibration vs. genome-wide calibration). The x-axis shows the gene-specific performance metric (AvgTPR − AvgFPR), and the y-axis shows the gene-specific REVEL AUC. Genes annotated with labels correspond to ACMG secondary finding genes. Higher values of AvgTPR − AvgFPR indicate better discrimination performance. The y-axis is displayed on a non-linear scale, with the upper range expanded to improve visualization of differences across genes.



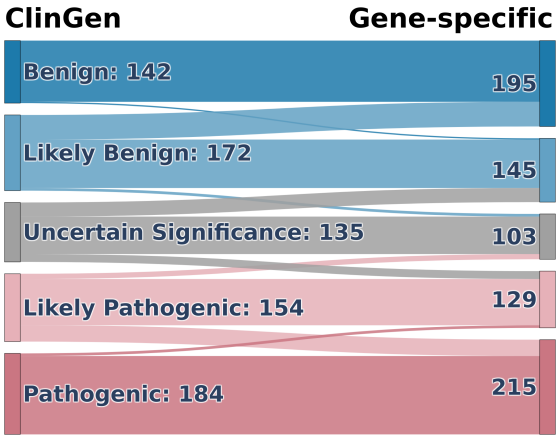
Extended Data Fig. 17: Gene-specific calibration performance and AlphaMissense AUC

Gene-specific performance (AvgTPR − AvgFPR) versus gene-specific AlphaMissense AUC shown as a dumbbell plot. Each gene is represented by a pair of connected points (gene-specific calibration vs. genome-wide calibration). The x-axis shows the gene-specific performance metric (AvgTPR − AvgFPR), and the y-axis shows the gene-specific AlphaMissense AUC. Genes annotated with labels correspond to ACMG secondary finding genes. Higher values of AvgTPR − AvgFPR indicate better discrimination performance. The y-axis is displayed on a non-linear scale, with the upper range expanded to improve visualization of differences across genes.

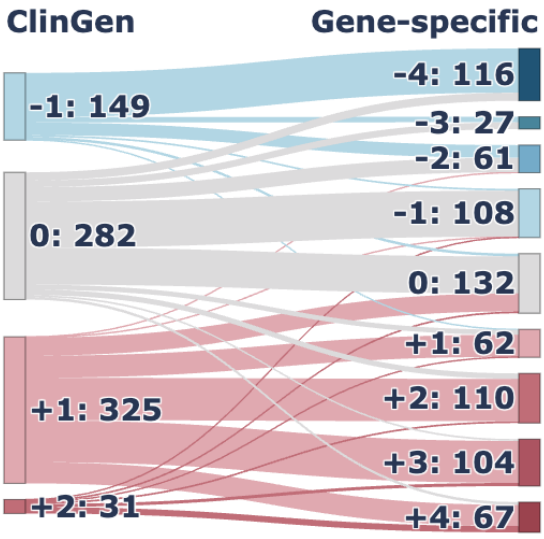


Extended Data Fig. 18: Gene-specific calibration performance and MutPred2 AUC Gene-specific performance (AvgTPR − AvgFPR) versus gene-specific MutPred2 AUC shown as a dumbbell plot. Each gene is represented by a pair of connected points (gene-specific calibration vs. genome-wide calibration). The x-axis shows the gene-specific performance metric (AvgTPR − AvgFPR), and the y-axis shows the gene-specific MutPred2 AUC. Genes annotated with labels correspond to ACMG secondary finding genes. Higher values of AvgTPR − AvgFPR indicate better discrimination performance. The y-axis is displayed on a non-linear scale, with the upper range expanded to improve visualization of differences across genes.

a ClinGen classification changes

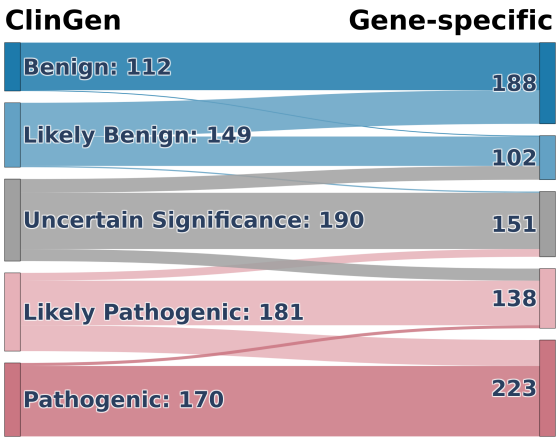


b ClinGen evidence point changes

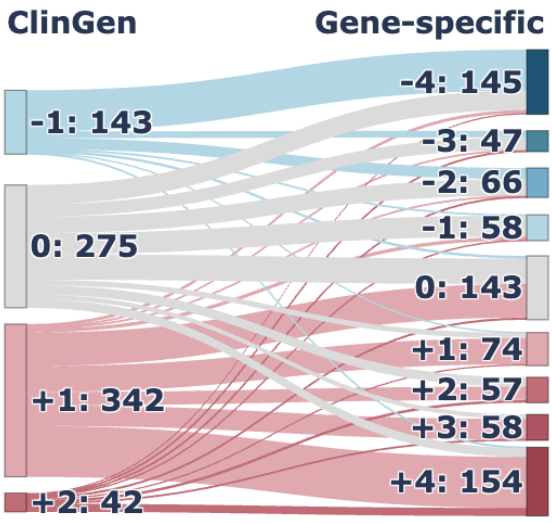


Extended Data Fig. 19: ClinGen Sankey by gene-specific calibration (REVEL) Sankey diagrams illustrating the impact of gene-specific calibration using REVEL scores. (a) Transitions in clinical classifications from original ClinGen assignments to recalibrated classifications. (b) Reassignment of computational evidence points under ClinGen guidelines compared with gene-specific calibration.

a ClinGen classification changes

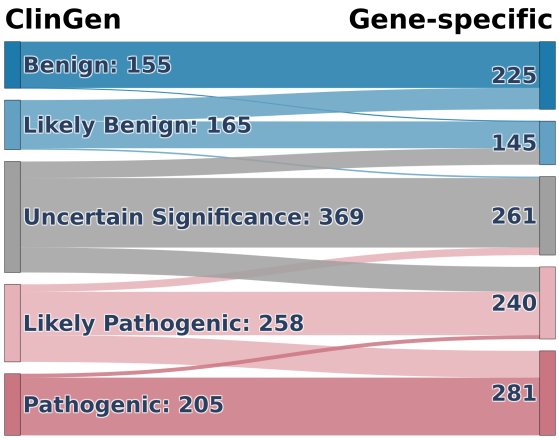


b ClinGen evidence point changes

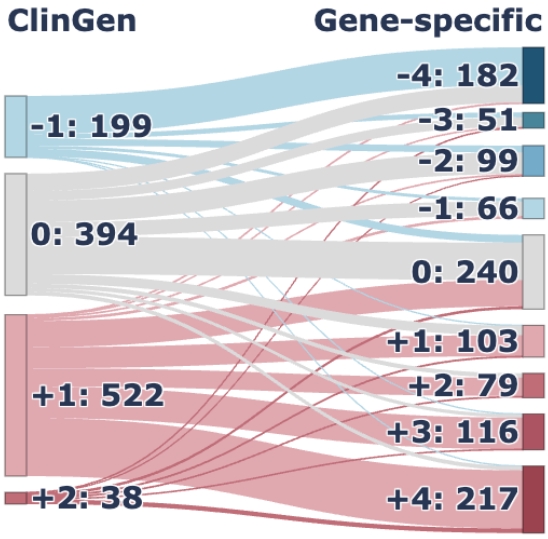


Extended Data Fig. 20: ClinGen Sankey by gene-specific calibration (AlphaMissense) Sankey diagrams illustrating the impact of gene-specific calibration using AlphaMissense scores. (a) Transitions in clinical classifications from original ClinGen assignments to recalibrated classifications. (b) Reassignment of computational evidence points under ClinGen guidelines compared with gene-specific calibration.

a ClinGen classification changes

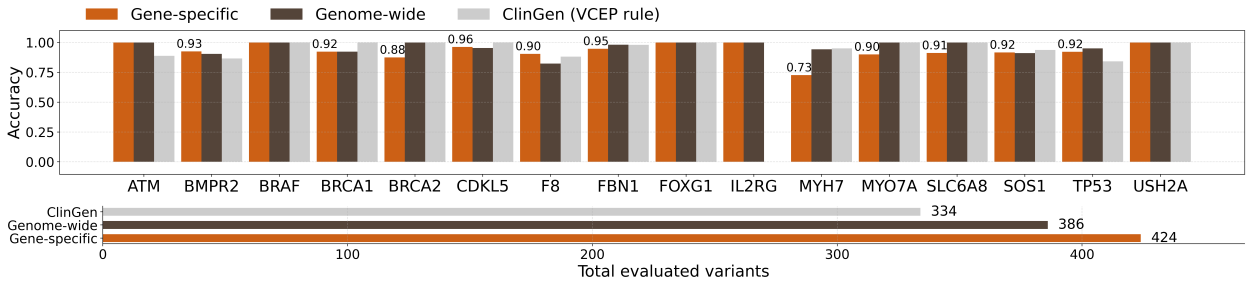


b ClinGen evidence point changes

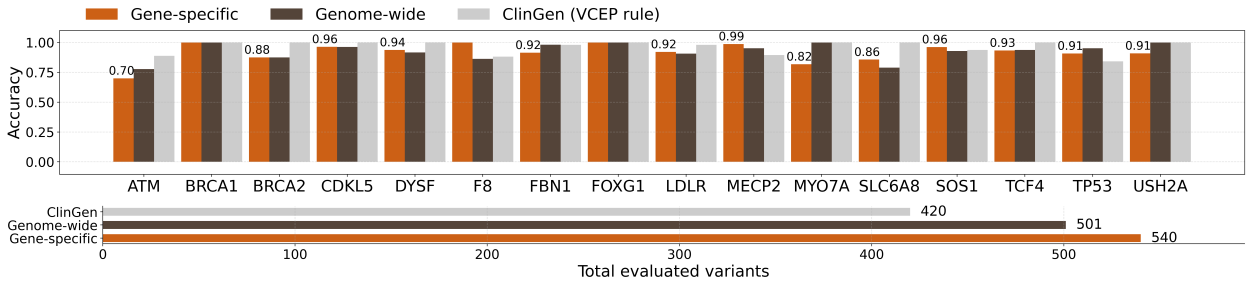


Extended Data Fig. 21: ClinGen Sankey by gene-specific calibration (MutPred2) Sankey diagrams illustrating the impact of gene-specific calibration using MutPred2 scores. (a) Transitions in clinical classifications from original ClinGen assignments to recalibrated classifications. (b) Reassignment of computational evidence points under ClinGen guidelines compared with gene-specific calibration.

a AlphaMissense



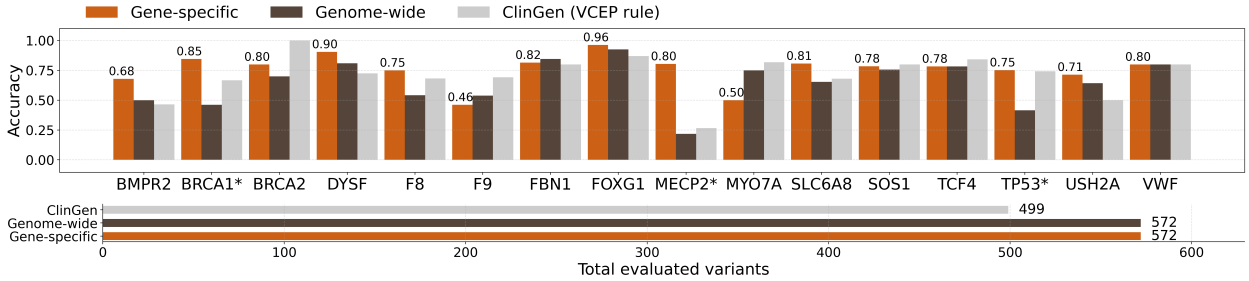
b MutPred2



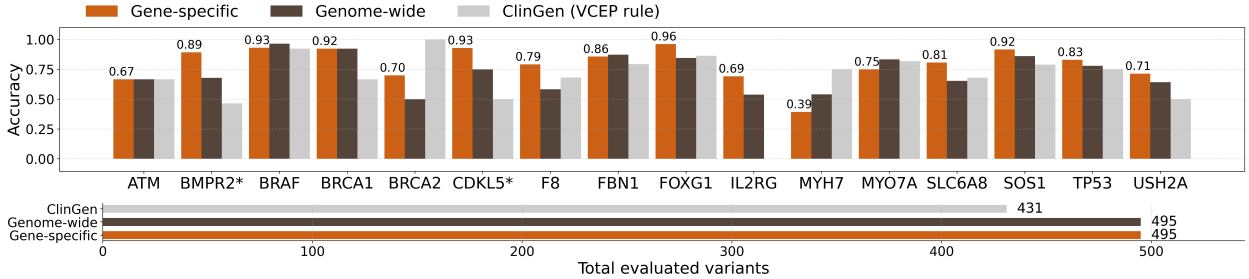
Extended Data Fig. 22: ClinGen non-circular set evaluation for gene-specific calibration (excluding zero-point assignments)

Evaluation of per-gene classification accuracy on the ClinGen non-circular variant set using (a) AlphaMissense and (b) MutPred2 scores under three approaches: ClinGen-provided computational evidence strengths, genome-wide calibration thresholds, and gene-specific calibration thresholds. Top bar plots show the fraction of variants correctly classified relative to ClinGen reference classifications after recomputing variant classifications using only non-computational evidence to maintain non-circularity. Variants assigned zero computational evidence points are excluded from accuracy calculations. Bottom bar plots show the total number of variants assigned non-zero computational evidence points by each method.

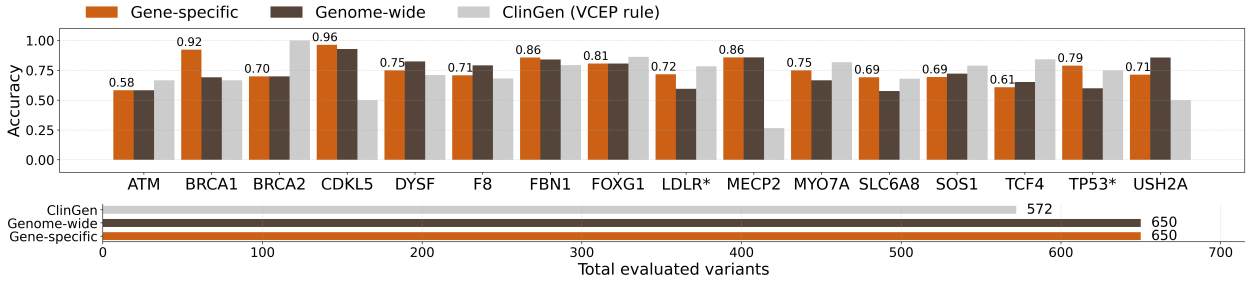
a REVEL



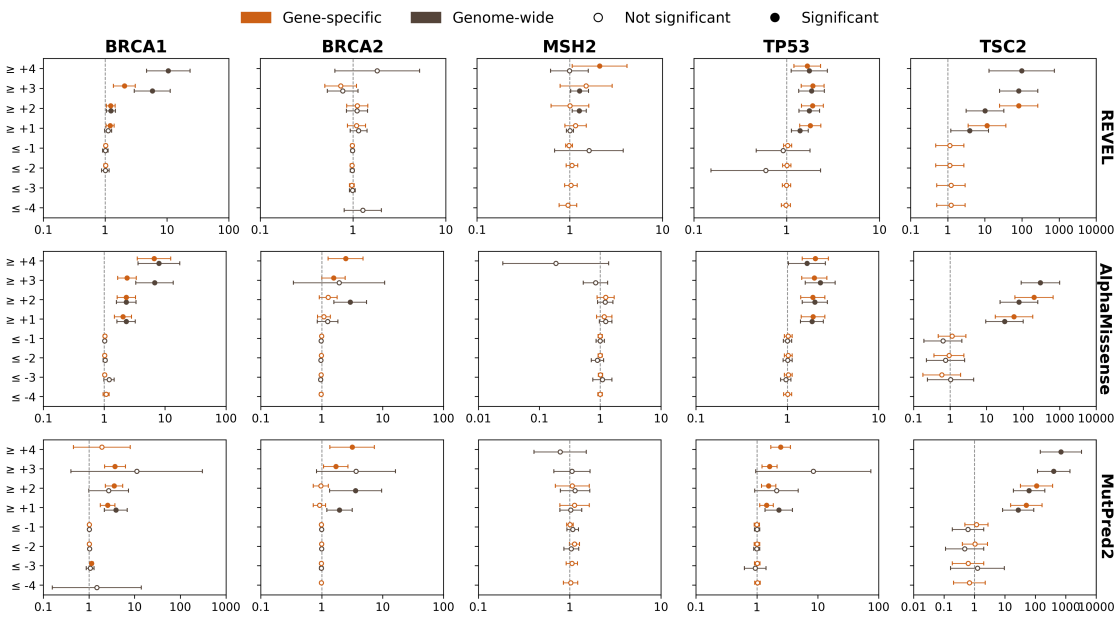
b AlphaMissense



c MutPred2

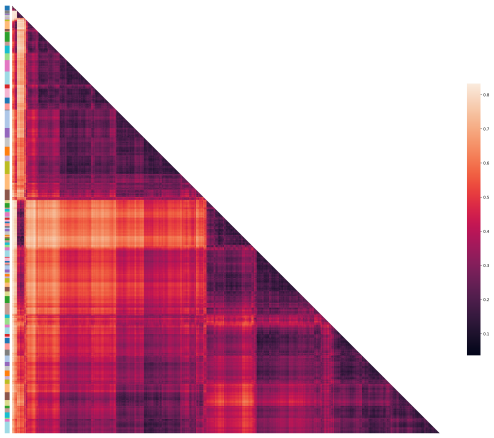


Extended Data Fig. 23: ClinGen non-circular set evaluation for gene-specific calibration (including zero-point assignments)
Evaluation of per-gene classification accuracy on the ClinGen non-circular variant set using (a) REVEL, (b) AlphaMissense, and (c) MutPred2 scores under three approaches: ClinGen-provided computational evidence strengths, genome-wide calibration thresholds, and gene-specific calibration thresholds. Top bar plots show the fraction of variants correctly classified relative to ClinGen reference classifications after recomputing variant classifications using only non-computational evidence to maintain non-circularity. Variants assigned zero computational evidence points are included in the accuracy calculations. Bottom bar plots show the total number of variants assigned computational evidence points by each method.

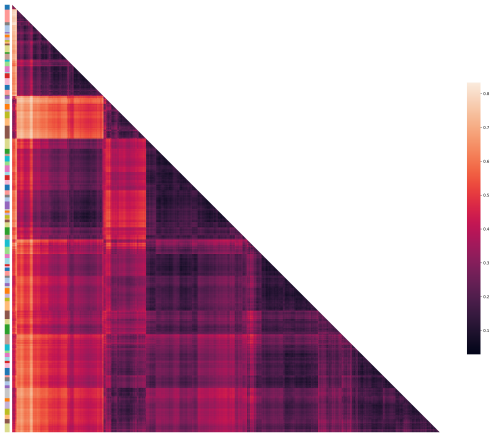


Extended Data Fig. 24: Odds ratios for disease occurrence in the All of Us biobank Odds ratios for disease occurrence in the All of Us biobank for variants meeting different evidence strength thresholds in example genes using gene-specific calibration compared with genome-wide calibration. The x-axis shows the odds ratio (vertical dashed line indicates OR = 1), and the y-axis shows total evidence points for variant sets. Circles represent estimated odds ratios with 95% confidence intervals (whiskers); filled circles indicate statistically significant associations.

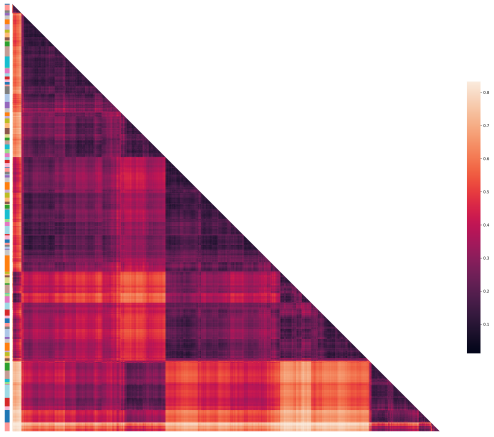
a REVEL



b AlphaMissense

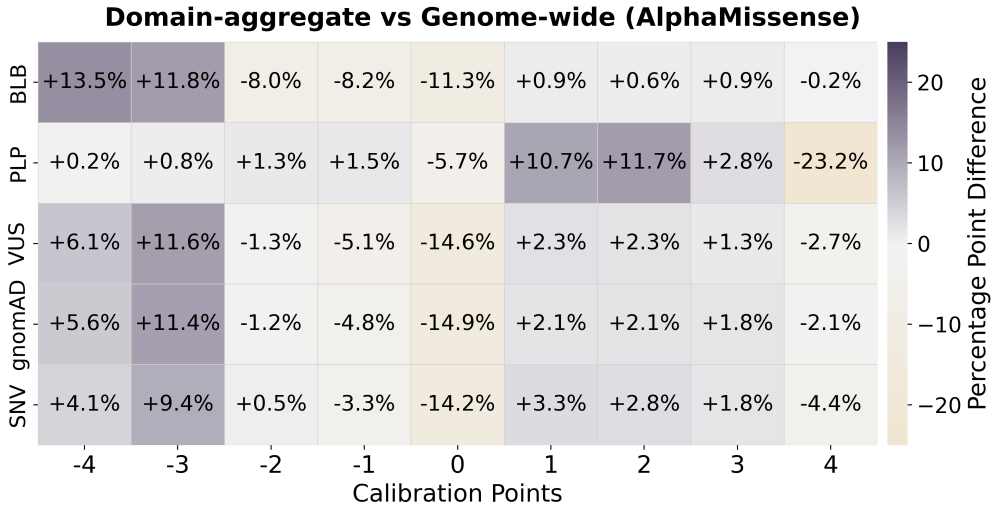


c MutPred2

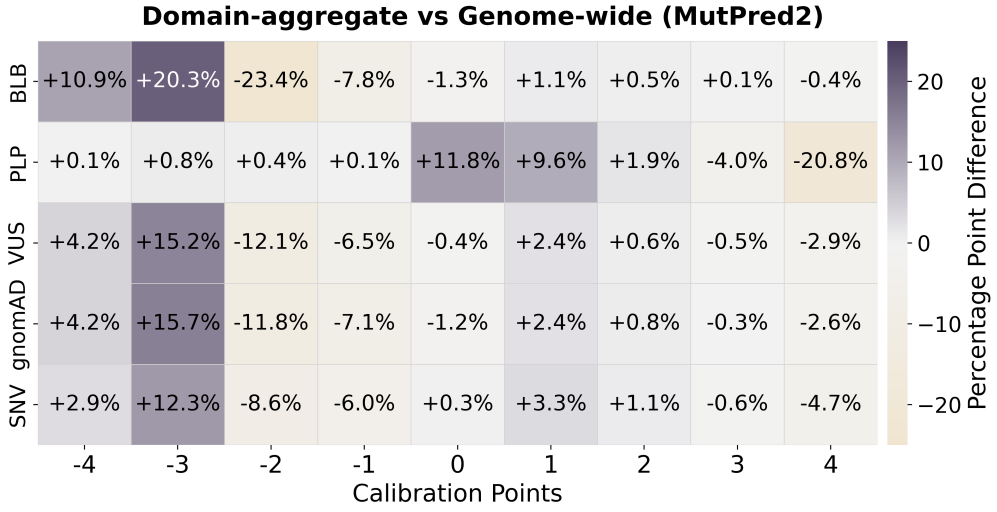


Extended Data Fig. 25: Domain-based clustering heatmap
Hierarchical clustering of domain score distributions for three variant effect predictors. Heatmaps show the full hierarchical clustering of domains based on Jensen–Shannon distance between score distributions, with sidebar colors indicating cluster assignments. (a) REVEL (98 clusters). (b) AlphaMissense (92 clusters). (c) MutPred2 (102 clusters).

a AlphaMissense



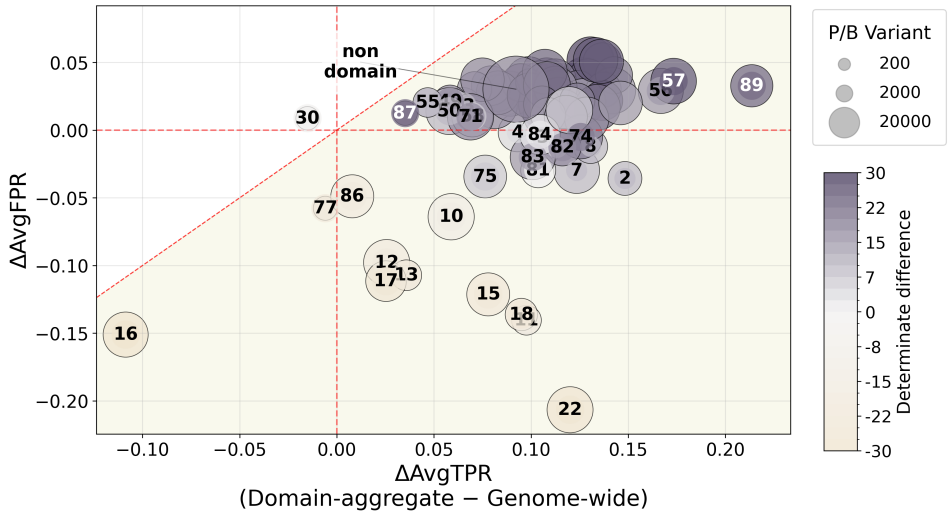
b MutPred2



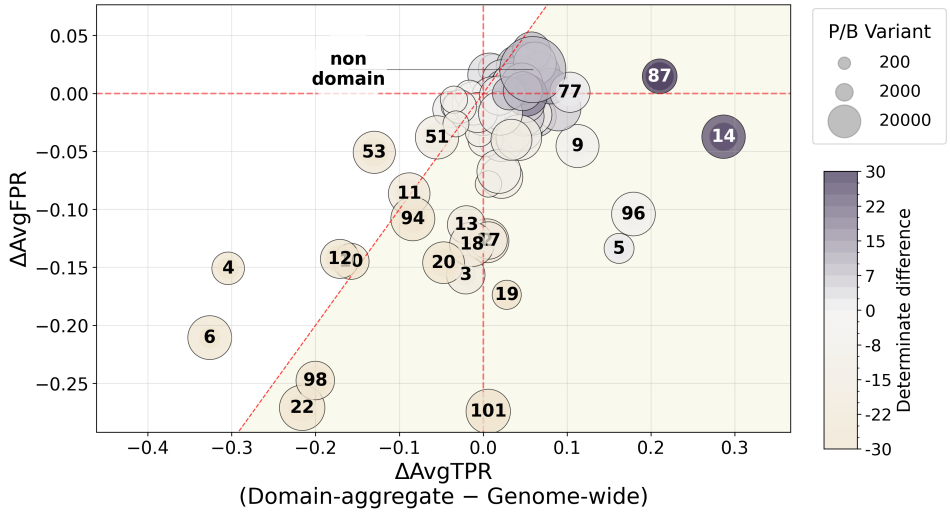
Extended Data Fig. 26: Evidence point assignment difference (domain-aggregate vs. genome-wide): AlphaMissense and MutPred2

Comparison of evidence point assignments between domain-aggregate and genome-wide calibration methods. (a) Results using AlphaMissense scores. (b) Results using MutPred2 scores. Heatmaps display the differences in the percentage of evidence point assignments stratified by variant classification or source. Grey indicates increased assignment rates by the domain-aggregate calibration method compared to genome-wide calibration.

a AlphaMissense

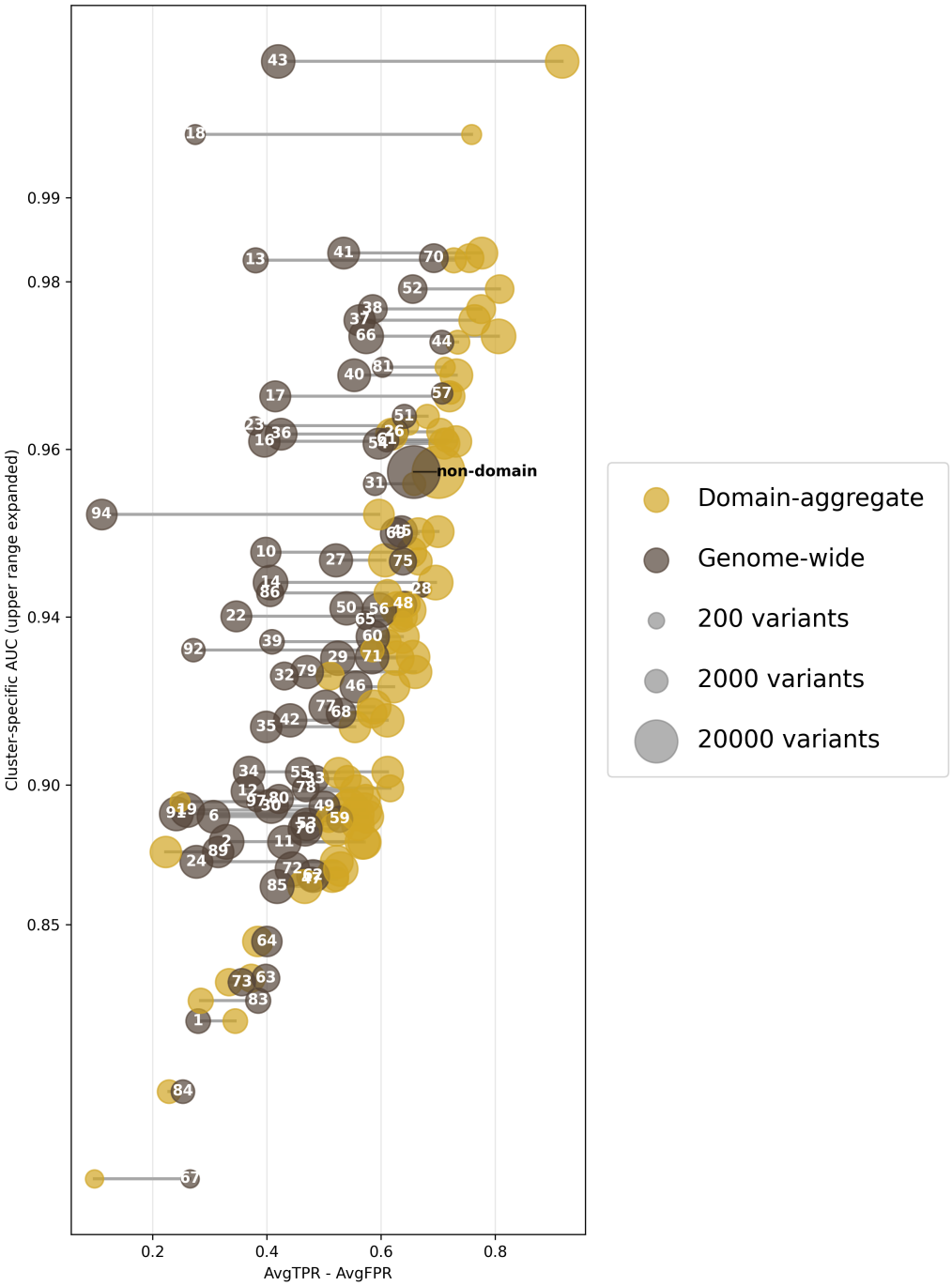


b MutPred2

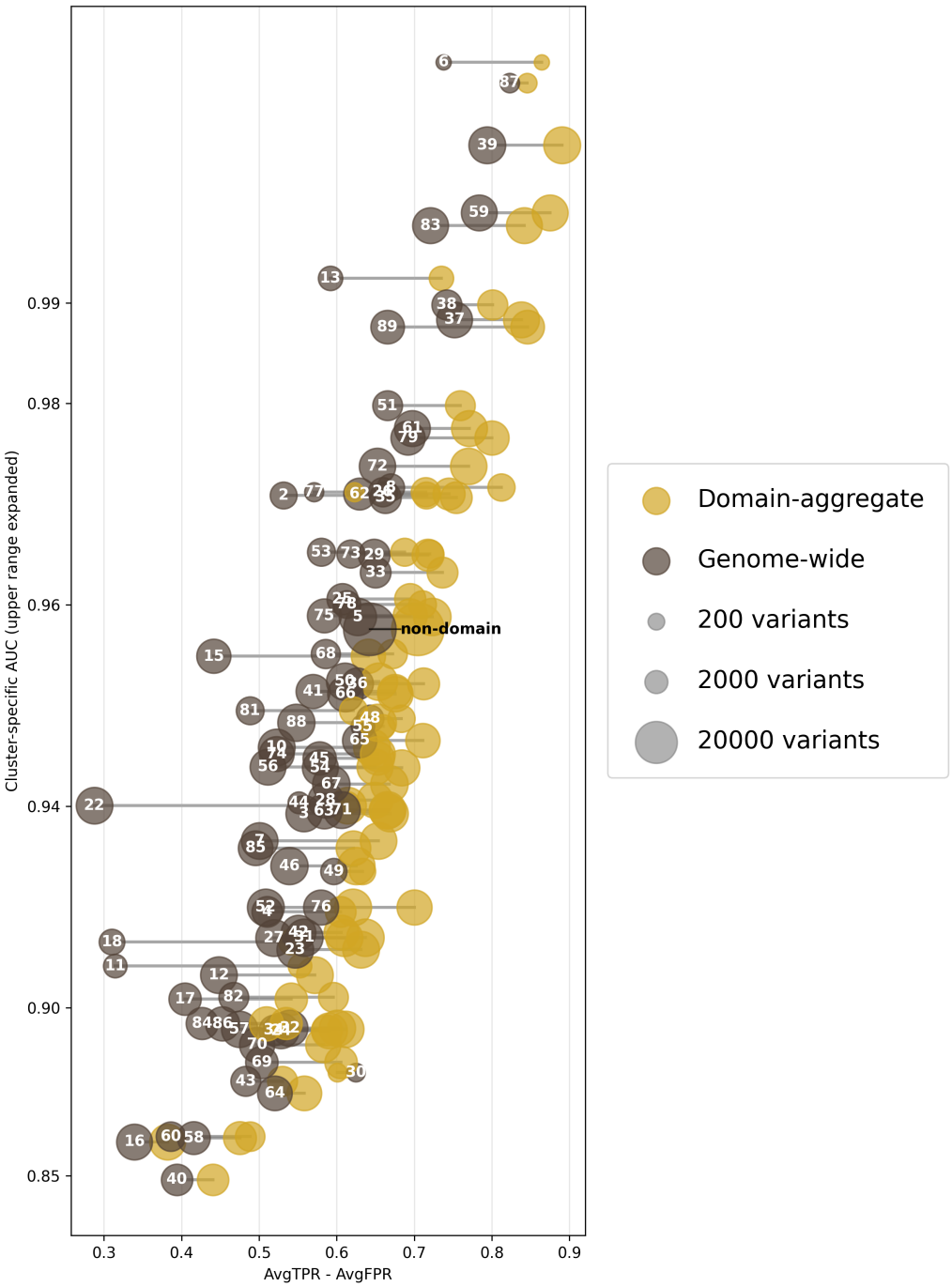


Extended Data Fig. 27: Calibration method comparison: AlphaMissense and MutPred2 (per-cluster)

Cluster-level changes in sensitivity and false positive rate following domain-aggregate calibration. (a) Results using AlphaMissense scores. (b) Results using MutPred2 scores. Each point represents a cluster. The x-axis shows the average change in true positive rate (ΔAvgTPR) between domain-aggregate and genome-wide calibration (domain-aggregate minus genome-wide), and the y-axis shows the corresponding change in false positive rate (ΔAvgFPR). Point color encodes the change in evidence coverage (fraction of variants receiving at least -/+1 point of evidence), with deeper grey indicating higher coverage under the domain-aggregate method. Point size is proportional to the number of pathogenic and benign variants (n_{PB}; P/LP and B/LB). The shaded area indicates increased performance (below the dashed red diagonal $x = y$).

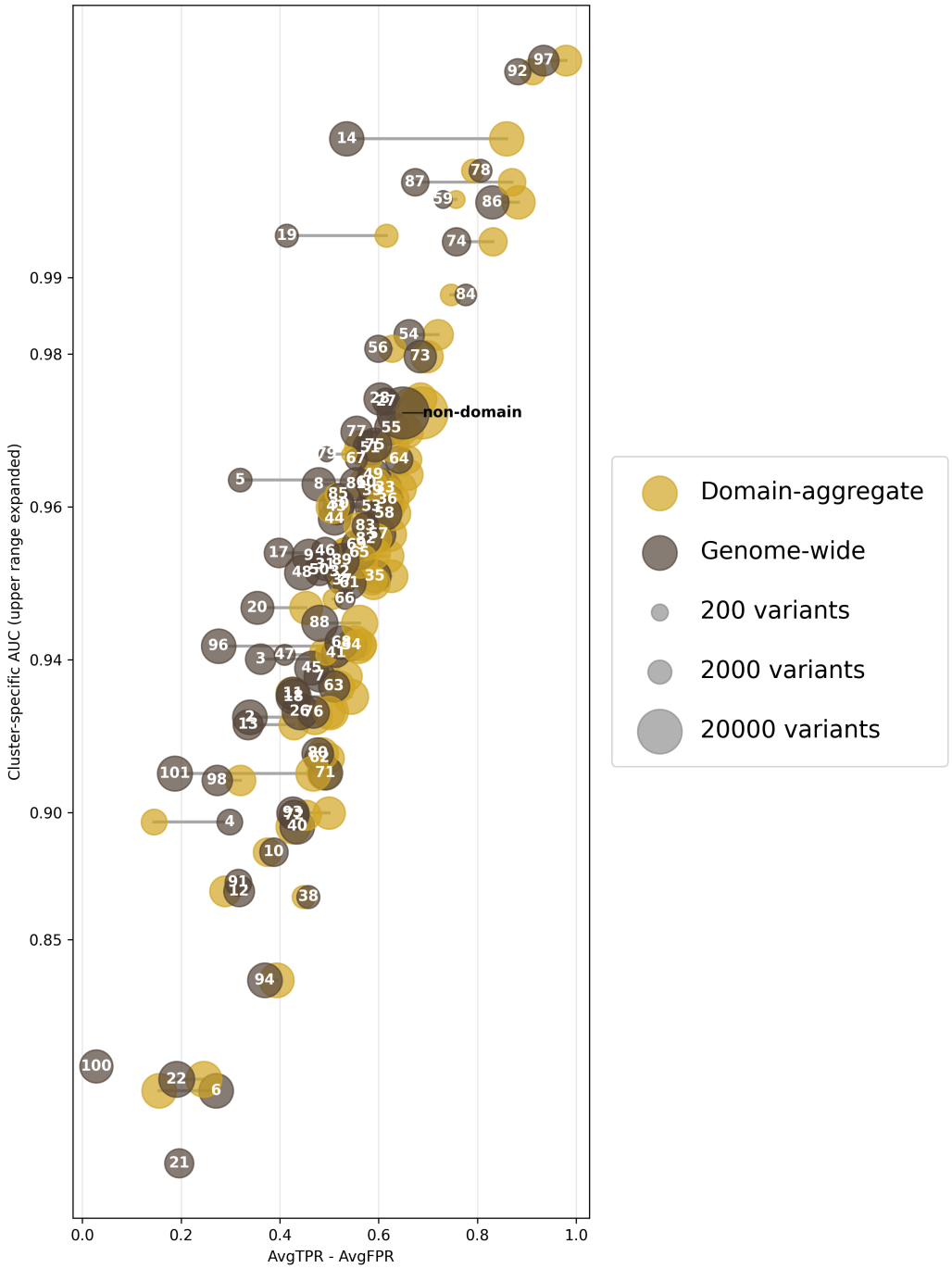


Extended Data Fig. 28: Cluster-specific calibration performance and REVEL AUC Cluster-specific performance (AvgTPR − AvgFPR) versus cluster-specific REVEL AUC shown as a dumbbell plot. Each cluster is represented by a pair of connected points (domain-aggregate calibration vs. genome-wide calibration). The x-axis shows the domain-aggregate performance metric (AvgTPR − AvgFPR), and the y-axis shows the cluster-specific REVEL AUC. Higher values of AvgTPR − AvgFPR indicate better discrimination performance. The y-axis is displayed on a non-linear scale, with the upper range expanded to improve visualization of differences across clusters.



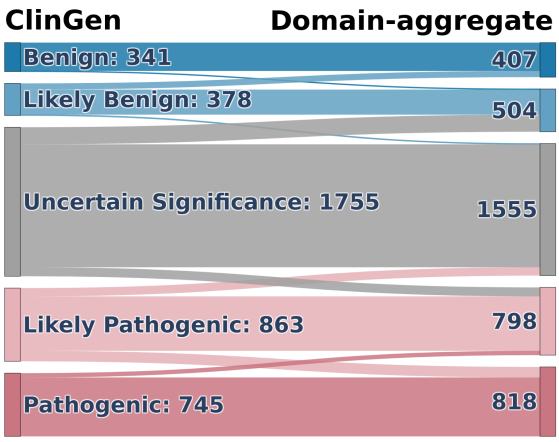
Extended Data Fig. 29: Cluster-specific calibration performance and AlphaMissense AUC

Cluster-specific performance (AvgTPR − AvgFPR) versus cluster-specific AlphaMissense AUC shown as a dumbbell plot. Each cluster is represented by a pair of connected points (domain-aggregate calibration vs. genome-wide calibration). The x-axis shows the domain-aggregate performance metric (AvgTPR − AvgFPR), and the y-axis shows the cluster-specific REVEL AUC. Higher values of AvgTPR − AvgFPR indicate better discrimination performance. The y-axis is displayed on a non-linear scale, with the upper range expanded to improve visualization of differences across clusters.

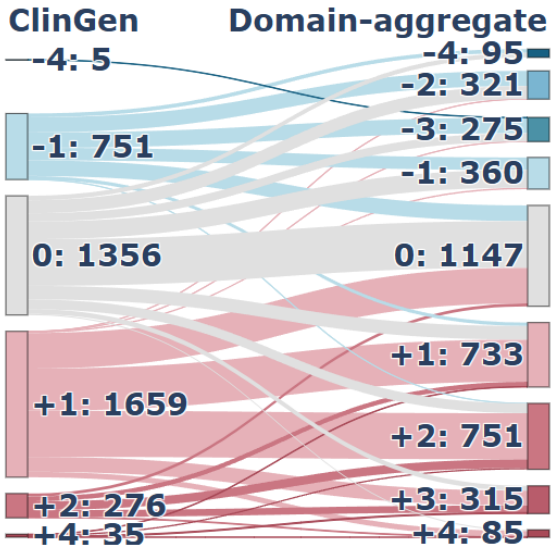


Extended Data Fig. 30: Cluster-specific calibration performance and MutPred2 AUC Cluster-specific performance (AvgTPR − AvgFPR) versus cluster-specific MutPred2 AUC shown as a dumbbell plot. Each cluster is represented by a pair of connected points (domain-aggregate calibration vs. genome-wide calibration). The x-axis shows the domain-aggregate performance metric (AvgTPR − AvgFPR), and the y-axis shows the cluster-specific REVEL AUC. Higher values of AvgTPR − AvgFPR indicate better discrimination performance. The y-axis is displayed on a non-linear scale, with the upper range expanded to improve visualization of differences across clusters.

a ClinGen classification changes

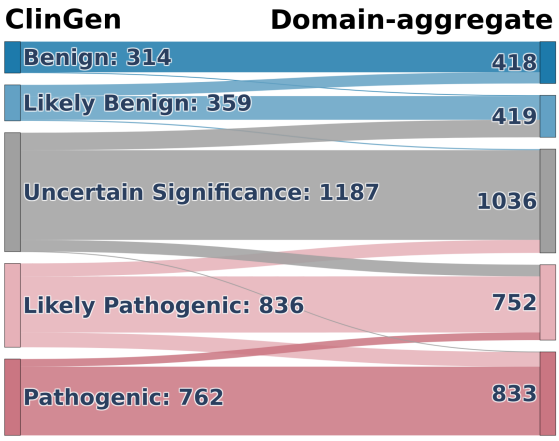


b ClinGen vs domain-aggregate evidence points

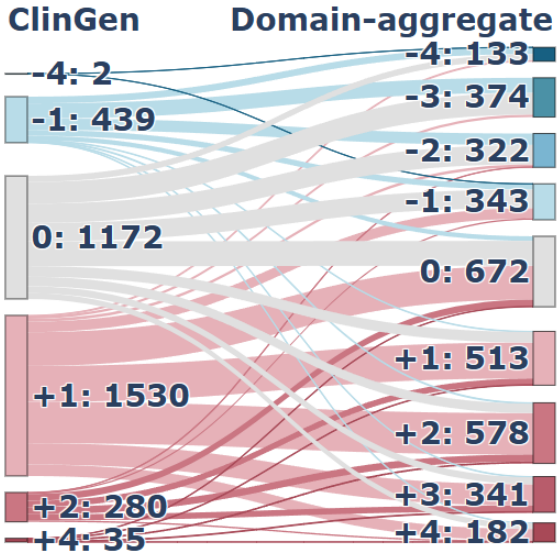


Extended Data Fig. 31: ClinGen Sankey by domain-aggregate calibration (REVEL) Sankey diagrams illustrating the impact of domain-aggregate calibration using REVEL scores. (a) Transitions in clinical classifications from original ClinGen assignments to recalibrated classifications. (b) Reassignment of computational evidence points under ClinGen guidelines compared with domain-aggregate calibration.

a ClinGen classification changes



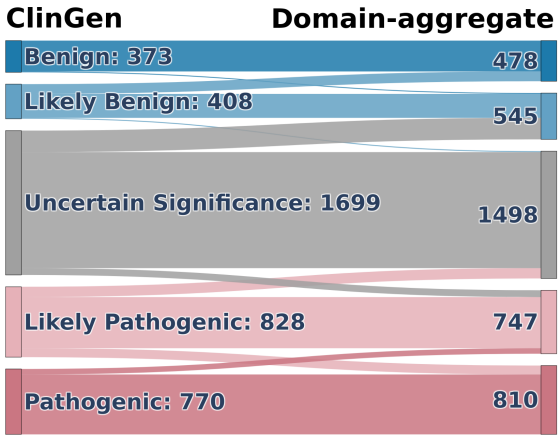
b ClinGen vs domain-aggregate evidence points



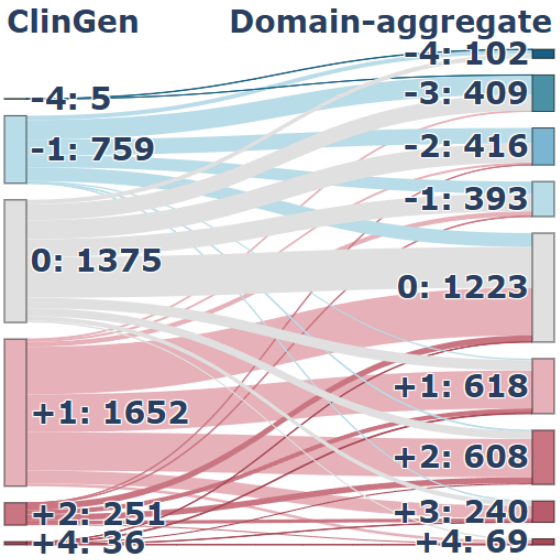
Extended Data Fig. 32: ClinGen Sankey by domain-aggregate calibration (AlphaMissense)

Sankey diagrams illustrating the impact of domain-aggregate calibration using AlphaMissense scores. (a) Transitions in clinical classifications from original ClinGen assignments to recalibrated classifications. (b) Reassignment of computational evidence points under ClinGen guidelines compared with domain-aggregate calibration.

a ClinGen classification changes

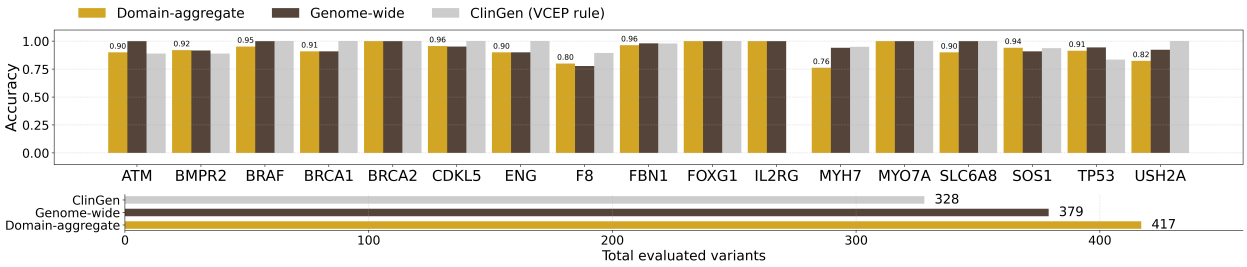


b ClinGen vs domain-aggregate evidence points

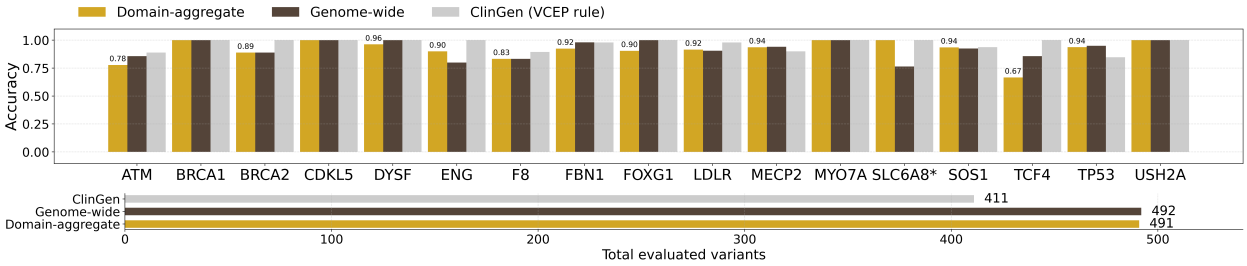


Extended Data Fig. 33: ClinGen Sankey by domain-aggregate calibration (MutPred2) Sankey diagrams illustrating the impact of domain-aggregate calibration using MutPred2 scores. (a) Transitions in clinical classifications from original ClinGen assignments to recalibrated classifications. (b) Reassignment of computational evidence points under ClinGen guidelines compared with domain-aggregate calibration.

a AlphaMissense

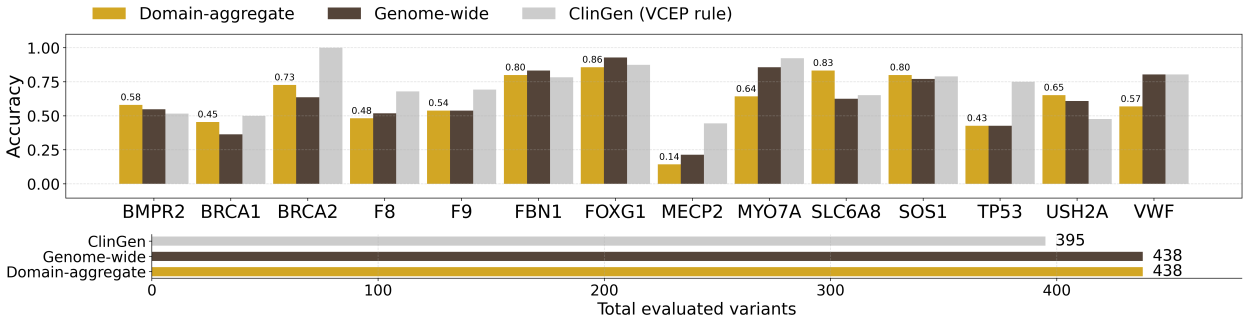


b MutPred2

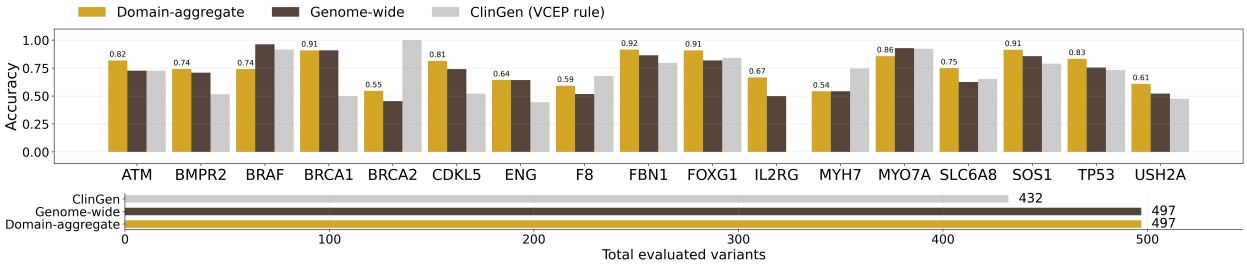


Extended Data Fig. 34: ClinGen non-circular set evaluation for domain-aggregate calibration (excluding zero-point assignments)
Evaluation of per-gene classification accuracy on the ClinGen non-circular variant set using (a) AlphaMissense and (b) MutPred2 scores under three approaches: ClinGen-provided computational evidence strengths, genome-wide calibration thresholds, and domain-aggregate calibration thresholds. Top bar plots show the fraction of variants correctly classified relative to ClinGen reference classifications after recomputing variant classifications using only non-computational evidence to maintain non-circularity. Variants assigned zero computational evidence points are excluded from accuracy calculations. Bottom bar plots show the total number of variants assigned non-zero computational evidence points by each method.

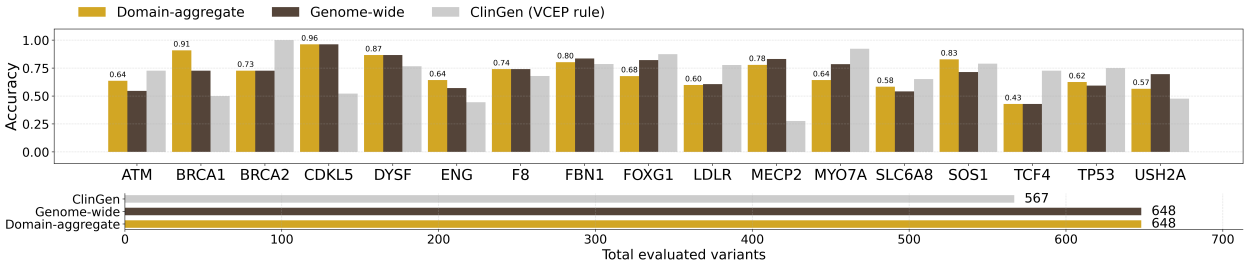
a REVEL



b AlphaMissense

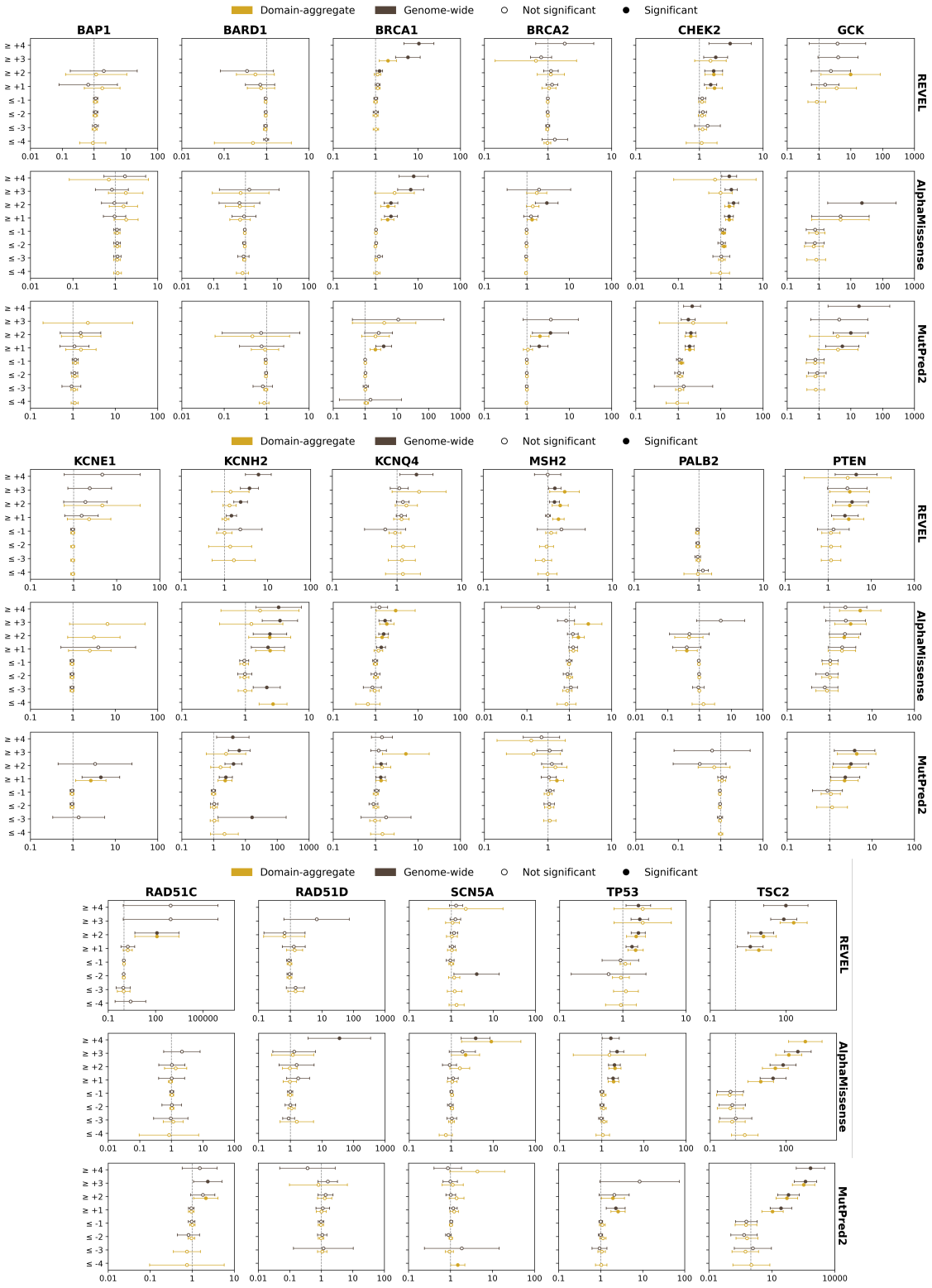


c MutPred2



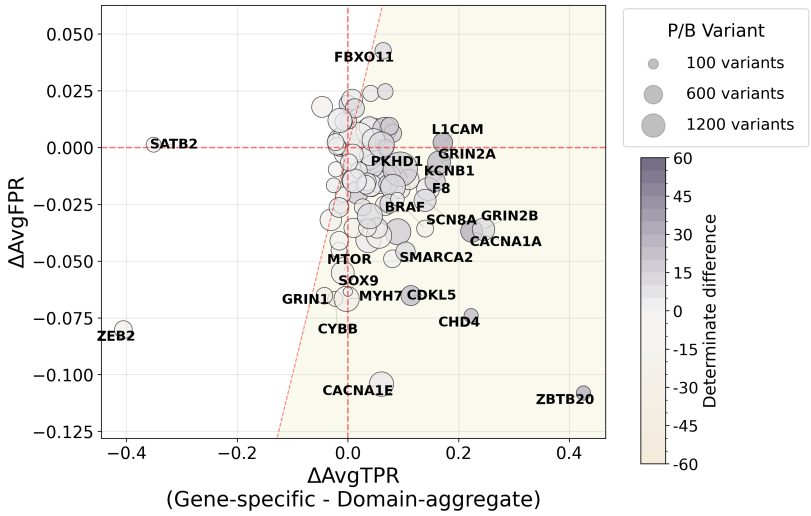
Extended Data Fig. 35: ClinGen non-circular set evaluation for domain-aggregate calibration (including zero-point assignments)

Evaluation of per-gene classification accuracy on the ClinGen non-circular variant set using (a) REVEL, (b) AlphaMissense, and (c) MutPred2 scores under three approaches: ClinGen-provided computational evidence strengths, genome-wide calibration thresholds, and domain-aggregate calibration thresholds. Top bar plots show the fraction of variants correctly classified relative to ClinGen reference classifications after recomputing variant classifications using only non-computational evidence to maintain non-circularity. Variants assigned zero computational evidence points are included in the accuracy calculations. Bottom bar plots show the total number of variants assigned computational evidence points by each method.

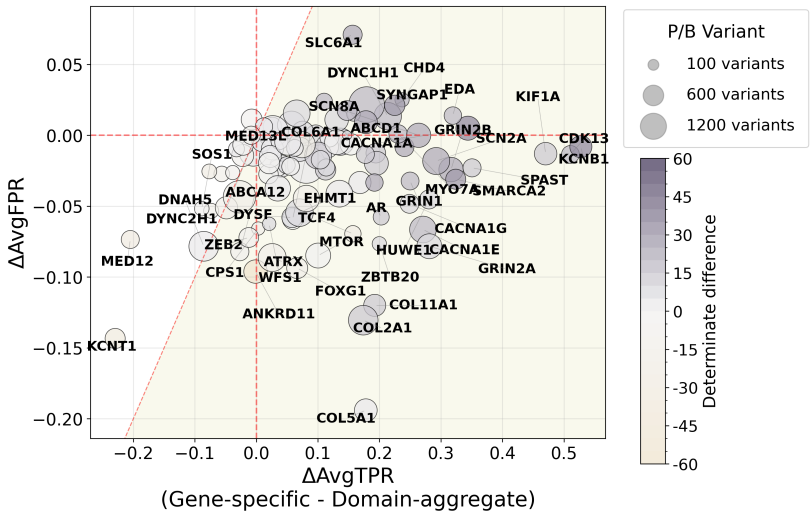


Extended Data Fig. 36: Odds ratios for disease occurrence in the All of Us biobank Odds ratios for disease occurrence in the All of Us biobank for variants meeting different evidence strength thresholds in example genes using domain-aggregate calibration compared with genome-wide calibration. The x-axis shows the odds ratio (vertical dashed line indicates OR = 1), and the y-axis shows total evidence points for variant sets. Circles represent estimated odds ratios with 95% confidence intervals (whiskers); filled circles indicate statistically significant associations.

a AlphaMissense



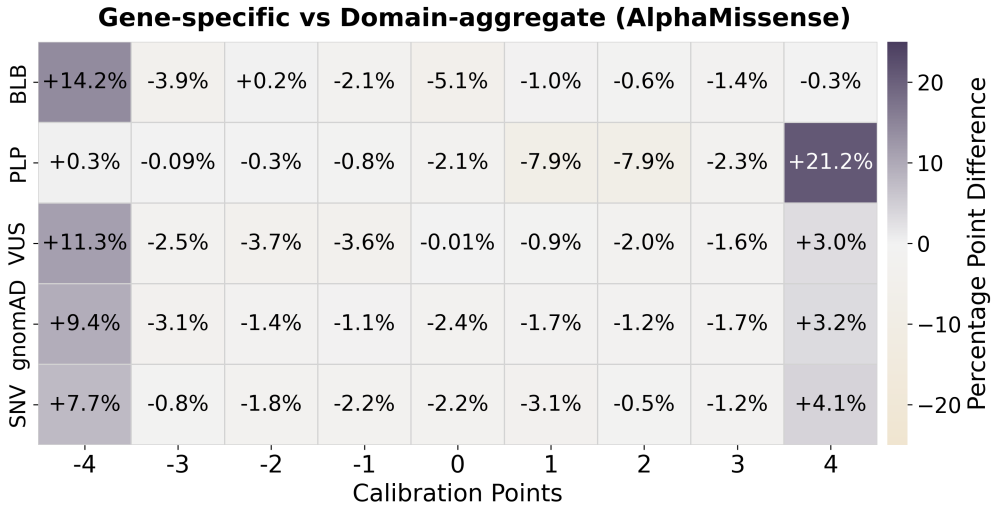
b MutPred2



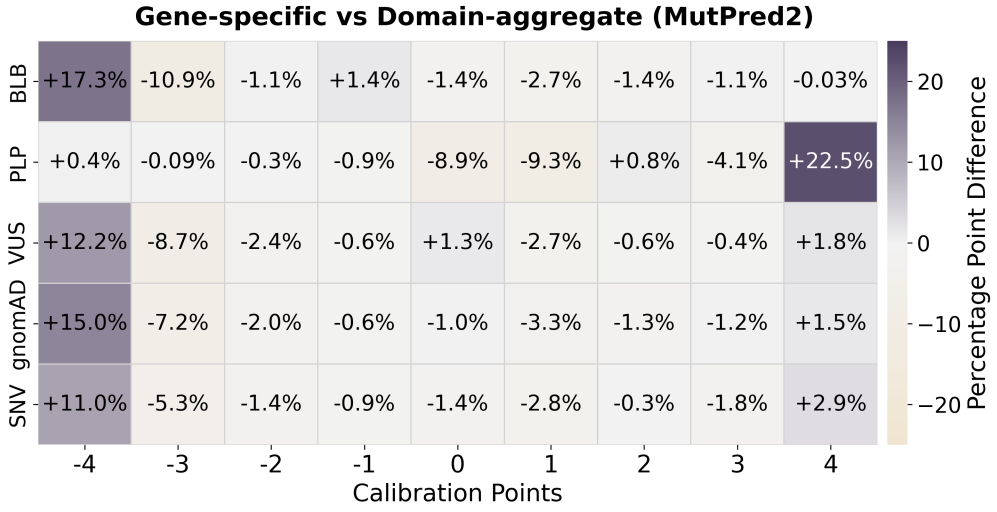
Extended Data Fig. 37: Calibration method comparison: AlphaMissense and MutPred2 (per-gene)

Gene-level changes in sensitivity and false positive rate comparing gene-specific calibration versus domain-aggregate calibration. (a) Results using AlphaMissense scores. (b) Results using MutPred2 scores. Each point represents a gene. The x-axis shows the average change in true positive rate (ΔAvgTPR) between gene-specific calibration and domain-aggregate calibration (gene-specific minus domain-aggregate), and the y-axis shows the corresponding change in false positive rate (ΔAvgFPR). Point color encodes the change in evidence coverage (fraction of variants receiving at least -/+1 point of evidence), with deeper grey indicating higher coverage under the gene-specific method. Point size is proportional to the number of pathogenic and benign variants (n_{PB}; P/LP and B/LB). The shaded area indicates increased performance (below the dashed red diagonal $x = y$).

a AlphaMissense

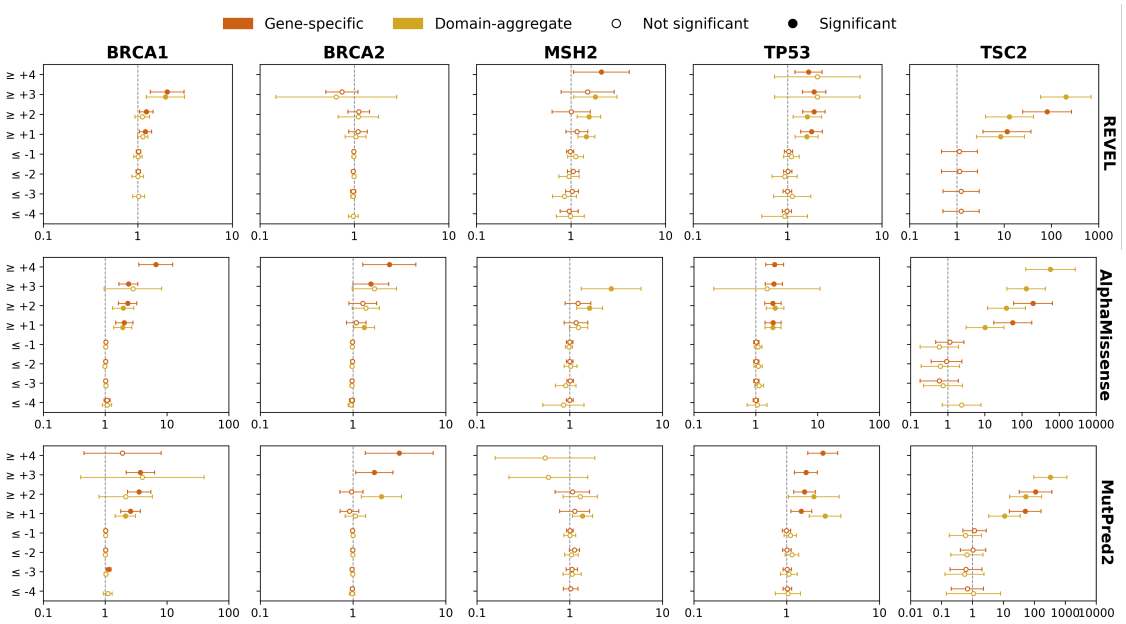


b MutPred2



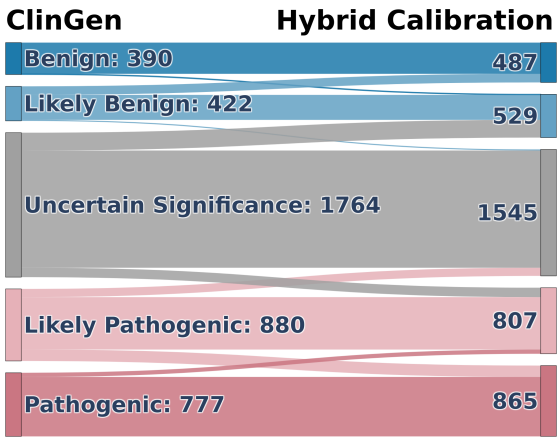
Extended Data Fig. 38: Evidence point assignment difference (gene-specific vs. domain-aggregate): AlphaMissense and MutPred2

Comparison of evidence point assignments between gene-specific and domain-aggregate calibration methods. (a) Results using AlphaMissense scores. (b) Results using MutPred2 scores. Heatmaps display percentage point differences in evidence point assignments stratified by variant classification. Darker grey shades indicate higher assignment rates under the gene-specific calibration method compared to domain-aggregate calibration.

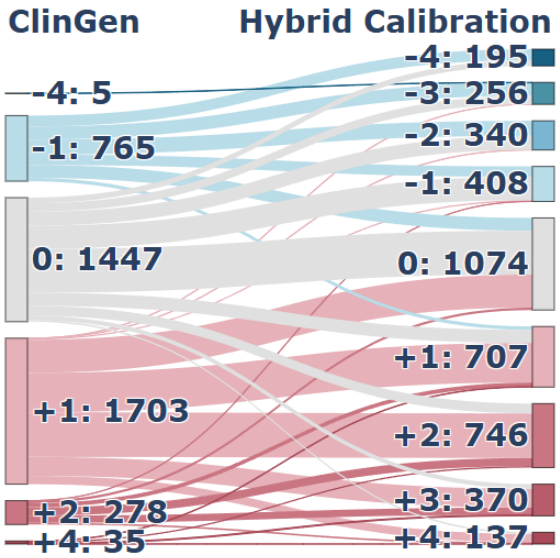


Extended Data Fig. 39: Odds ratios for disease occurrence in the All of Us biobank Odds ratios for disease occurrence in the All of Us biobank for variants meeting different evidence strength thresholds in example genes using gene-specific calibration compared with domain-aggregate calibration. The x-axis shows the odds ratio (vertical dashed line indicates OR = 1), and the y-axis shows total evidence points for variant sets. Circles represent estimated odds ratios with 95% confidence intervals (whiskers); filled circles indicate statistically significant associations.

a ClinGen classification changes

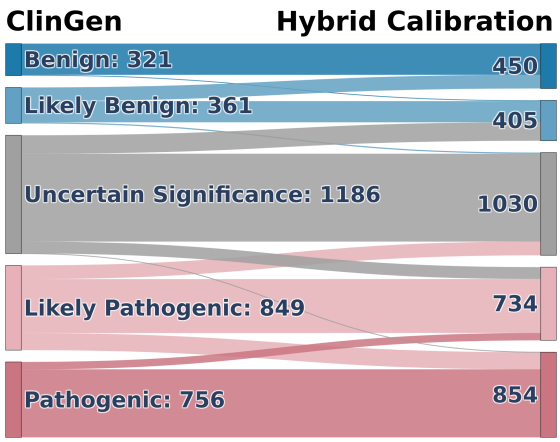


b ClinGen evidence point changes

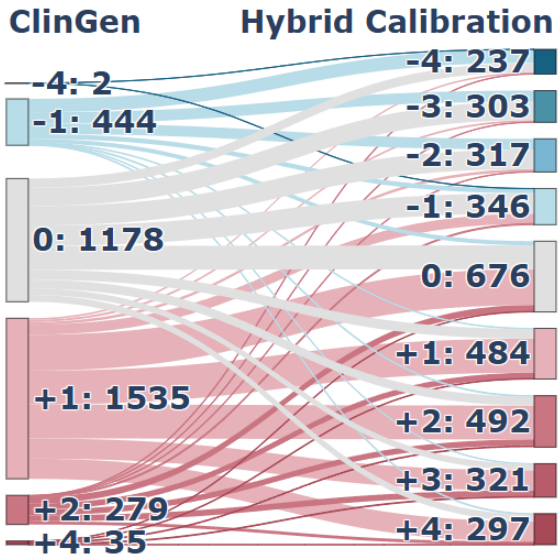


Extended Data Fig. 40: ClinGen Sankey by hybrid calibration (REVEL)
Sankey diagrams illustrating the impact of hybrid calibration using REVEL scores. (a) Transitions in clinical classifications from original ClinGen assignments to recalibrated classifications. (b) Reassignment of computational evidence points under ClinGen guidelines compared with hybrid calibration.

a ClinGen classification changes

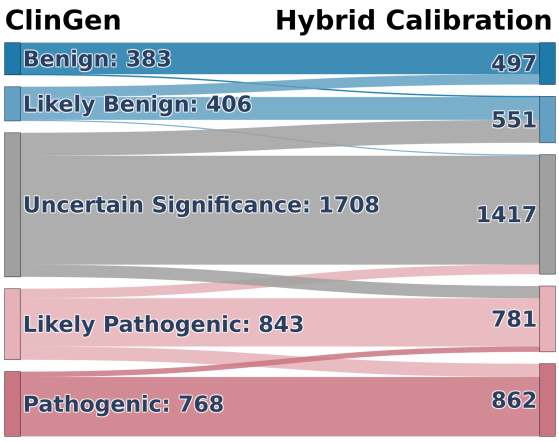


b ClinGen evidence point changes

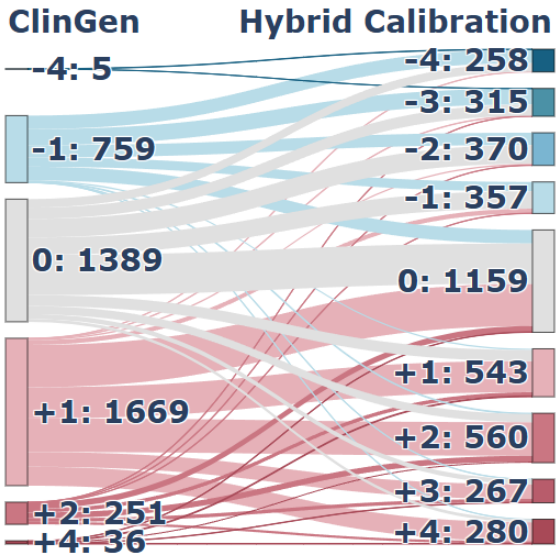


Extended Data Fig. 41: ClinGen Sankey by hybrid calibration (AlphaMissense) Sankey diagrams illustrating the impact of hybrid calibration using AlphaMissense scores. (a) Transitions in clinical classifications from original ClinGen assignments to recalibrated classifications. (b) Reassignment of computational evidence points under ClinGen guidelines compared with hybrid calibration.

a ClinGen classification changes



b ClinGen evidence point changes



Extended Data Fig. 42: ClinGen Sankey by hybrid calibration (MutPred2)
Sankey diagrams illustrating the impact of hybrid calibration using MutPred2 scores. (a) Transitions in clinical classifications from original ClinGen assignments to recalibrated classifications. (b) Reassignment of computational evidence points under ClinGen guidelines compared with hybrid calibration.