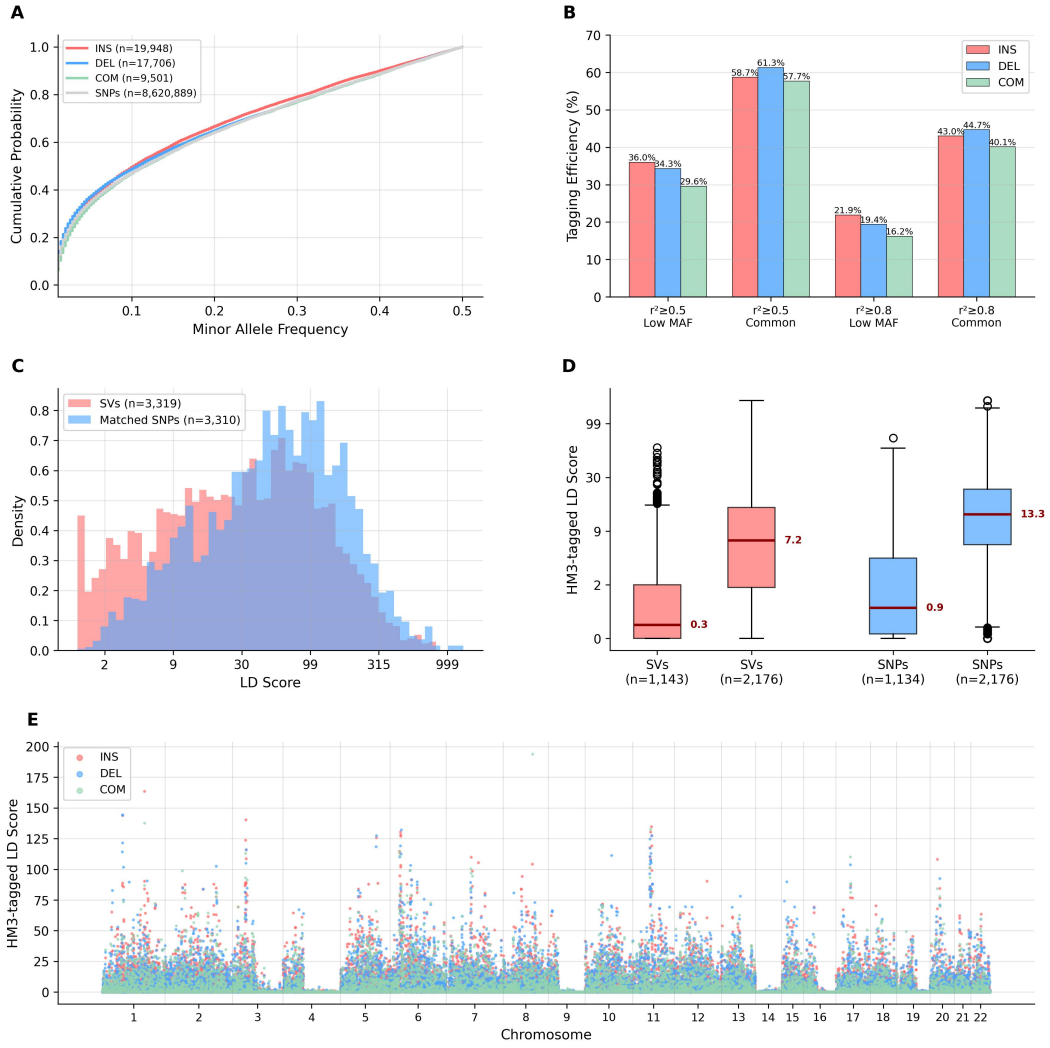
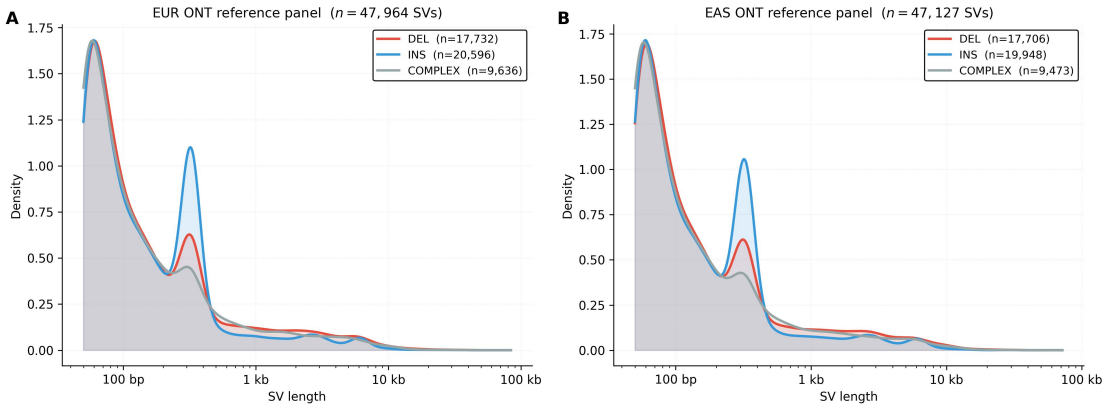
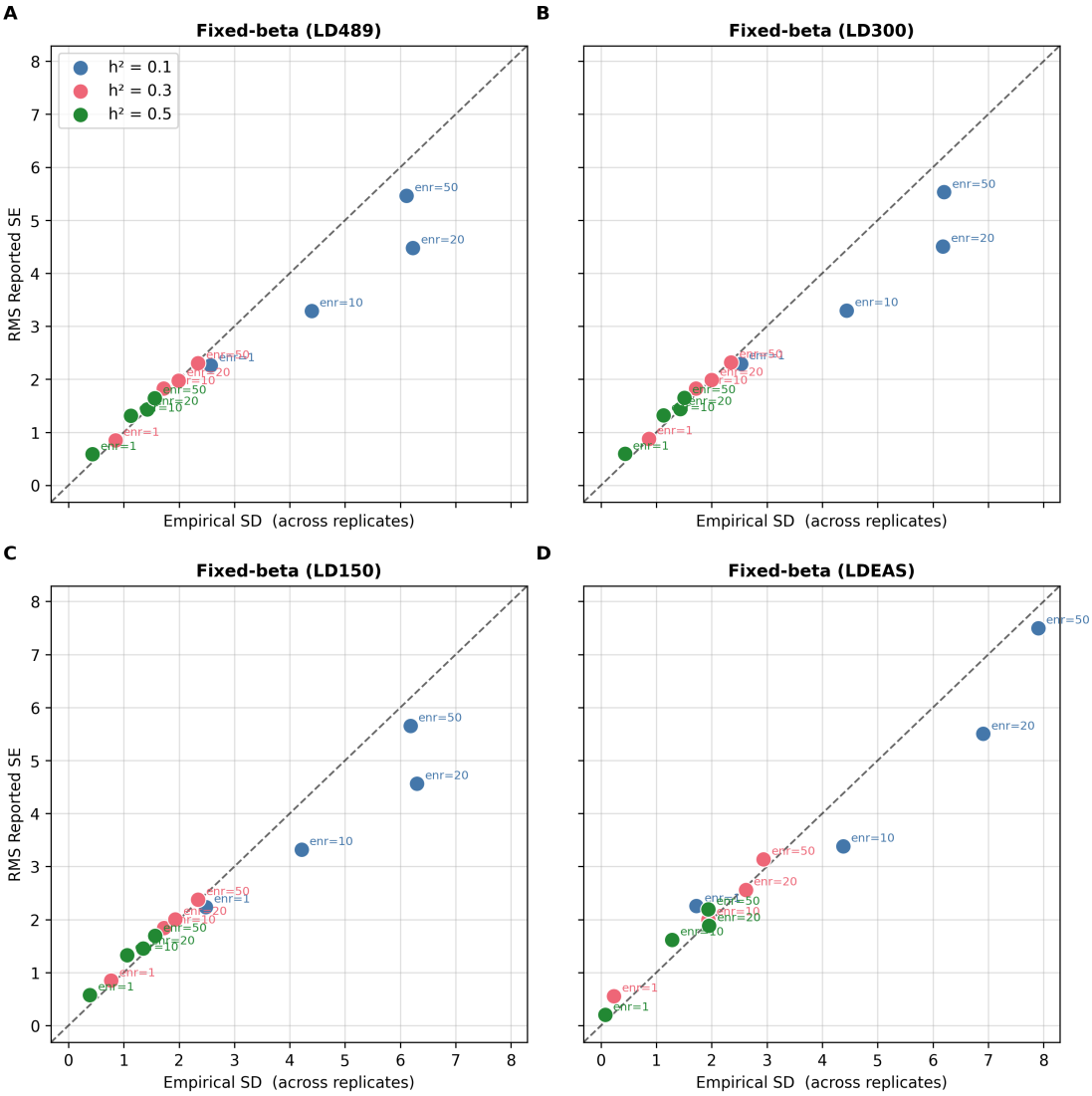
Supplementary Figures



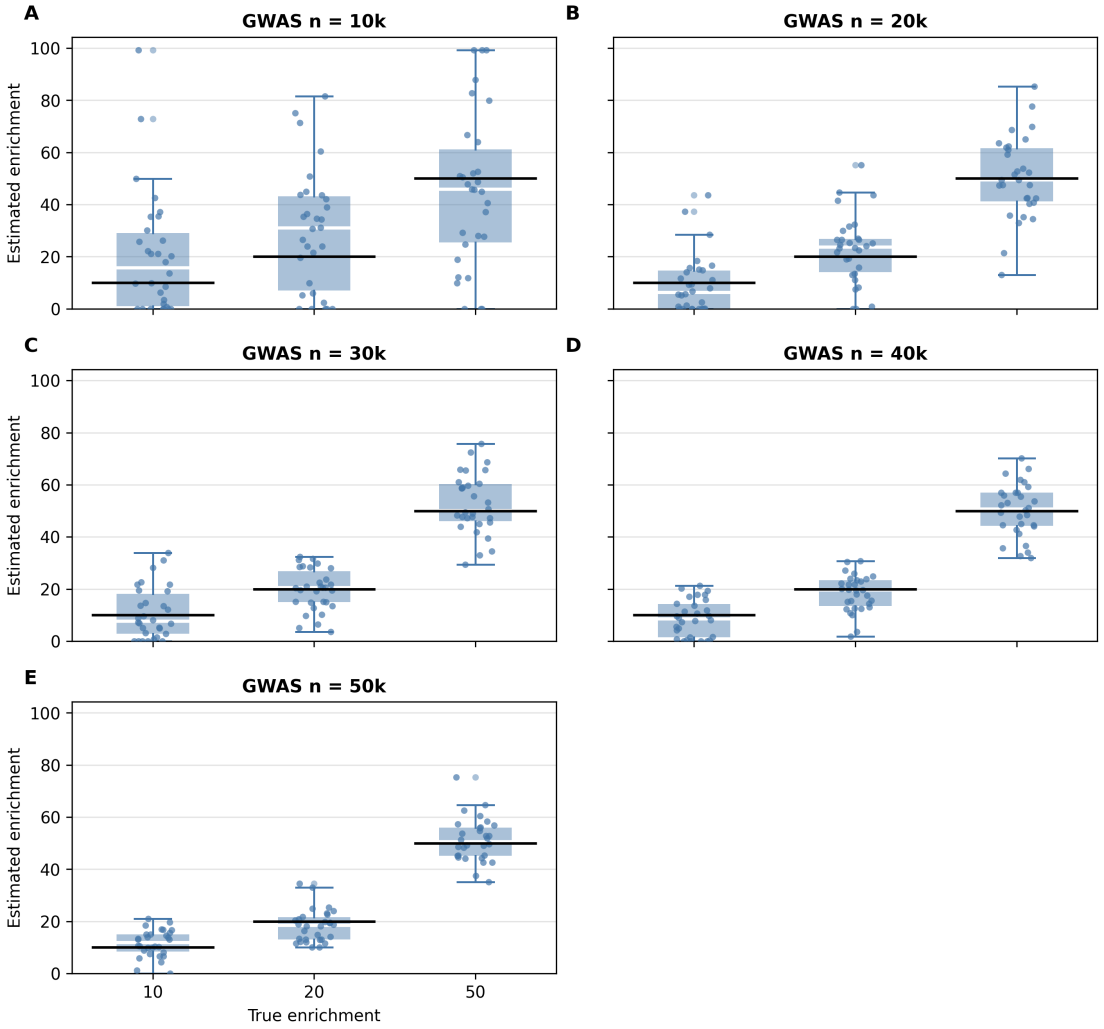
Supplementary Figure S1: **Structural variants exhibit substantial SNP tagging in East Asian ancestry.** **(A)** Cumulative MAF distributions show insertions enriched for rare alleles, while deletions and complex rearrangements resemble SNPs ($N = 47{,}155$ SVs: 19,948 insertions, 17,706 deletions, 9,501 complex rearrangements). **(B)** SV tagging efficiency using all reference-panel SNPs as potential tags, stratified by MAF and r^2 threshold. Common SVs (MAF$\geq$0.05) show 57.7 to 61.3% tagging at $r^2 \geq 0.5$, whereas low-MAF SVs show 29.6 to 36.0%. At the stringent threshold ($r^2 \geq 0.8$), common SVs show 40.1 to 44.7% tagging and low-MAF SVs 16.2 to 21.9%. **(C)** Reference LD score density distributions on chromosome 1 (SVs $n = 3{,}319$, MAF-matched SNPs $n = 3{,}310$) using all SNPs as tags, showing that SVs have somewhat lower but overlapping LD scores compared with SNPs. **(D)** HM3-tagged LD scores on chromosome 1 after restricting tag variants to HapMap3 SNPs, stratified by MAF (SVs low MAF $n = 1{,}143$, SVs common $n = 2{,}176$, SNPs low MAF $n = 1{,}134$, SNPs common $n = 2{,}176$). **(E)** Genome-wide HM3-tagged LD scores for all 47,155 SVs reveal LD hotspots and inter-chromosomal heterogeneity in SV tagging by HapMap3 SNPs.



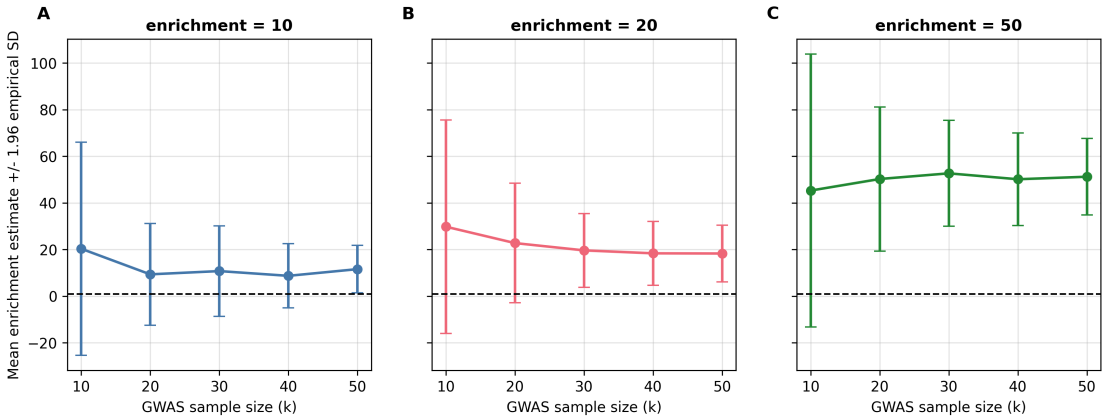
Supplementary Figure S2: **Size distribution of SV classes in the EUR and EAS ONT reference panels.** Kernel density estimates (log-scaled x-axis) of SV length for DEL (red), INS (blue), and COM (gray) in **(A)** the EUR ONT reference panel ($n = 47{,}964$ SVs) and **(B)** the EAS ONT reference panel ($n = 47{,}127$ SVs). All three classes are predominantly short (class medians 88–105 bp), with lengths ranging from 50 bp to approximately 84 kb. Class composition is consistent between ancestries ($\sim$43% INS, $\sim$37% DEL, $\sim$20% COM).



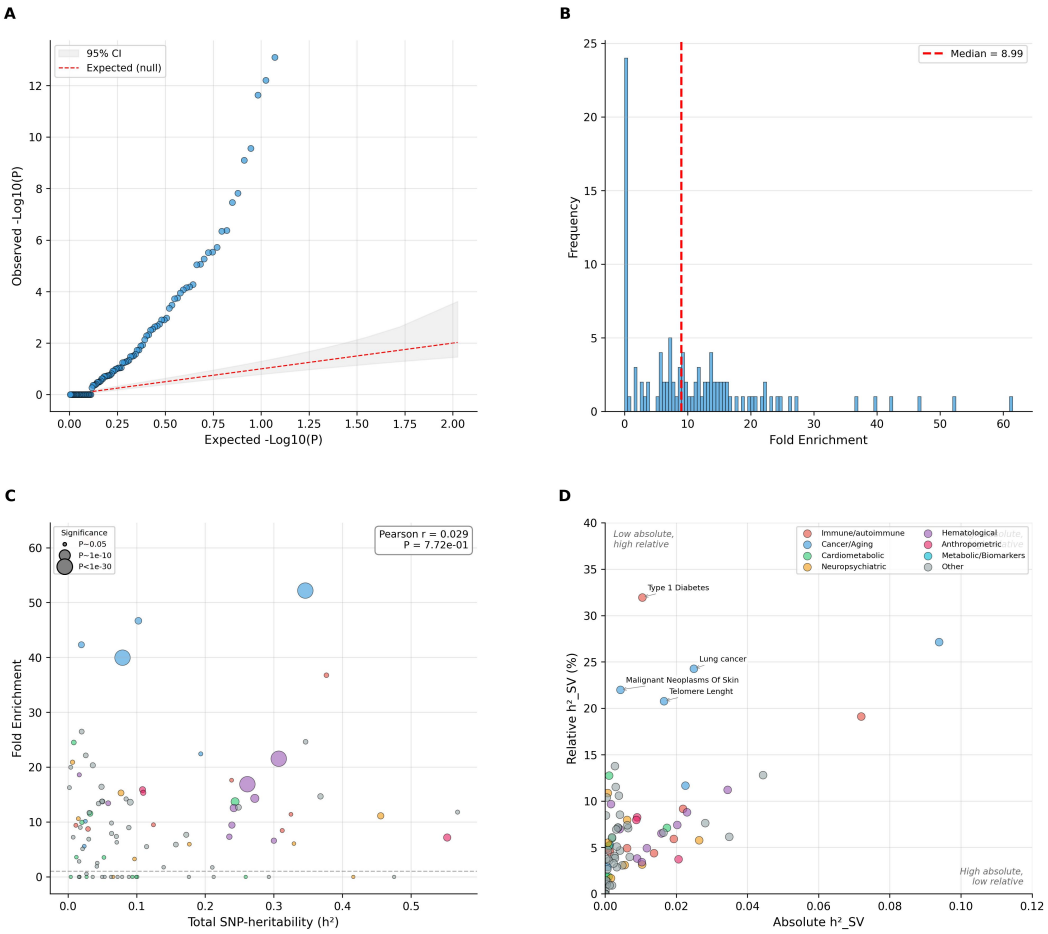
Supplementary Figure S3: **SE calibration of MiXeR-SV enrichment estimates across LD reference panel configurations.** Scatter plots of RMS reported SE (y-axis) versus empirical SD across 30 fixed-β replicates (x-axis) for SV enrichment estimates, stratified by LD reference panel: (A) EUR HGSVC3 full panel ($N = 489$; LD489), (B) EUR HGSVC3 $N = 300$ (LD300), (C) EUR HGSVC3 $N = 150$ (LD150), and (D) EAS HGSVC3 panel ($N = 481$) applied to EUR-simulated phenotypes (LDEAS). Each point corresponds to one (h^2, enrichment) simulation setting, colored by heritability level ($h^2 = 0.1$ blue, $h^2 = 0.3$ pink, $h^2 = 0.5$ green). The dashed diagonal line indicates perfect calibration (RMS SE = empirical SD). In the fixed-β design, genetic effect sizes are held constant across replicates so that only environmental noise (ϵ) varies, matching the assumption of the Hessian-based SE estimator. Points close to the diagonal indicate well-calibrated uncertainty estimates. Supplementary Table S3 provides the full numerical summary for both the random-β (point estimate) and fixed-β (SE calibration) simulation designs.



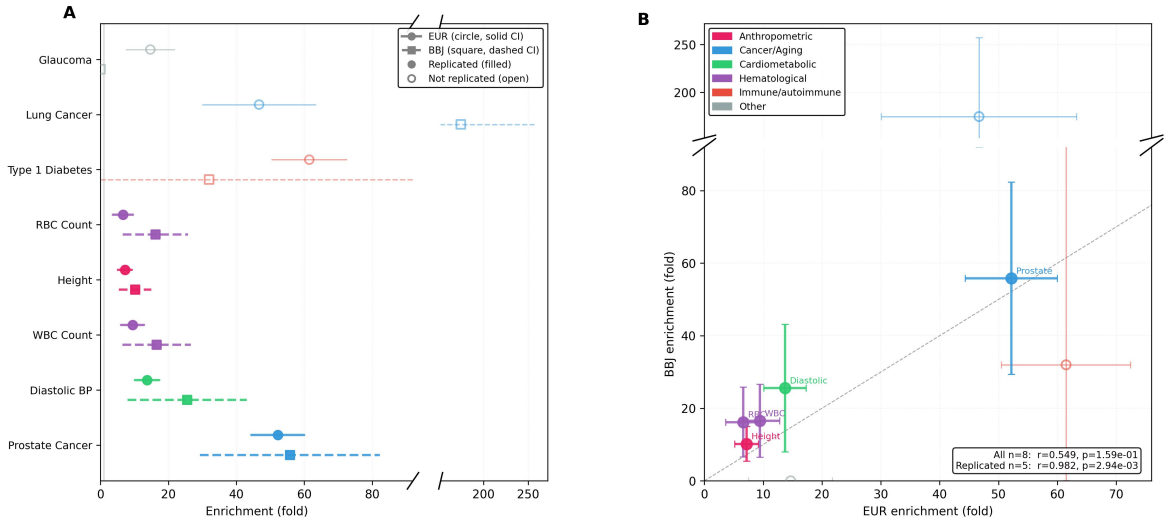
Supplementary Figure S4: **MiXeR-SV SV enrichment estimation across GWAS sample sizes.** Box-and-jitter plots of estimated SV enrichment versus true enrichment for 30 simulation replicates, across three true enrichment levels (10×, 20×, 50×; columns) and five GWAS sample sizes (rows): N = 10,000, 20,000, 30,000, 40,000, and 50,000. Simulations used $h^2 = 0.1$, 1% causal variant proportion, and the EUR HGSVC3 full panel (N = 489; LD489). Black horizontal bars indicate the median estimate across replicates. Supplementary Table S4 provides the full numerical summary.



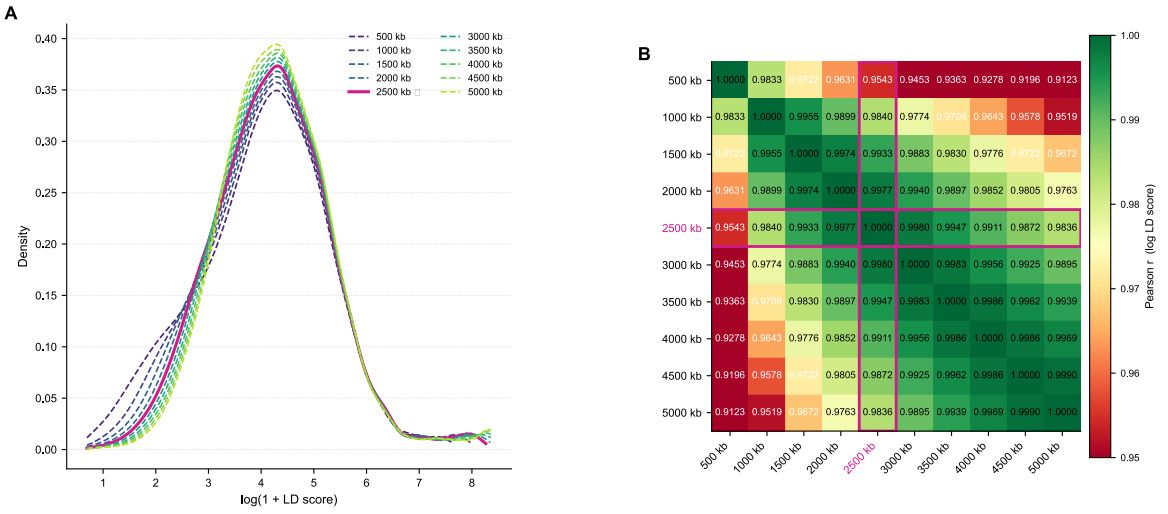
Supplementary Figure S5: **Mean estimated SV enrichment with 95% CIs across GWAS sample sizes.** Line plots of mean estimated SV enrichment (point) ± 95% CIs (shaded ribbon) across 30 simulation replicates, stratified by true enrichment level (10×, 20×, 50×; panels) and GWAS sample size (N = 10,000–50,000; x-axis). Simulations used $h^2 = 0.1$, 1% causal variant proportion, and the EUR HGSVC3 full panel ($N = 489$; LD489). Dashed horizontal lines indicate the true enrichment values. The shaded bands illustrate how estimation uncertainty narrows with increasing GWAS sample size, complementing the replicate-level distribution shown in Supplementary Figure S4.



Supplementary Figure S6: **Extended analysis of SV enrichment patterns and heritability partitioning.** **(A)** Quantile-quantile plot of enrichment P-values across all 105 traits, comparing observed versus expected distributions under the null hypothesis of no enrichment. Strong deviation from the identity line (red dashed) indicates systematic enrichment signal beyond chance expectation. Gray shading represents 95% CI for the null distribution. **(B)** Distribution of fold enrichment values across all 105 analyzed traits, showing a median of 8.99-fold (red dashed line) with a long tail of extreme effects concentrated in specific trait categories. **(C)** Scatter plot of fold enrichment versus total variant heritability for all 105 traits. Point sizes scale with statistical significance ($-\log_{10} P$); colors indicate trait category. Pearson correlation ($r = 0.029$, $P = 7.7 \times 10^{-1}$) demonstrates that enrichment is independent of total variant heritability, reflecting trait-specific genetic architecture rather than statistical power. **(D)** Absolute versus relative SV heritability contributions reveal distinct biological scenarios. X-axis shows absolute h^2_{SV}; y-axis shows the percentage of total heritability from SVs. Type 1 diabetes exemplifies the top-left quadrant (low absolute, high relative: 32.0%), while anthropometric traits occupy the bottom-right (high absolute, low relative), reflecting their highly polygenic backgrounds.



Supplementary Figure S7: **Cross-ancestry replication of SV enrichment in Biobank Japan.** All eight overlapping traits are shown; five replicated (filled markers) and three did not meet replication criteria (open markers, lighter error bars). **(A)** Forest plot of EUR discovery (circles, solid 95% CI) and BBJ replication (squares, dashed 95% CI) enrichment estimates for all eight traits. The x-axis is broken (//) to accommodate the extreme Lung Cancer BBJ CI (∼175-fold, upper CI ∼257-fold). Traits are ordered with non-replicated traits at top. **(B)** Scatter plot of EUR versus BBJ fold-enrichment for all eight traits with 95% CIs. The y-axis is broken (//) to show Lung Cancer BBJ (∼175-fold) without compressing the main cluster. The dashed line indicates the identity ($y = x$). Pearson correlation among all eight traits: $r = 0.549$, $P = 0.16$; among replicated traits only: $r = 0.982$, $P = 2.9 \times 10^{-3}$, $n = 5$. See Supplementary Table S9 for complete numerical results. Colors indicate trait categories.



Supplementary Figure S8: **LD-score sensitivity to genomic window size.** Window lengths from 500 kb to 5,000 kb were evaluated using chromosome 10 data from the EUR population in the ONT dataset (480,076 SNPs common to all windows). **(A)** Distribution of log(1+LD score) for each window size. As the window widens, the density curves shift only marginally and converge, with successive curves nearly overlapping beyond 2,500 kb (bold). **(B)** Pairwise Pearson correlation (r) of log LD scores across window sizes. Correlations relative to the focal 2,500 kb window (highlighted) exceed 0.998 for all neighbouring windows ($\geq$2,000 kb), confirming that LD scores stabilize at 2,500 kb and that wider windows add negligible additional information.