

Supplementary Information

Candidate Pharmacogenomic and HLA Risk Alleles for Drug-Induced Liver Injury in Vietnamese Cancer Patients: A Pre-Specified Case-Control Study

Van Thai Nguyen, Thi Hang Tong, Thi Thanh Huyen Tran, Thi Thu Hien Tran, Thi Huyen Dao, Thao To Pham, Tien Thinh Nguyen, Vinh C. Duong, Huong T. Vuong, Nguyen Anh Vu Mai, Nam S. Vo, Thi Ly Le, Huu Song Le

Summary

This file contains supplementary methods detail, robustness analyses, one supplementary table, and twelve supplementary figures covering secondary and exploratory pre-specified hypotheses (H2, H4-H9, H13-H15) and methods-validation content that support, but are not required to interpret, the primary findings reported in the main text. All citations use the numbering of the main-text reference list.

Supplementary Methods

Second imputation branch (Vietnamese native-panel reference). A second, independent branch ran the same two-step phase-then-impute procedure against a Vietnamese reference panel, executed at the genome center that holds that panel, on genotypes we had already quality-controlled. The input was our quality-controlled cohort callset: 211 samples and 37 306 biallelic GRCh38 variants, split by chromosome, with per-sample sex derived from chromosome X heterozygosity rather than from the clinical record. We applied no minor-allele-frequency filter to that input; 12 171 of those variants sit below MAF 0.01, and because most loss-of-function pharmacogene alleles are rare, filtering at input would remove the variant class of interest. The data provider ran this branch, estimating haplotypes with SHAPEIT4 v4.2.2 (ref. 23) and performing imputation with Minimac4 v4.1.6 (ref. 22) on GRCh38; no run record on file states its phasing seed. This branch did not carry

array-typed positions through the imputer: we spliced them back afterwards as the complement of the typed callset, then concatenated them with the original typed VCF, so the spliced typed records carry the original array genotype fields rather than a computed imputation R2 and are not subject to the $R2 \geq 0.30$ gate that removes 1 036 low-quality typed records on the reporting branch. The reference was the 1 008-genome intermediate freeze of the VN1K Vietnamese whole-genome resource ¹⁴ merged with 1000 Genomes ²⁴; the merge is that panel's own published design. The provider returned output as a single tier filtered on imputation R2 alone ($R2 \geq 0.30$); we apply the $MAF \geq 0.01$ gate downstream at the analysis stage, as on the reporting branch. This branch returned 17 090 493 records across chromosomes 1-22 for all 202 analytic samples and served as the sensitivity axis for H1, H2, and H5. It does not cover chromosome X, which on the reporting branch comes from a separate dedicated run; no candidate-panel variant falls on chromosome X, so the sensitivity axis covers the candidate panel in full. Neither branch's phasing step is seed-reproducible: we ran the reporting branch's Eagle2 run and the Vietnamese-panel branch's SHAPEIT4 run each once, with no recorded seed; every analysis in this study anchors to the frozen, checksummed imputation output we used throughout rather than to a re-runnable pipeline.

The merge with 1000 Genomes is load-bearing rather than cosmetic. The native panel alone carries 2016 haplotypes and its build removes singletons before phasing, so we cannot impute panel-level singleton alleles at any coverage. When run without the merge on this array, it returned 1.05 million imputed records and placed a record at only 48 of the 93 candidate-panel positions, against 17.09 million records and 92 of 93 once merged (13.53 million and 93 of 93 on the reporting branch); we therefore did not use the unmerged output for any analysis. Of the 93 variants on the as-run candidate panel, the array types 27 directly and imputation supplies 66, the latter at a median imputation R2 of 0.591 (27 below 0.5, 52 below 0.8). Because the effective sample size for an

imputed variant scales roughly with R^2 , we compute the detection limits reported throughout on genotype counts without an imputation-quality correction, so they run correspondingly optimistic for imputed variants; the supplementary table and Table 3 give per-variant R^2 .

Array-based pharmacogenomic calling caveats. We could not call CYP2D6 phenotype because the array carries no copy-number probes or common East Asian reduced-function alleles; PyPGx served as the structural/copy-number caller for CYP2D6 where a call was possible. We imputed UGT1A1*28, a promoter repeat with no on-array probe, and validated it against the on-array tag SNP rs887829: complete concordance on the reporting branch (0 of 202 samples discordant; cohort frequency 0.1224) and 4 of 202 samples discordant on the Vietnamese-panel branch. For TPMT, we measure *3C (no probe) only through imputation. The TPMT probes on the array illustrate why we read star-allele calls from the array genotypes rather than the imputed callset: both *2 (rs1800462) and *3B (rs1800460) are on-array and absent from the imputed VCF, but for different reasons. We call *2 in all 211 samples at 100% call rate; it is invariant (imputation $R^2=0$, so the $R^2 \geq 0.30$ filter removes it; the East Asian reference panel is likewise invariant there, so this is a genuine absence of variation). The variant-level call-rate filter instead removes *3B, which is polymorphic (one heterozygous carrier) and missing in 8 of 211 samples. PharmCAT read the array genotypes directly, so neither removal reached the clinical calls: 209 of 211 samples are *1/*1, one is *1/*3A or *1/*3B, and one is *1/*20.

HLA imputation caveats. We imputed HLA alleles from the array's chromosome 6 MHC SNPs against the same merged native reference panel we used for the variant sensitivity branch, using an in-house tool rather than a published caller. We attempted HIBAG, but it did not yield a usable comparison: every URL for a GRCh38 model returned HTTP 403, and the model available on disk was an hg19 Illumina model that shared no SNPs with this array at HLA-A and DRB1 (no

overlapping SNPs), while at HLA-B the maximum posterior reached was 0.066, with no call reaching the 0.7 threshold. We applied neither the 1000 Genomes East Asian panel nor the Eagle2/Minimac4 route to HLA, so the $R^2 \geq 0.30$ post-imputation filter of the variant branch does not apply here, and HLA calls carry a posterior probability rather than an imputation R^2 . We made calls at two-field resolution across 7 loci in all 202 samples, defining call confidence as the smaller allele posterior at heterozygous loci and the single posterior at homozygous loci. We applied a primary threshold of ≥ 0.5 (the HIBAG default,^{31,32}) and report a ≥ 0.7 sensitivity pass alongside. The primary threshold itself already excludes 31.7% of samples at locus B (138 of 202 retained); raising it to ≥ 0.7 excludes a further 15.2% of those retained calls (21 of 138), bringing cumulative loss at that locus to 42.1% of all 202 samples - the locus carrying HLA-B*57:01 and HLA-B*58:01. We measured concordance with an independent caller, CookHLA³³, on the 190 samples the two callers share, as Cohen's kappa over sorted allele pairs. At two-field resolution, kappa was 0.807 at A, 0.856 at B, 0.967 at C, 0.851 at DRB1, and 0.962 at DQB1 (lowest at A, not at DRB1); at one-field resolution the same comparison gives 0.937, 0.888, 0.988, 0.939, and 0.980. Confidence in the 187 homozygous calls is a ceiling rather than a point estimate, and the HLA imputation engine is unpublished, proprietary, and carries no version identifier; no one has independently benchmarked it in a Kinh-ancestry population, and the CookHLA comparison is the only external check available for this branch.

Genomic inflation as a descriptive statistic, not a stopping gate. We reported the genomic inflation factor (λ_{GC}) descriptively at every tier but never used it as a stopping gate, since a candidate panel selected for expected effect directly violates its validity assumption - that most tested markers are unassociated with the phenotype; we instead controlled population stratification through the mandatory PC1-10 covariates. An early run of the exploratory genome-wide scan (H9)

had shown lambdaGC exceeding 1.10 across all six sample subsets before we enforced true Firth estimation for every variant; correcting this gave lambdaGC=0.971 for that scan (a separate figure from the lambdaGC=0.7787/1.0292 we report for the candidate panel, H1, in the main text), confirming the inflation was an estimation artifact, not population stratification.

Multiple-testing family definition and amendment history. The pre-declared hypothesis set defines a testing family;⁵⁸ we fixed denominators at pre-specification rather than on the number of variants surviving cohort-specific filters, which is conservative. The 2 June 2026 lock carried M=63 for H1; amendment v1.1 set M=115 on 14 June 2026, and amendment v1.6 set H3's M=24. A documented waiver permitted the 2026-08-15 re-analysis pass to proceed on six named tiers, including the primary hypothesis H3, without a prior amendment; the numbers we report for H3 throughout are post-waiver.

Full case exclusion criteria. All cases underwent protocol-specified exclusion of viral hepatitis (HBV, HCV, HAV, HEV), autoimmune hepatitis, Wilson disease, hemochromatosis, alcoholic cirrhosis, and primary or secondary liver malignancy; we applied these exclusions to cases only (see main text Limitations).

Genotype-based sex determination and reconciliation. We derived sex for every model from chromosome X heterozygosity outside the pseudoautosomal region rather than from clinical records. The distribution was cleanly bimodal, with no sample falling between the two bands, indicating unambiguous genotype-based sex calls throughout. Three samples had a clinical-record sex in conflict with genotype-inferred sex. We traced two to a single sample-identity swap between two adjacent wells during plate preparation; correcting the sample-ID-to-genotype linkage restored agreement for both. Two independent, orthogonal methods called the third sample female - array probe intensity (X/Y mean-signal ratio 0.559, consistent with the female background level) and

chromosome X genotype (PLINK --check-sex F=0.018 and 0.052 across two independent computations) - indicating the clinical record (male) was in error. After correction, this sample's pair is the only one of 101 that no longer matches on sex; because sex is a mandatory covariate in every model, and the conditional model does not require within-pair equality of matching variables to remain valid, this does not bias the associations we report.

Post hoc hepatotoxicity-annotated panel (H7). Because CPIC does not index any of the four key-panel gene-drug pairs to hepatotoxicity as its primary outcome - in the frozen CPIC gene-drug pair table used throughout (accessed 14 August 2026) (ref. 9), NUDT15 is paired with the thiopurines, SLCO1B1 with the statins, CYP2C19 with clopidogrel/proton-pump inhibitors/several antidepressants/voriconazole, and UGT1A1 with irinotecan and atazanavir (neutropenia and hyperbilirubinemia, not liver injury) - we built a second, post hoc key set by filtering the candidate panel to variants carrying a PharmGKB clinical annotation in the toxicity category that explicitly names a liver-injury phenotype. This retained eight variants, in ABCB1, GSTP1, TNF, MTHFR, NR1H2, CYP3A5, TYMS, and DHFR. Because the annotation source carries no allele-direction column, we could not assign a risk-direction to build a combined burden score for this set, so we report each component individually rather than as an aggregate (results in Supplementary Results). Of the four CPIC key-panel index drugs, only irinotecan appears in this cohort (an irinotecan-containing regimen in 3 cases and their 3 matched controls, 6 of 202 samples), too few to support a drug-specific analysis; H7 therefore measures dose-adjustment-relevant burden in general, not hepatotoxicity-specific burden.

Exploratory-tier methods (H6, H8-H14). We used elastic-net regression via glmnet⁵⁹ and discrimination via pROC⁶⁰ for H8, running two separate branches: a 109-feature elastic-net summarized in Supplementary Table S1, and a 1 203-feature stability-selection branch whose

within-pair label-permutation null (2 000 permutations, FWER threshold 0.85) appears in Supplementary Figure S10; neither branch selected a feature. We used gene-level collapsing burden and SKAT-O⁶¹ and gene-set enrichment via MAGMA⁶² for H10-H11; Bayesian fine-mapping via SuSiE⁶³ for H12; and SNP heritability via GCTA⁶⁴ for H14. Two methodological caveats apply to the chromosome X branch (H13): it excludes one QC-flagged sample with no chrX calls at all, leaving 201 samples, and we built it on a pseudo-diploid contig without separating the pseudoautosomal region from the non-PAR region or encoding X-inactivation, so its imputation-quality metrics and lambdaGC do not compare to the autosomal analyses; we defer a methodologically adequate chrX scan to a separately pre-specified analysis. The heritability estimate (H14) did not incorporate pair-matching, and we report it as a curve over plausible population prevalence (K=2.5%-29%, refs. 65-67) rather than a point estimate; at every point on that curve the standard-error-to-estimate ratio is 4.58, making it uninterpretable regardless of K.^{68,69}

Supplementary Results

Post hoc hepatotoxicity-annotated panel (H7) - full results. The eight-variant post hoc panel described in Supplementary Methods (ABCB1, GSTP1, TNF, MTHFR, NR1I2, CYP3A5, TYMS, DHFR) showed no component reaching its Bonferroni threshold; the smallest raw association was rs1800629/TNF (OR 2.371, 95% CI 0.945-6.399, p=0.066, Bonferroni-corrected p=0.530). Because the annotation source carries no allele-direction column, we could not assign a risk-direction to build a combined burden score for this set, so we report it component-by-component rather than as an aggregate, and we do not pool it with the four-allele CPIC key panel we report in the main text.

Vietnamese pharmacogene allele-frequency profile (H2) - extended comparison. Of the three alleles differing from the gnomAD v4.1 East Asian reference (CYP2D6*10, CYP2C19*3, UGT1A1*6), we compared the same three alleles against the whole-genome-sequenced

Vietnamese reference VN1K¹⁴ (n=1 011). CYP2D6*10 agreed closely (cohort 0.628 vs VN1K 0.634, counting the *36+*10 tandem haplotypes as *10); UGT1A1*6 showed the smallest gap (cohort 0.099 vs VN1K 0.104); CYP2C19*3 showed the largest (cohort 0.040 vs VN1K 0.053). We derived these comparison values from the star-allele calls the VN1K release ships, generated with PyPGx at caller_version 0.28.0 as declared in the provider's release files, a version number no public PyPGx release carries; cross-checked against the population-level frequencies reported for the same three alleles in genomes.vn,⁵⁷ a separate whole-genome Vietnamese resource (406 unrelated Kinh individuals): of the three alleles, only UGT1A1*6 is represented in that database, at 0.0779; rs1065852 (CYP2D6*10) and rs4986893 (CYP2C19*3) are absent from it, so no genomes.vn comparison is available for those two. Across the 40 shared panel variants with allele frequency ≥ 0.05 , VN1K and genomes.vn themselves differ by a mean of +0.016 (95% CI +0.005 to +0.027, paired t p=0.004), and equivalence between those two references within a +/-0.010 margin is not established (TOST p=0.87). The cohort's differences from gnomAD at CYP2C19*3 and UGT1A1*6 therefore sit inside the spread between two independent Vietnamese references; we did not run an equivalence test of the cohort against gnomAD. One methodological caveat applies throughout this comparison: the cohort values are single-rsID allele frequencies, while the VN1K values are star-allele/haplotype frequencies that a different caller (PyPGx) generated, so where VN1K reports a like-for-like single-variant frequency the gap is wider than the haplotype-level comparison suggests. Differences of this size sit within the spread between two independent Vietnamese references (VN1K and genomes.vn) - up to 0.026 at UGT1A1*6 between those two alone - though we did not establish equivalence within a +/-0.010 margin. We therefore read the agreement observed here as consistency supporting the genotyping and imputation pipeline, not as a novel population-frequency finding or a precision estimate.

Ancestry validation (H15) - full detail. This tier ran on two independent axes with different verdicts. On the pre-specified primary axis - principal-component cluster position against the 1000 Genomes East Asian panel - the cohort centroid sat 0.00269 from the 99-sample Kinh (KHV) reference centroid, versus 0.04912 from the other East Asian populations combined (486 samples), an 18-fold difference in distance (Supplementary Figure S6); we plot the full PCA in Supplementary Figure S6. On the secondary, allele-frequency axis, the cohort showed no corrected-significant difference from the 99-sample Kinh (KHV) reference across the 7 tested candidate-panel alleles carrying a KHV comparison value (smallest raw $p=0.170$, Bonferroni-corrected $p=1.0$), and a two-one-sided equivalence test found no variant reaching its pre-specified margin at this sample size - an underpowered comparison rather than a demonstration of equivalence. Two pre-registered negative controls ran on this allele-frequency axis rather than on the PCA axis: 1 of 7 variants reached corrected significance against Han Chinese (CHB) and 1 of 6 against West Javanese, while the cohort's own comparison against KHV reached 0 of 7. The axis therefore retains some discriminative ability, but at one variant per control it is a weak demonstration. What separates the two axes is the corrected-significance count on the allele-frequency comparison (0 of 7) against the clear, quantitatively large separation on the PCA axis.

A sixth comparison family. We added a sixth comparison family - the cohort against a set of 101 unrelated KHV genomes drawn from the high-coverage 1000 Genomes release²⁰ - post hoc on 2026-09-23, under an amendment we signed that day; it carries its own denominator ($M=7$), and we do not pool it with the five pre-specified families, and it does not enter the negative-control gate. This family uses 101 rather than all 122 KHV genomes because 21 of the 122 are offspring whose parent is also in the panel (trios the high-coverage release added); including them would raise the reference allele number to 244 and narrow the detectable difference spuriously (measured:

median minimum detectable allele-frequency difference of 0.1314 at the 101 unrelated genomes versus 0.1257 at all 122, at 80% power on the primary control-only branch). None of the 7 tested alleles in this family reached Bonferroni significance (smallest raw $p=0.152$, corrected $p=1.0$).

PCA cluster position. On the primary PCA axis, joint principal-component analysis of the cohort with the 1000 Genomes East Asian panel placed the cohort centroid 0.00269 from the 99-sample pre-specified KHV comparison group and 0.04912 from the other East Asian populations combined (486 samples; Supplementary Figure S6), a pooled centroid that therefore bounds the separation from any single population from above. Because the panel carries 122 KHV genomes in total, 23 sit inside the comparison group and pull it toward the cohort, making the pre-specified comparison conservative. A sensitivity branch anchored on all 122 KHV genomes (rather than the pre-specified 99) gives 0.00258 versus 0.05173 - materially unchanged. We can state one evidence-supported conclusion: the cohort's genetic cluster position matches the Kinh population, established on this PCA axis rather than the underpowered allele-frequency axis.

Exploratory tier (H6, H8-H14) - full results. We pre-specified eight hypotheses in full, but they are underpowered at $n\sim 200$ on a targeted panel, so we report them for hypothesis-generation only. The H8 elastic-net branch (109 features, Supplementary Table S1; out-of-sample AUROC 0.380, 95% CI 0.303-0.458, and a per-repeat mean of 0.464) and the separate 1 203-feature stability-selection branch (Supplementary Figure S10; maximum selection frequency 0.690 against a permutation-null mean of 0.672, global permutation $p=0.430$) both selected zero features. The genome-wide scan (H9, exploratory) found zero genome-wide-significant signals and ten suggestive ones, falling in four positional clusters (Supplementary Figures S10-S12); these are groupings by physical position only, since we ran no linkage-disequilibrium clumping, and readers should not read them as four independent signals. Of the 9 123 220 variants tested, 3 266 Firth fits

did not converge, and we exclude them from the p-value distribution; the remaining 9 119 954 variants with a valid, converged p-value are the ones we plot in Supplementary Figure S11. Genotyped/imputed concordance was 97.7% direction-concordant, supporting technical validity despite the expected null on a targeted array. Supplementary Methods gives caveats for H13 (chromosome X) and H14 (heritability).

Detection limits and robustness. At 101 cases and 101 controls with 80% power, this design resolves odds ratios of 2.81 (nominal) and 4.37 (Bonferroni-corrected at $M=115$) at MAF 0.05, and 6.45/12.38 at MAF 0.01 (Figure 3); we compute these detection limits on genotype counts without an imputation-quality correction, so they run correspondingly optimistic for the 66 imputed panel variants. A pre-specified rule specified that if the corrected MDOR at $MAF \geq 0.10$ exceeded 6, we would reframe the multivariable elastic-net/discrimination branch we report under H1 as exploratory; it did (6.17, computed at Bonferroni $M=91$ on an earlier 91-variant panel build), though this cost no findings since the penalized and stability-selection analyses had already selected zero variants. The selection result was re-confirmed on the final 93-variant panel on 2026-09-27, which again selected zero variants; the within-pair concordance figures we report under H1 still come from the earlier 91-variant build, which is the registered sensitivity panel for this branch.

Subsampling learning curve. A subsampling learning curve for the gradient-boosted-tree branch (histogram gradient boosting; the LightGBM curve is materially identical) yields no estimate of the sample size needed for stable feature selection: mean within-pair concordance rises modestly and non-monotonically from 0.500 at 20 pairs to 0.553 at 60 pairs and 0.550 at the full 101 pairs, with wide, overlapping confidence intervals at every subsample size; we present this curve descriptively, not as evidence that a specific larger sample would yield feature selection.

Every hypothesis tier was re-run across six sample subsets. We re-ran every hypothesis tier across six sample subsets, and no pre-specified conclusion changed. We ran the second imputation panel only for the candidate-variant tiers (H1, H2, H5) and for the tier-1 SNP biomarker family below; the exploratory genome-wide, chromosome X, gene-burden, and heritability scans never ran on it and rest on the reporting branch alone. On the candidate panel, all 93 variants carry an estimate on the reporting branch and 92 on the Vietnamese-panel branch (one candidate-panel position present on the reporting branch's 1000 Genomes-based reference is absent from the merged Vietnamese-panel reference); median imputation R^2 over the imputed panel subset was 0.591 on the reporting branch (66 variants) and 0.603 on the Vietnamese-panel branch (65 variants), so the Vietnamese-merged reference imputed this panel marginally better despite carrying fewer haplotypes than full 1000 Genomes.

Tier-1 SNP biomarker head-to-head. We also compared the two branches head-to-head on a tier-1 SNP biomarker family of 67 candidate variants, computing multiplicity correction separately per branch rather than pooling it. The Vietnamese panel retrieved more of that set than the public one: 45 of 67 positions estimable on the Vietnamese-panel branch versus 39 of 67 on the reporting branch (Bonferroni $M=42$ and $M=37$ respectively, reflecting each branch's own retrieval count). The smallest raw p -value was 0.021 on the reporting branch and 0.010 on the Vietnamese-panel branch, neither surviving correction; genomic inflation on this subfamily was $\lambda=1.16$ (reporting branch) and $\lambda=0.87$ (Vietnamese-panel branch). Of the 67 candidates, 36 were estimable on both branches, with cohort allele frequencies agreeing closely between branches (median absolute difference 0.025, maximum 0.073, with 9 of 36 variants differing by more than 0.05).

Direction-of-effect instability. Direction of effect is markedly less stable across panels for imputed positions than for typed ones: of the 92 candidate variants estimable on both branches, 21 of the

65 imputed ones reverse the direction of their odds ratio between panels, against 0 of the 27 directly-typed ones, including reversals at R2 as high as 0.78-0.93. One reversing variant, rs13181, goes from OR 0.391 (95% CI 0.138-1.025) on the reporting branch to OR 1.208 (95% CI 0.465-3.201) on the Vietnamese-panel branch - both non-significant after correction, and the reversal reflects the small discordant-pair count (9 and 17 carriers, as noted in the main Results) rather than a meaningful shift in evidence. No reversing variant reaches significance after correction on either branch, so no conclusion turns on the direction of any of them, but the instability is a property of the imputed estimates themselves rather than a small tail of poorly measured ones. That the two branches disagree on which variants are estimable, but agree on every conclusion, is the result we designed the sensitivity axis to produce.

Supplementary Table S1. Exploratory-tier results (H6, H8-H14), reported for hypothesis-generation only

Hypothesis	Result	Interpretation
H6 - allele $\times$ sex/age interaction	0/186 reached threshold ($\alpha=2.69\times 10^{-4}$); λ_{GC} 0.985 (sex axis), 0.946 (age axis); smallest raw $p=0.0125$ (rs1544410) and 0.0243 (rs2236624)	Interaction model is underpowered; the sex axis has only 1 discordant pair, so the conditional model is inestimable
H8 - machine-learning prediction	Elastic-net selected 0/109 features; AUROC 0.464 (mean across iterations), 0.380 out-of-sample. A separate 1 203-feature stability-selection branch retained none (Supplementary Figure S10)	Below chance; permutation null (global $p=0.430$, max selection freq. 0.690 vs null 0.672) in Supplementary Figure S10
H9 - genome-wide association scan	9 123 220 variants tested, 9 119 954 valid; λ_{GC} 0.971; 0 reached 5×10^{-8} ; 10 reached suggestive 10^{-5} (four positional clusters); lead signal 20:53 274 212, OR 0.144, $p=2.91\times 10^{-6}$	Absence of genome-wide signal is the expected result on this targeted array

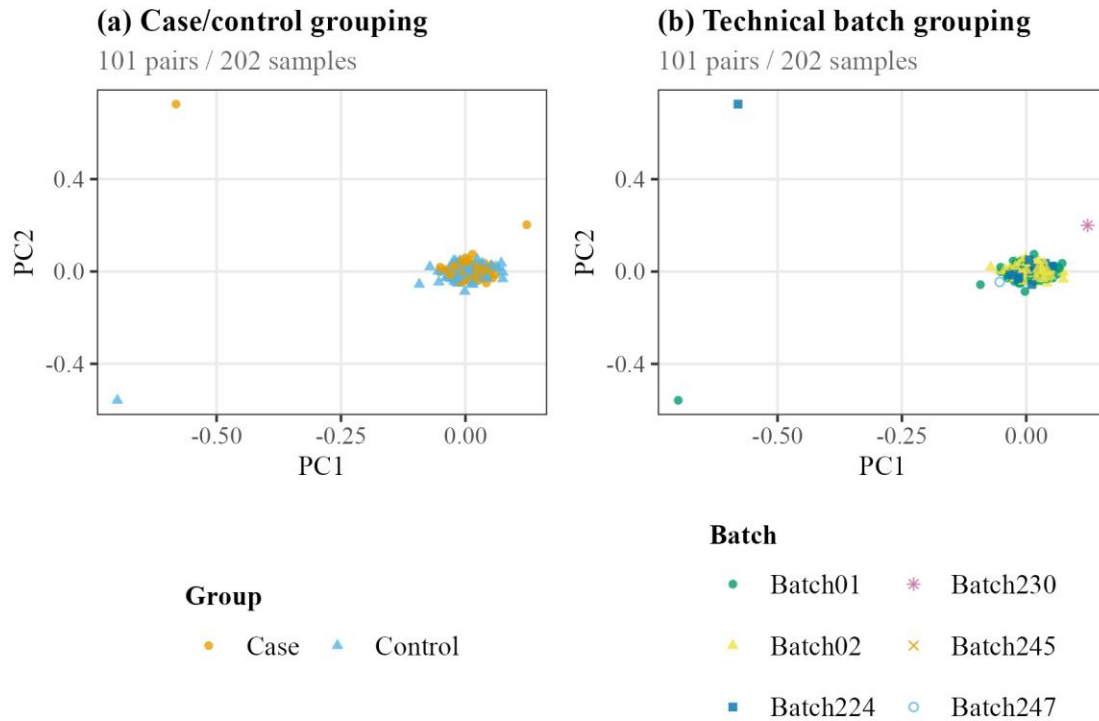
H10 - gene-level burden	0/64 genes by both collapsing burden and SKAT-O; lead gene CYP2D6 (raw $p=0.0534$ and 0.0312); smallest $q=0.799$ and 0.896	Downstream of an already-null candidate scan
H11 - gene-set / pathway enrichment	0/4 pre-declared gene sets; lead set (mitochondrial) $p=0.100$, $\beta=0.289$	Denominator counts pre-declared gene sets, not surviving ones
H12 - fine-mapping of lead loci	No 95% credible set identified; maximum posterior inclusion probability 0 in the chr20 lead window and 0.006 in the MHC window	No locus reached genome-wide significance, so there is nothing to fine-map
H13 - chrX association scan	143 608 variants tested (143 390 with a valid p); 0 reached threshold; λ_{GC} 0.858	See chrX caveat in Supplementary Methods (pseudo-diploid contig, no PAR/non-PAR separation)
H14 - SNP heritability	$V(G)/V_p$, observed scale: 0.3081 (SE 1.4100, $p=0.415$)	Uninterpretable at this sample size; see Limitations

Notes: We pre-specified all eight hypotheses in this tier in full, but they are explicitly exploratory at this sample size ($n \approx 200$ on a targeted array), and we report them for hypothesis-generation only, not as confirmatory findings; see Supplementary Methods and Supplementary Results above for full methodological caveats on H13 and H14. Exploratory-tier methods: elastic-net regression via glmnet and discrimination via pROC (H8); gene-level collapsing burden and SKAT-O and gene-set enrichment via MAGMA (H10-H11); Bayesian fine-mapping via SuSiE (H12); SNP heritability via GCTA (H14).

Abbreviations: MAF, minor allele frequency; SE, standard error; λ_{GC} , genomic inflation factor; PAR, pseudoautosomal region.

Supplementary Figures

Figure 1. Principal component analysis (PCA) of population structure



202 samples (101 cases, 101 controls) after QC filtering; Affymetrix Axiom VinGen V1 array; no clear stratification (flat PC spectrum).

Genome-wide $\lambda_{GC} = 1.110$ is explained by the small sample size, not by population stratification (see the covariate-adjusted model, PRIMARY).

Supplementary Figure S1. Principal component analysis of the cohort, first two axes.

Imputation quality — R^2 distribution by MAF bin

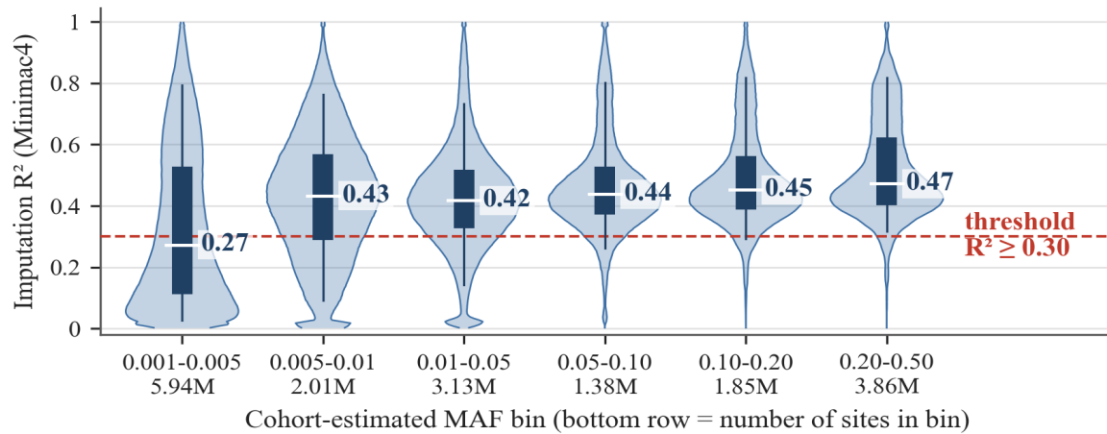
Autosomal sites from the cohort imputation, split typed versus imputed · analysis
threshold $R^2 \geq 0.30$

18,179,821 imputed sites on chr1-22 with $MAF \geq 0.001$, of which 13,472,714 reach $R^2 \geq 0.30$ (74.1%).

Not plotted in panels (a)-(b): 52,543,935 sites with $MAF < 0.001$ (near-monomorphic in the cohort) and 25,520 typed sites on the array.

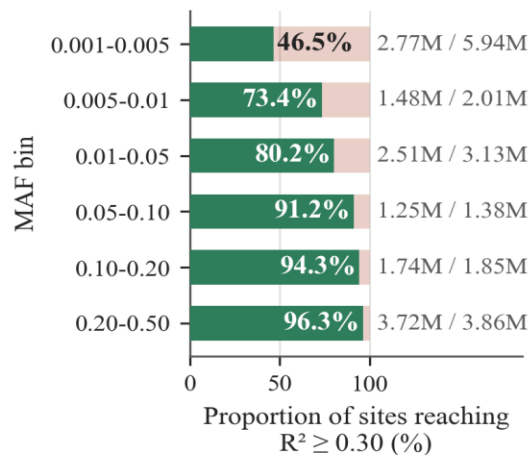
(a) R^2 distribution by MAF bin — 18,179,821 imputed sites (chr1-22)

box = interquartile range · white bar = median · whiskers = 5th-95th percentile



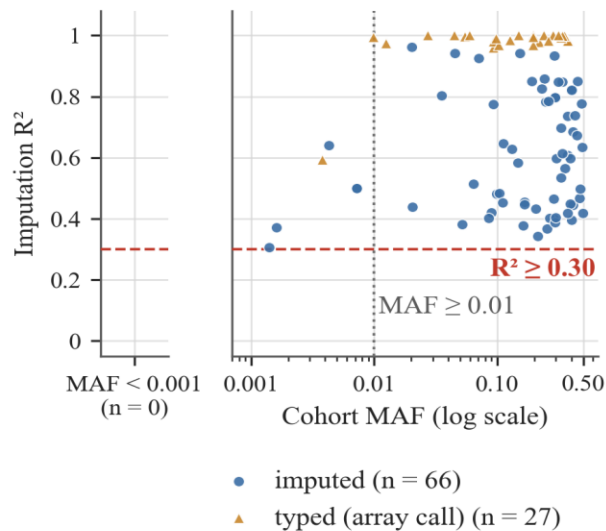
(b) Proportion reaching threshold by MAF bin

number on the right = sites passing / total sites in bin



(c) Candidate oncology-PGx panel variants — 93 variants, not genome-wide

R^2 threshold applies to the imputed group only:
66/66 pass · typed variants use the array call, not imputation

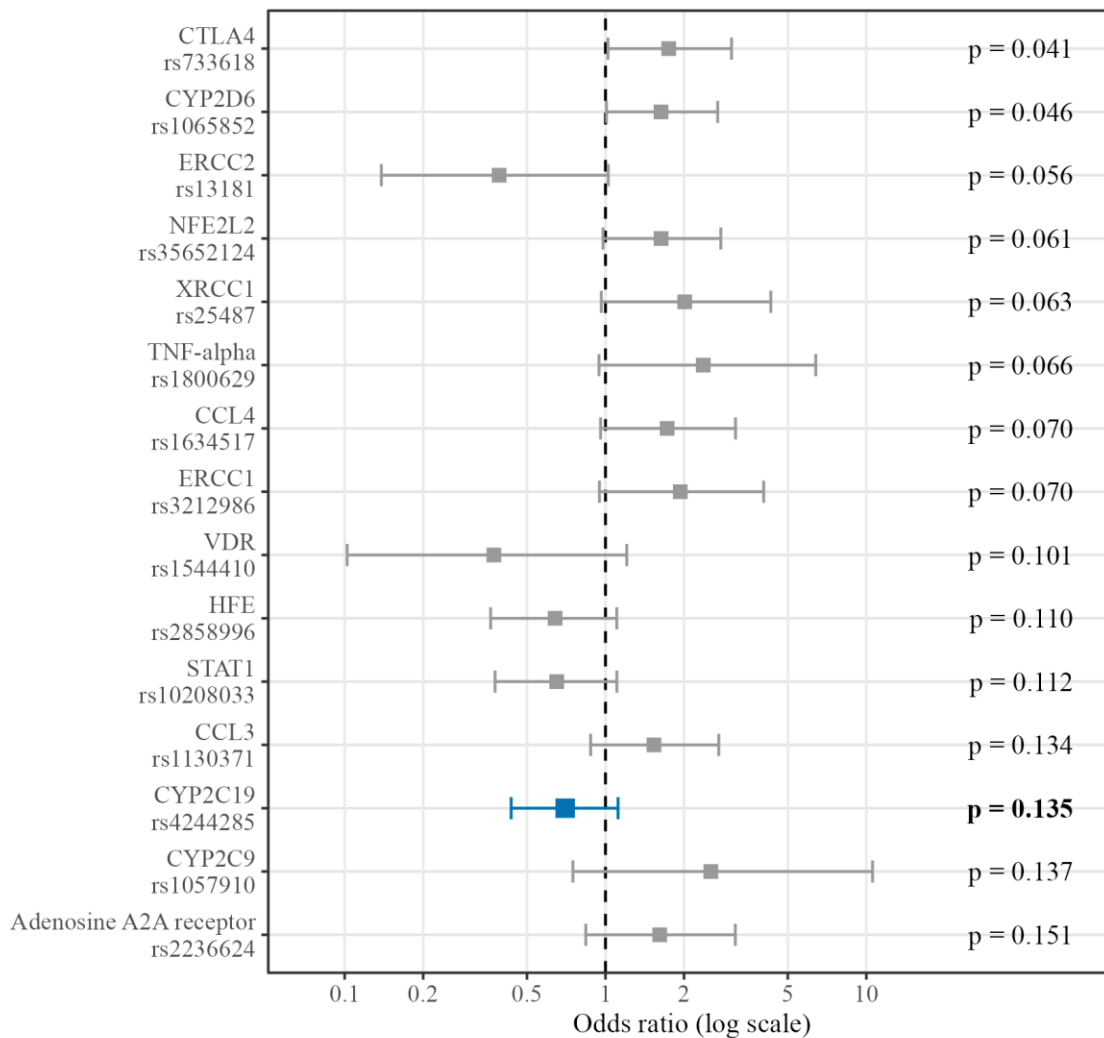


Panels (a)-(b): all imputed autosomal sites, unfiltered by R^2 . Panel (c): candidate oncology-PGx panel variants only.

Supplementary Figure S2. Imputation quality: R^2 distribution by minor-allele-frequency bin, proportion of sites reaching $R^2 \geq 0.30$ per bin, and candidate-panel coverage.

Top 15 variants by raw p-value — Firth regression (H1)

101 pairs / 202 samples



Note: blue point = CYP2C19*2 (rs4244285, p = 0.135; not nominally significant). Smallest raw p-value: CTLA4 rs733618 (p = 0.041).

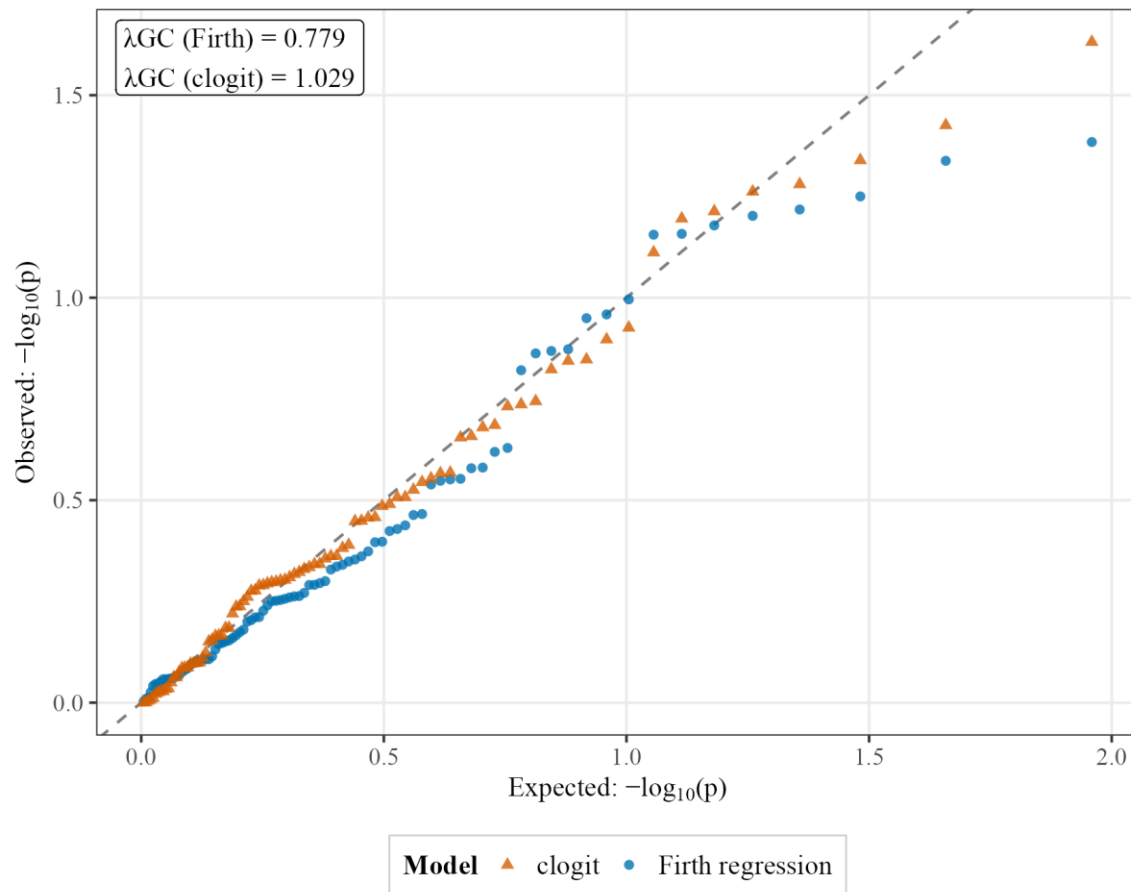
No variant exceeded the Bonferroni threshold $\alpha = 4.348\text{e-}04$ (M = 115).

95% CI from Firth regression. Excluded 0 variant(s) with non-converged OR/CI.

Supplementary Figure S3. H1 — odds ratios and 95% confidence intervals for the fifteen variants with the smallest raw p-values. No variant shown reached significance under Bonferroni or Benjamini-Hochberg correction.

Quantile-quantile (QQ) plot — H1 candidate-panel variants

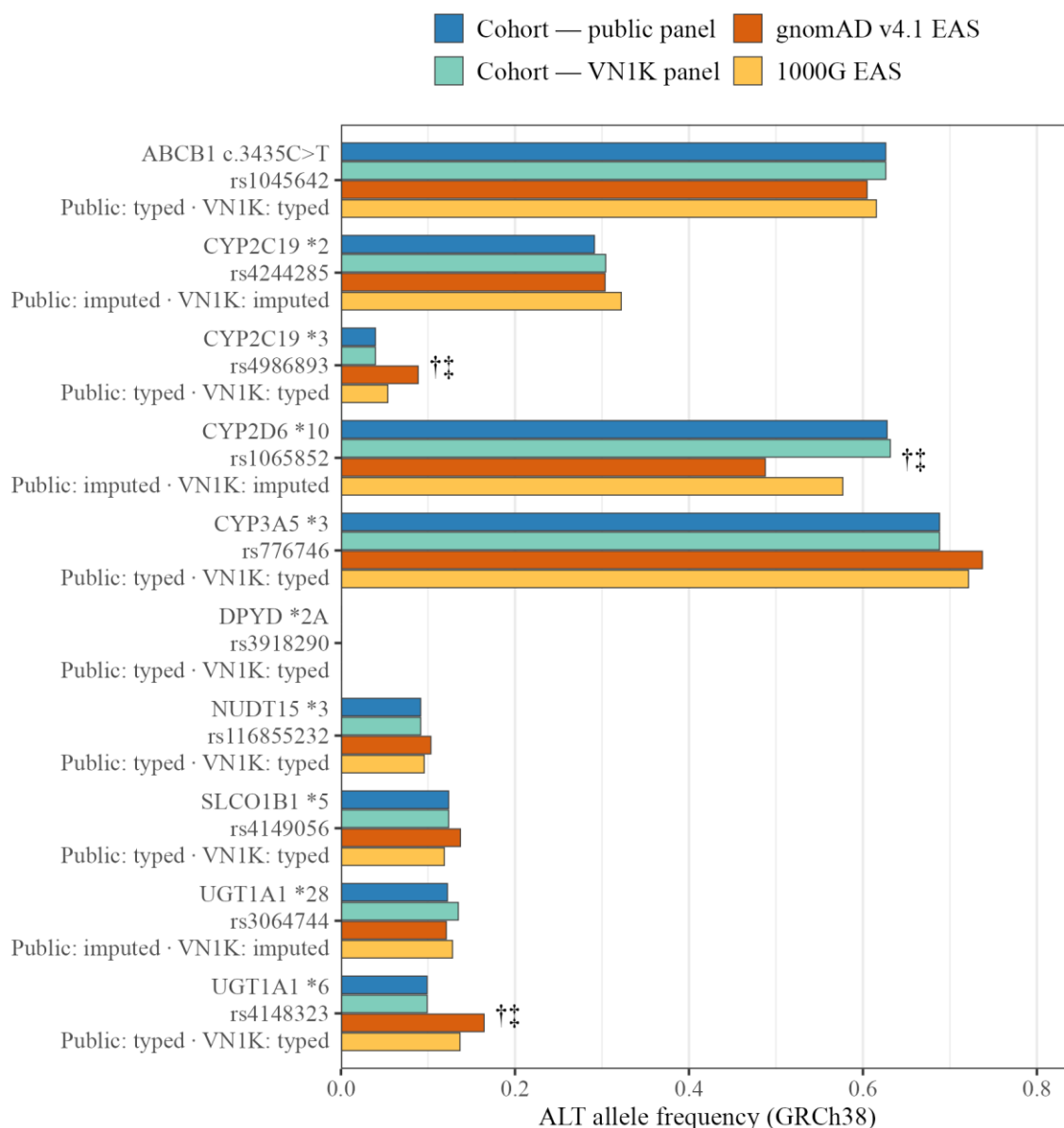
101 pairs / 202 samples



Supplementary Figure S4. H1 — quantile-quantile plot of the candidate-panel p-value distribution.

H2 — CPIC Tier A allele frequencies: two reference-panel arms vs two East Asian references

Public panel: n_chrom = 404 · VN1K panel: 404

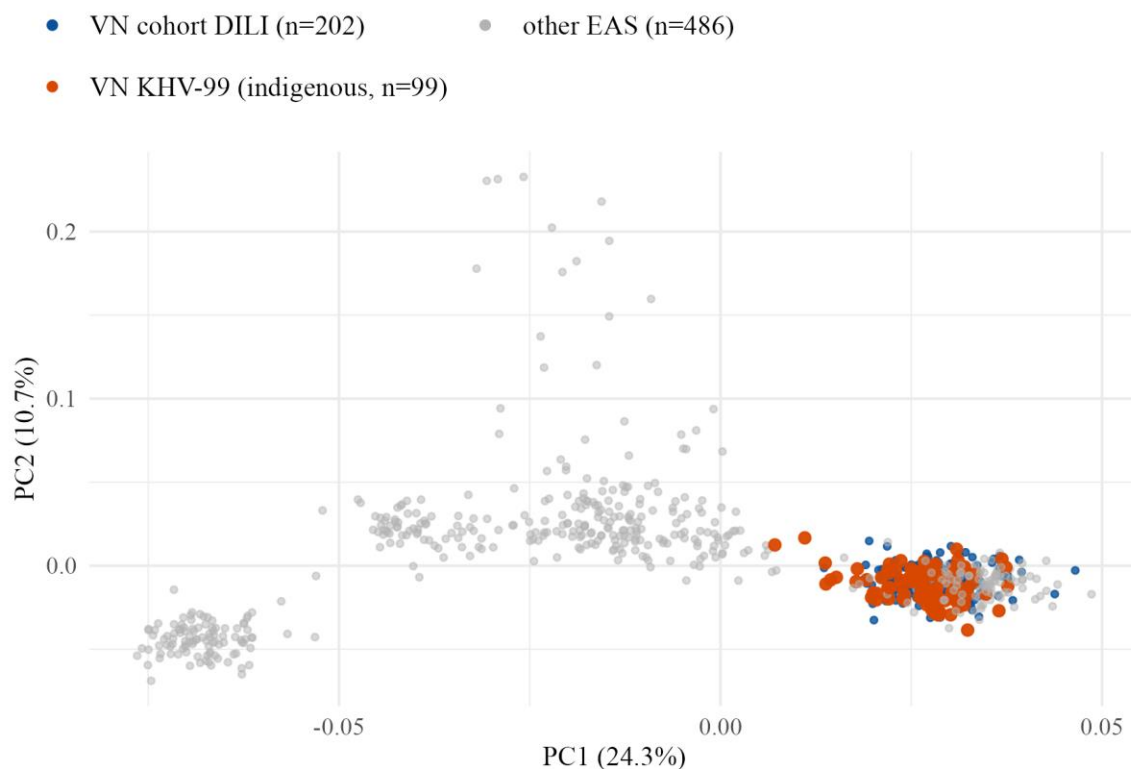


† Bonferroni < 0.05 vs gnomAD v4.1 EAS ‡ BH-FDR < 0.05 vs gnomAD v4.1 EAS.
Last line of the axis label: how the allele was measured on each arm — typed (an array probe exists) or imputed (Public = public panel, VN1K = native panel).

Supplementary Figure S5. H2 — cohort allele frequencies on both reference branches, versus the gnomAD v4.1 and 1000 Genomes East Asian references.

Vietnam-anchored PCA — DILI cohort vs. indigenous Kinh reference (KHV-99)

Joint PCA on a shared array site, palindromic SNPs and long-range LD regions excluded.
 $d(\text{cohort}, \text{KHV})=0.00269 \cdot d(\text{cohort}, \text{other EAS})=0.04912$

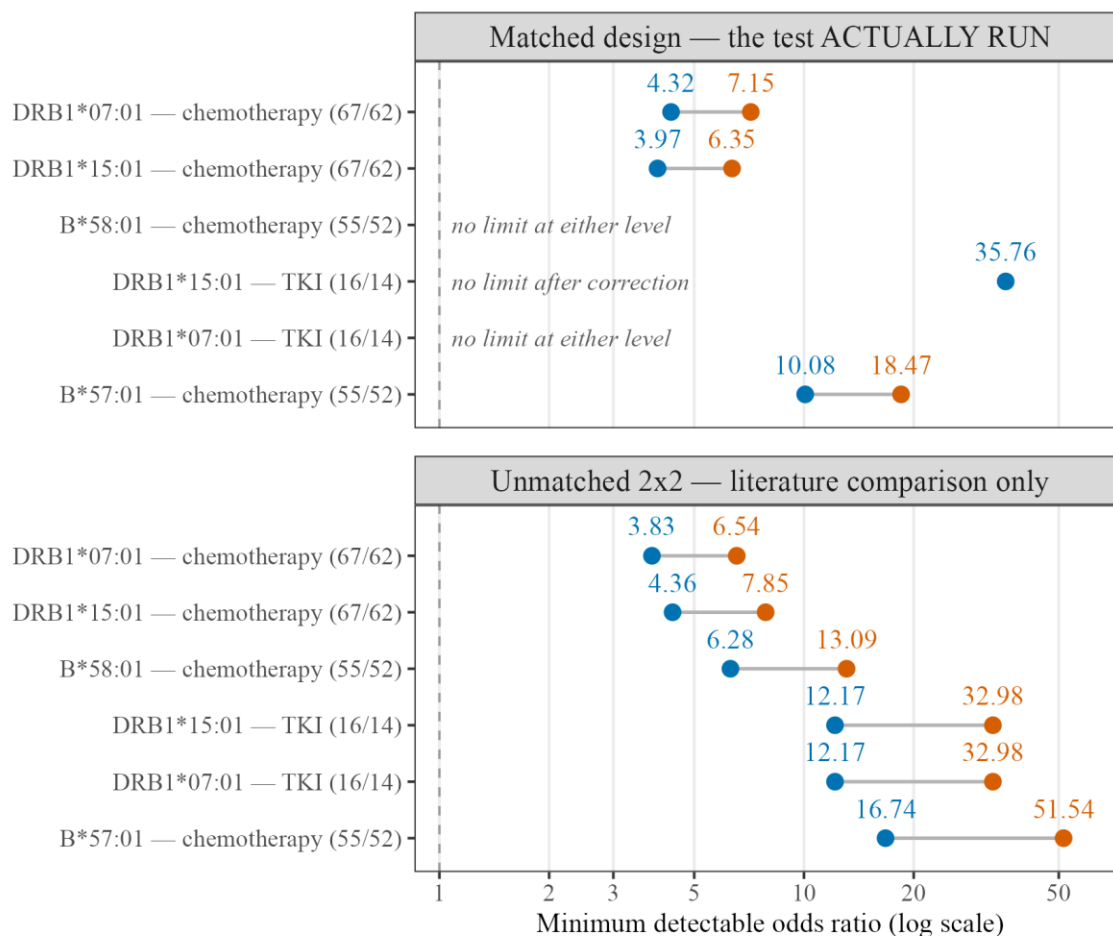


Population-structure description · joint PCA, not a projection.

Supplementary Figure S6. H15 — joint principal component analysis of the cohort and the 1000G-EAS panel, with the Kinh (KHV) population as anchor.

H3 — minimum detectable odds ratio (MDOR) for each estimable HLA-by-drug-class cell: matched design vs 2x2 approximation

80% power; 6 of 24 pre-registered cells estimable. In parentheses: cases / controls in the drug tier.



Significance level

- Nominal (alpha = 0.05)
- Bonferroni-corrected (alpha = 0.05/24)

18 of the 24 pre-registered cells were non-estimable (no carriers in one or both arms, or insufficient sample size) and have NO defined detection limit.

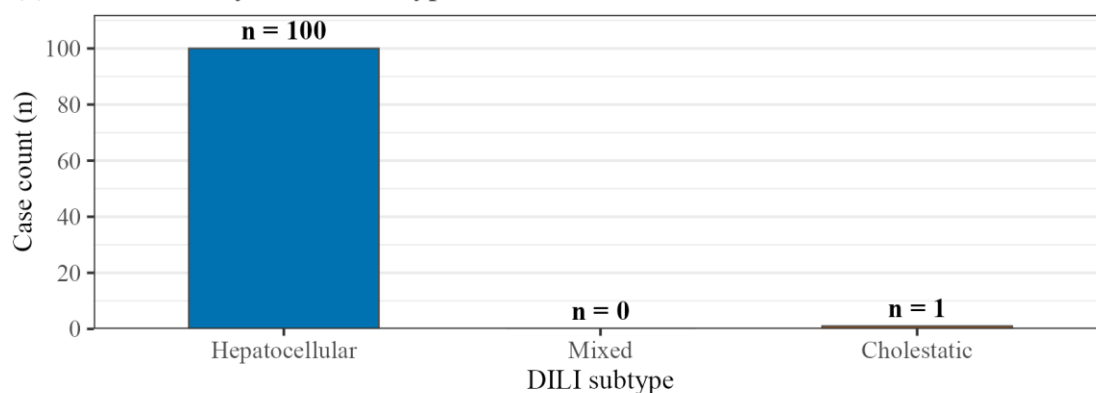
Top panel: the design actually run — exact McNemar route, Dupont parameterisation by the measured within-pair exposure correlation phi. Bottom panel: the 2x2 normal approximation conventional in this literature; it ignores covariates and pair structure and on this design is NOT a lower bound — it exceeds the matched limit 2.8-fold at B*57:01/chemotherapy and falls below it at DRB1*07:01/chemotherapy.

"No limit" means no odds ratio reaches 80% power at that level.

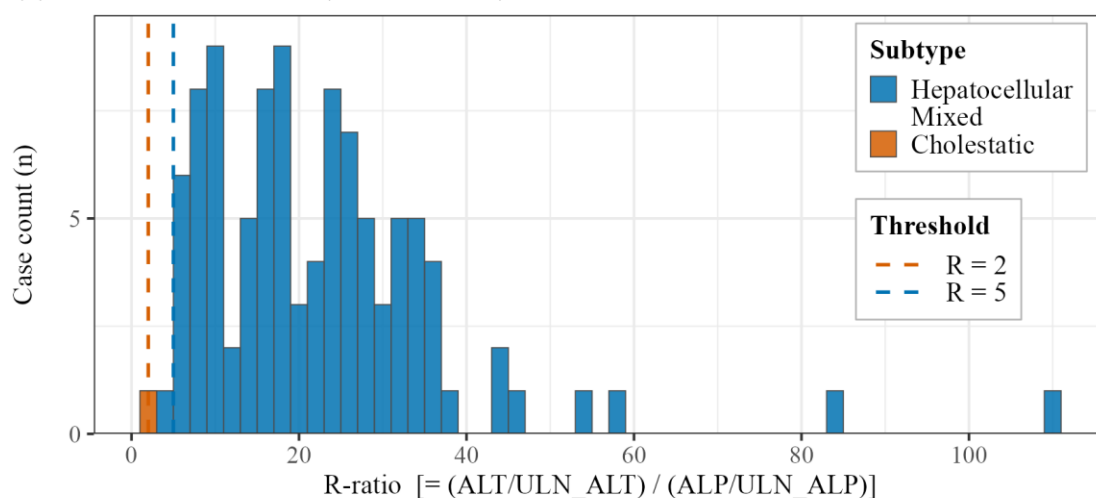
Supplementary Figure S7. H3 — minimum detectable odds ratio at 80% power for each estimable HLA-by-drug-class cell, at nominal and corrected significance levels. Top: the matched design actually run. Bottom: the unmatched 2×2 approximation, shown only for comparability with the published literature; it ignores covariates and pair structure and is neither an upper nor a lower bound on the matched limit.

H4: Distribution of DILI phenotype by R-ratio

(a) Case counts by clinical subtype



(b) R-ratio distribution (n = 101 cases)

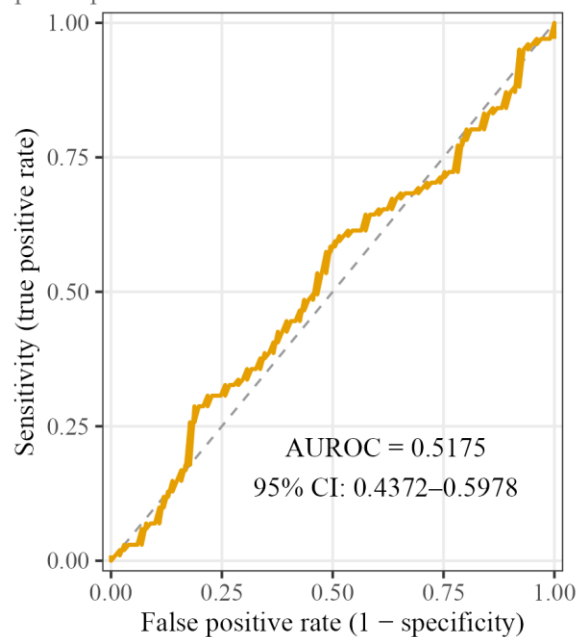


Note: 0 case(s) could not have R-ratio computed (missing ALP) and were excluded from panel (b).
 λ GC: Hepatocellular = 0.6941; Mixed = not estimable (stratum below the minimum size for testing).
 Subtype thresholds: hepatocellular $R \geq 5$; mixed $2 < R < 5$; cholestatic $R \leq 2$; ULN lab_specific;
 R rounded to 1 decimal place(s) BEFORE the thresholds are applied.

Supplementary Figure S8. H4 — distribution of DILI phenotype by R-ratio in 101 cases, under the pre-specified primary convention of per-sample laboratory upper limits.

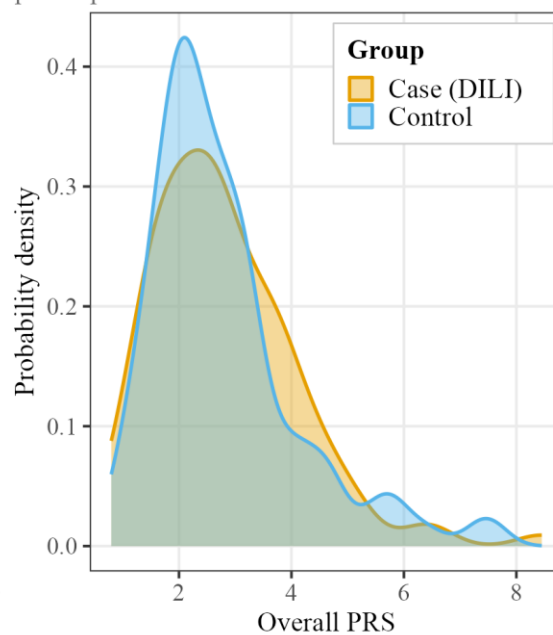
(a) ROC curve — overall PRS

101 cases / 101 controls — 202 samples ·
public panel

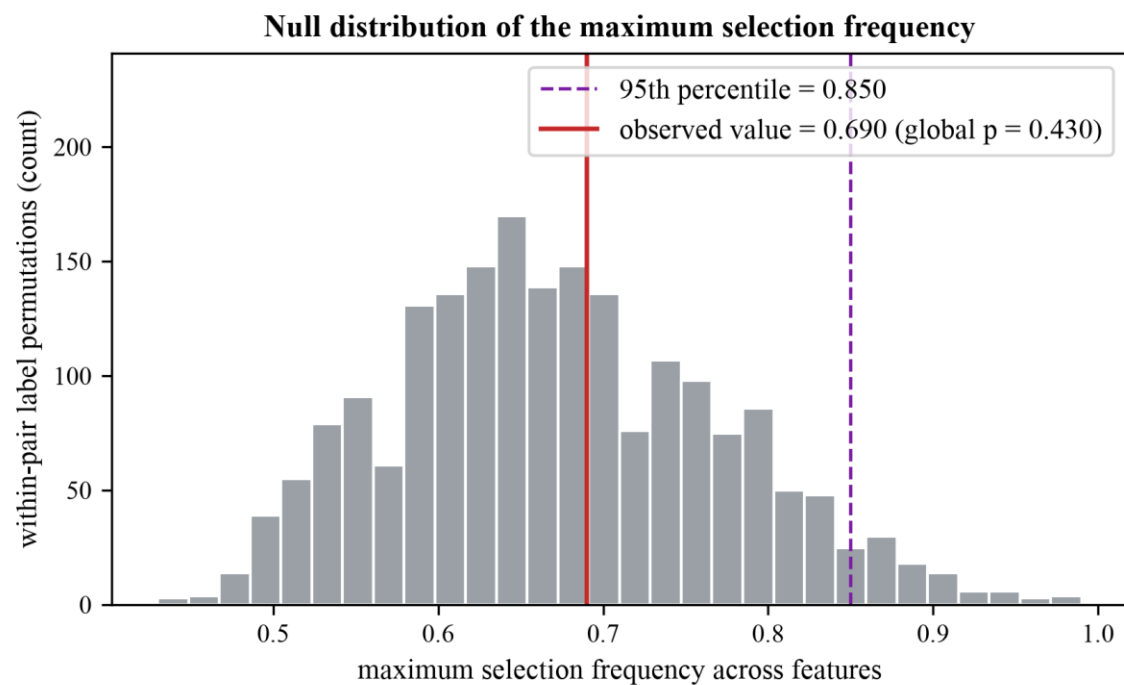


(b) Overall PRS distribution by case/control status

101 cases / 101 controls — 202 samples ·
public panel



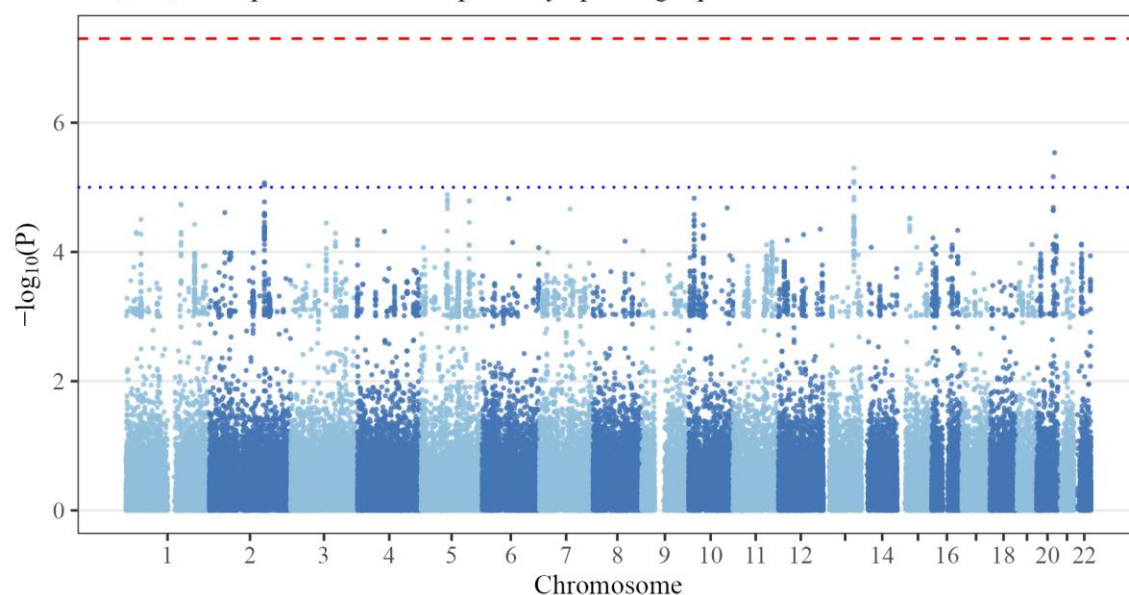
Supplementary Figure S9. H5 — receiver-operating-characteristic curve for the candidate-panel polygenic risk score.



Supplementary Figure S10. H8 — null distribution of the maximum selection frequency under within-pair label permutation, with the observed value marked.

H9 — Genome-wide GWAS scan (Firth) · $\lambda_{GC} = 0.971$ (descriptive)

n=202 · 9,119,954 imputed variants · Exploratory · pending replication



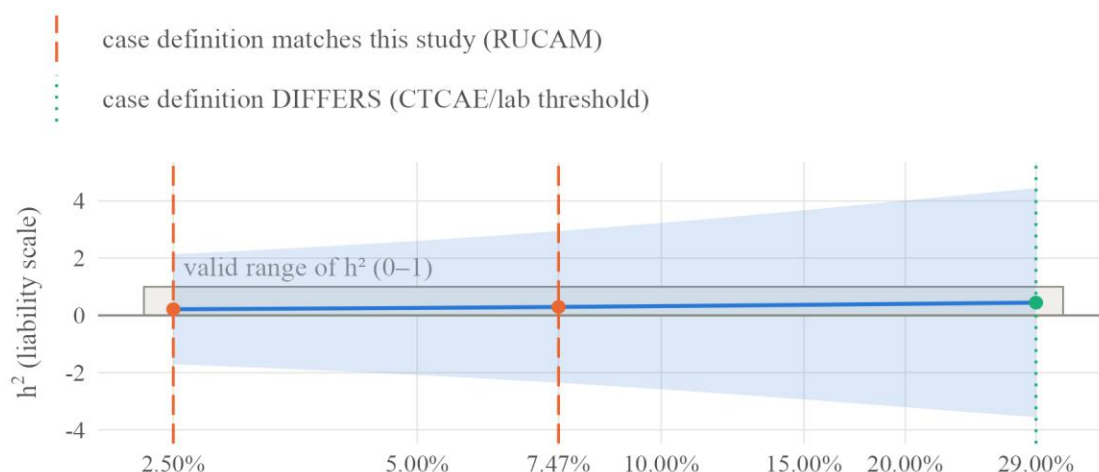
Red: $5e-8$ (genome-wide significance) · Blue: $1e-5$ (suggestive threshold). λ_{GC} is reported, not a stopping gate. n=202.

Supplementary Figure S11. H9 — Manhattan plot of the genome-wide association scan. Of the 9 123 220 variants tested after the $R^2 \geq 0.30$ floor, 3 266 Firth fits did not converge; we plot the remaining 9 119 954 with a valid p-value here, with both significance thresholds marked.

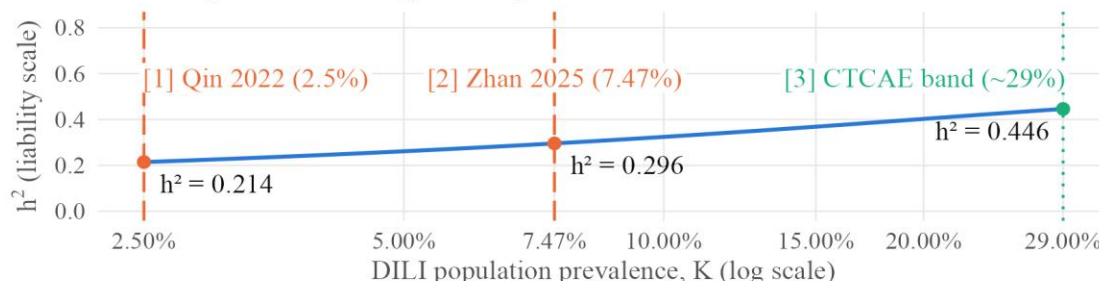
H14 — SNP-based h^2 on the liability scale vs. population prevalence K

Reporting the CURVE, not a single K: no study reports RUCAM-based DILI prevalence for the Vietnamese or Southeast Asian cancer population.

Band: 95% confidence interval. It covers both negative values and values above 1 at EVERY K, so no value in the valid range is excluded. SE/estimate = 4.58, and this ratio is INVARIANT to K: the Lee 2011 conversion is a constant multiplier, so choosing K does not buy precision.



Y-axis zoomed in, CONFIDENCE BAND REMOVED — only to show the shape of $h^2(K)$. Do not read this panel as a finding: see the panel above.



[1] Qin FL, Sang GY, Zou XQ, Cheng DH. Can J Gastroenterol Hepatol 2022;2022:5914593. DOI 10.1155/2022/5914593; PMID 35369115 — Childhood acute lymphoblastic leukaemia, not adult solid tumours; and a per-course rate, not a cumulative per-patient rate.

[2] Zhan H, Deng Y, Huang C, Yang JX, Yang J. Medicine 2025;104(25):e42881. DOI 10.1097/MD.0000000000042881; PMID 40550018 — Closest match on population and case definition (Chinese adults, chemotherapy, RUCAM), but it covers ONE drug class only (docetaxel), whereas this cohort spans chemotherapy + TKI + ICI.

[3] Joel S, et al. Cancer Manag Res 2026;18:589840. DOI 10.2147/CMAR.S589840; PMID 41913766 (29.2%, 63/216) — upper edge of the CTCAE/lab-threshold band — The case definition DIVERGES from RUCAM, so the denominator answers a different question; population is Ugandan, not Asian. Included only to mark the upper edge of the band.

Observed-scale → liability-scale conversion: Lee SH, Wray NR, Goddard ME, Visscher PM. Am J Hum Genet 2011;88(3):294-305, DOI 10.1016/j.ajhg.2011.02.002, eq.(23), with ascertainment correction via the case fraction in the sample (0.500). Cross-checked independently against GCTA --prevalence at all three anchors.

LIMITATION: the matched-pair structure was never entered into the GREML model — 101 matched pairs is being treated as 202 independent observations. GCTA has no option for a matched-pair design, and the GREML/PCGC/LDSC literature has no published method for it (Golan, Lander & Rosset, PNAS 2014, DOI 10.1073/pnas.1419064111, calls this an open problem).

$n = 202$, $V(G)/V_p$ observed scale = $0.308102 \pm \text{SE } 1.41$.

Supplementary Figure S12. H14 — liability-scale heritability as a function of population prevalence K, with 95% confidence band and three literature-anchored reference points.

Supplementary References

9. Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin Pharmacol Ther.* 2011;89(3):464-467. doi:10.1038/clpt.2010.279
14. Tran TTH, Hoang TH, Tran MH, Pham TM, Nguyen NN, Vu GM, et al. VN1K is a pangenome-informed multi-omics and phenomics resource for the Vietnamese population. *Nat Commun.* Jul 22 2026;17(1):6679. doi:10.1038/s41467-026-75375-0
20. Byrska-Bishop M, Evani US, Zhao X, Basile AO, Abel HJ, Regier AA, et al. High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell.* Sep 1 2022;185(18):3426-3440 e19. doi:10.1016/j.cell.2022.08.004
22. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet.* Oct 2016;48(10):1284-1287. doi:10.1038/ng.3656
23. Delaneau O, Zagury JF, Robinson MR, Marchini JL, Dermitzakis ET. Accurate, scalable and integrative haplotype estimation. *Nat Commun.* Nov 28 2019;10(1):5436. doi:10.1038/s41467-019-13225-y
24. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature.* Oct 1 2015;526(7571):68-74. doi:10.1038/nature15393
31. Zheng X, Shen J, Cox C, Wakefield JC, Ehm MG, Nelson MR, et al. HIBAG--HLA genotype imputation with attribute bagging. *Pharmacogenomics J.* Apr 2014;14(2):192-200. doi:10.1038/tpj.2013.18

32. Karnes JH, Shaffer CM, Bastarache L, Gaudieri S, Glazer AM, Steiner HE, et al. Comparison of HLA allelic imputation programs. *PLoS One*. 2017;12(2):e0172444. doi:10.1371/journal.pone.0172444
33. Cook S, Choi W, Lim H, Luo Y, Kim K, Jia X, et al. Accurate imputation of human leukocyte antigens with CookHLA. *Nat Commun*. Feb 24 2021;12(1):1264. doi:10.1038/s41467-021-21541-5
57. Le VS, Tran KT, Bui HTP, Le HTT, Nguyen CD, Do DH, et al. A Vietnamese human genetic variation database. *Hum Mutat*. Oct 2019;40(10):1664-1675. doi:10.1002/humu.23835
58. Bender R, Lange S. Adjusting for multiple testing--when and how? *J Clin Epidemiol*. Apr 2001;54(4):343-9. doi:10.1016/s0895-4356(00)00314-0
59. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*. 2010;33(1):1-22
60. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. Mar 17 2011;12:77. doi:10.1186/1471-2105-12-77
61. Lee S, Wu MC, Lin X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics*. Sep 2012;13(4):762-75. doi:10.1093/biostatistics/kxs014
62. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. Apr 2015;11(4):e1004219. doi:10.1371/journal.pcbi.1004219
63. Wang G, Sarkar A, Carbonetto P, Stephens M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J R Stat Soc Series B Stat Methodol*. Dec 2020;82(5):1273-1300. doi:10.1111/rssb.12388

64. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* Jan 7 2011;88(1):76-82. doi:10.1016/j.ajhg.2010.11.011
65. Qin FL, Sang GY, Zou XQ, Cheng DH. Drug-Induced Liver Injury during Consolidation Therapy in Childhood Acute Lymphoblastic Leukemia as Assessed for Causality Using the Updated RUCAM. *Can J Gastroenterol Hepatol.* 2022;2022:5914593. doi:10.1155/2022/5914593
66. Zhan H, Deng Y, Huang C, Yang JX, Yang J. Real-world data of docetaxel-induced liver injury among hospitalized patients in Wuhan, China: A retrospective analysis. *Medicine (Baltimore).* Jun 20 2025;104(25):e42881. doi:10.1097/MD.00000000000042881
67. Joel S, Bukke SPN, Mamilla Mugaiahgari BK, et al. Incidence and Risk Factors of Chemotherapy-Induced Hepatotoxicity: A Cross-Sectional Study. *Cancer Manag Res.* 2026;18:589840. doi:10.2147/CMAR.S589840
68. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet.* Mar 11 2011;88(3):294-305. doi:10.1016/j.ajhg.2011.02.002
69. Golan D, Lander ES, Rosset S. Measuring missing heritability: inferring the contribution of common variants. *Proc Natl Acad Sci U S A.* Dec 9 2014;111(49):E5272-81. doi:10.1073/pnas.1419064111