

Ancestry-matched GWAS removes the chromosome 6 MHC dominance in consumer-chip psychiatric polygenic risk scoring: a single-subject 23andMe v5 case analysis in an East Asian (Han Chinese) subject

David Liu*

May 2026

Abstract

In a Han Chinese subject's 23andMe v5 chip-typed genotypes cross-referenced against two ancestry-stratified schizophrenia GWAS, the chromosome 6 MHC region behaves qualitatively differently across cohorts. From PGC3 wave 3 EUR (Trubetsky et al. 2022, $n_{cases} = 76,755$), **990 of 1,491 chip-typed genome-wide-significant (GWS) SNPs (66.4%) fall within the chromosome 6 extended MHC region (25–34 Mb)**. From Lam et al. 2019 EAS ($n_{cases} = 22,778$), **0 of 54 chip-typed GWS SNPs – and 0 of all 1,730 GWS SNPs in the EAS sumstats – fall within the MHC**. Lam et al. note this cross-population discrepancy directly: rs13194504 has minor allele frequency $< 1\%$ in EAS vs 9% in EUR, and the C4-BS allele is uncommon in Han Chinese populations. We document the downstream methodological consequence for single-subject chip-PRS: the chromosome 6 MHC tag-SNP dominance that drives consumer-chip psychiatric polygenic risk scores is specific to European-cohort GWAS at current sample sizes, not a structural property of chip-PRS as such.

In the EUR-cohort analysis, the subject's uncalibrated centered chip-PRS (defined as $\sum_i \beta_i (d_i - 1)$ on the GWAS-reported log-odds scale, where $d_i \in \{0, 1, 2\}$ is effect-allele dosage; this is a convenience centering on a heterozygous-everywhere baseline, not a population-deviation metric) drops by approximately two orders of magnitude (and changes sign, from $+60.33$ to -0.67) when MHC SNPs are excluded. In the EAS-cohort analysis, the dominant per-locus signal lives outside the MHC – primarily at chromosome 10q24 (BORCS7/AS3MT/NT5C2 cluster; subject homozygous for the risk-direction allele at 8 of 13 typed tag SNPs), chromosome 12q24, chromosome 2p16 (VRK2/ZFP36L2), and chromosome 3 loci. The per-locus signal landscape is therefore less LD-block-dominated for East Asian subjects than for European subjects when ancestry-matched GWAS are used – a statement about which loci dominate the weighted sum, not about predictive accuracy of either score.

We complement this analysis with a per-variant pharmacogenomic profile and a coverage audit identifying CPIC-actionable East-Asian-relevant variants not typed by 23andMe v5: *ADH1B* *2, *CYP2D6* *10, *NUDT15* *3, *HLA-B**15:02, and *HLA-B**58:01. These omissions limit the chip's utility as a substitute for a CPIC-aligned clinical pharmacogenomic panel in East Asian patients.

1 Introduction

The Psychiatric Genomics Consortium (PGC) and related efforts have produced increasingly powerful genome-wide association study (GWAS) summary statistics for major depressive disorder, schizophrenia (SCZ), bipolar disorder, anxiety, and other phenotypes [1, 2, 3, 4]. A common downstream use case is the construction of a polygenic risk score (PRS) for an individual subject, often computed from consumer direct-to-consumer (DTC) genotyping data

*Independent research. Code and supplementary data: <https://github.com/daliu/daliu.github.io>

such as 23andMe v5 (Build GRCh37/hg19, ~631,000 typed sites) for educational, research, or speculative-clinical purposes.

This practice has well-documented limitations. Naïvely computed PRS from a chip's typed subset of GWS SNPs, without linkage-disequilibrium (LD) clumping or population-matched calibration, is uninterpretable as a clinical risk number. The single-subject use case compounds this with a missing-reference-distribution problem: a centered weighted sum has no meaning without a population reference. Standard PRS pipelines (PRSice-2 [10], LDpred / LDpred2 [8, 9], PRS-CS [7]) address most of these concerns but require imputed dense genotypes and population-reference panels that DTC raw downloads cannot directly support.

A less-commonly-discussed limitation is the *anatomical* bias of chip-typed GWS SNPs toward the chromosome 6 major histocompatibility complex (MHC) region. The MHC (~25–34 Mb on chromosome 6, GRCh37) is one of the strongest reproducible psychiatric-disorder loci, particularly for schizophrenia, where complement-mediated synaptic pruning via increased *C4A* expression has been proposed as a downstream mechanism (Sekar et al. 2016 [5]; subsequent work has reported partial replication and methodological challenges, particularly around *C4* imputation in non-European populations [23]). It is also a region of unusually long-range linkage disequilibrium, dense gene content, and exceptional medical interest, and is therefore intensively typed on consumer chips for HLA imputation. The combination produces a chip-PRS in which a single LD block — one set of correlated tag SNPs — can dominate the weighted sum, an LD-inflation source noted in passing in the broader PRS-methodology literature [24] but rarely characterized empirically for the single-subject DTC use case.

A third under-discussed issue is that of *ancestry-cohort mismatch*. The major psychiatric GWAS releases are predominantly European-cohort. Their effect sizes, allele frequencies, and even the identity of GWS SNPs do not transfer cleanly to East Asian, African, or admixed individuals [6]. A subject of non-European ancestry who computes a PRS against an EUR-cohort GWAS receives a number that is doubly suspect: the ancestry-mismatched effect-size weights and the population-mismatched MAF distribution interact in ways that the literature has not fully characterized for the single-subject case.

The Lam et al. 2019 EAS-cohort schizophrenia GWAS [2] observes the cross-population MHC discrepancy directly: several MHC variants are genome-wide significant in EUR cohorts but not in EAS, attributed to allele-frequency differences (e.g., rs13194504 minor allele frequency < 1% in EAS vs 9% in EUR; the *C4* schizophrenia-associated allele being uncommon in Chinese and Korean samples). This case analysis contributes the practical downstream consequence for single-subject chip-PRS, which to our knowledge has not been demonstrated explicitly: **the chromosome 6 MHC dominance in consumer-chip psychiatric PRS is observed only when European-cohort GWAS is used as the reference; it is essentially absent when an East-Asian-cohort GWAS is used.** For the East Asian subject in this case study, the EAS-cohort schizophrenia analysis yields a substantively different per-locus signal landscape, dominated not by the MHC but by chromosomes 10q24, 12q24, 2p16, and 3q26. The non-MHC portion of the chip-PRS becomes interpretable in a way that the EUR-cohort analysis is not, because the LD inflation source has been replaced.

We document this observation, characterize the per-locus signal landscape, complement it with a per-variant pharmacogenomic profile, and audit the 23andMe v5 chip's coverage of clinically actionable East Asian pharmacogenomic variants. We close with a discussion of the methodological implications for single-subject chip-PRS users.

2 Methods

2.1 Subject and genotype quality control

A single male subject of self-reported and genomically-consistent 100% Han Chinese ancestry provided their 23andMe v5 raw download (build GRCh37/hg19, 631,970 typed sites). Indels (II/DD/DI), no-calls (--/00), and entries with non-{A,C,G,T} alleles were filtered, yielding 594,952 biallelic autosomal+X SNP genotypes (94.1% retention). All analyses were performed at 23andMe-typed sites only; no imputation was applied.

2.2 GWAS reference cohorts

We use two schizophrenia GWAS summary-statistics releases:

- **PGC3 wave 3 European-cohort schizophrenia** (Trubetsky et al. 2022) — 76,755 cases / 243,649 controls, primary European-ancestry cohort with no 23andMe sub-cohort [1].
- **Lam et al. 2019 East Asian schizophrenia** — 22,778 cases / 35,362 controls of East Asian descent, with autosomal and chrX SNPs processed via the RICOPILI pipeline [2].

We additionally maintain a curated table of 19 well-characterized single-variant pharmacogenomic positions covered by the 23andMe v5 chip, drawn from CPIC-aligned guidelines and PharmGKB clinical annotations [13, 14, 15, 16, 17, 18, 19]. Variants and frequencies are referenced against gnomAD East-Asian (EAS) population frequencies where available.

2.3 GWS extraction and unified schema

For each GWAS file, we stream the gz-compressed sumstats once and filter rows to a genome-wide-significance threshold of $p < 5 \times 10^{-8}$. Source columns are normalized to a unified schema (rsid, chr, pos, effect_allele, other_allele, freq_effect, beta, or, se, p, n_eff, info). Files where only a Z-score is reported are handled by treating the Z as a signed effect-size proxy — a fallback we flag explicitly because Z-scores are sample-size-dependent ($Z = \beta/SE$) and not directly comparable to β across studies. For the two SCZ files used in this report (PGC3 wave 3 reports BETA, Lam 2019 reports OR), the Z fallback is never triggered; the headline EUR-cohort and EAS-cohort centered-PRS numbers are computed on β and $\log(OR)$ respectively, which are commensurable on the log-odds scale.

2.4 chr 6 MHC partition

We define the *extended* MHC region as chr 6: 25–34 Mb (GRCh37 coordinates), spanning the classical class-I/II/III region (28.5–33.4 Mb) with a conservative buffer to capture extended-MHC LD. For each disorder, we partition typed GWS SNPs into MHC and non-MHC sets and compute centered PRS for each:

$$centered_PRS = \sum_i \beta_i \cdot (d_i - 1)$$

where $d_i \in \{0, 1, 2\}$ is the effect-allele dosage at SNP i and β_i is the GWAS-reported effect size. A positive value indicates the subject’s average dosage is enriched for effect alleles relative to a heterozygous-everywhere baseline; a negative value indicates depletion.

Caveats on the centered-PRS metric. This is a convenience centering, not a calibrated population-deviation metric. The expected value under random sampling from a population at Hardy-Weinberg equilibrium is $\sum_i \beta_i(2p_i - 1)$, where p_i is the per-SNP effect-allele frequency – not zero. The metric should not be compared to standardized PRS outputs from PRSice-2, LDpred2, or PRS-CS, which return scores calibrated against population reference panels. We use this convenience centering because (a) it is unambiguous to compute from raw chip data without a reference panel, and (b) the headline contrast in this paper is between the MHC and non-MHC partitions of the same metric on the same subject’s chip, which is meaningful regardless of absolute calibration. Cross-cohort magnitude comparisons (e.g., the EUR centered PRS magnitude vs the EAS centered PRS magnitude) should not be over-interpreted: they reflect the different SNP sets and different effect-size distributions of the two underlying GWAS, not a directly comparable risk metric.

2.5 Pharmacogenomic per-variant lookup

Per-variant lookups are performed on a curated table of 19 SNPs spanning CYP2C19, CYP2C9, VKORC1, SLCO1B1, DPYD, CYP2B6, UGT1A1, CYP3A5, IFNL3 (IL28B), ALDH2, CYP1A2, ADORA2A, APOE, COMT, BDNF, DRD2/ANKK1, OPRM1, OXTR, and HTR2A. Each genotype is matched against the variant-specific dbSNP forward-strand allele convention; a strand-flip caveat is applied to CYP3A5 rs776746, where 23andMe v5 has been observed to use reverse-strand encoding inconsistently across releases. Findings are framed against East Asian (CHB/CHS) population frequencies from gnomAD where available, falling back to 1000 Genomes Phase 3 EAS otherwise.

2.6 Code availability

Analysis scripts are open-source on the author’s GitHub:

- [scripts/build_genomics_chip_prs_mhc.py](#) – the centered-PRS-with-MHC-partition computation that backs the EUR and EAS schizophrenia results in §3. Reproduces the headline numbers (Lam 2019 EAS: full +1.17, MHC 0, non-MHC +1.17) directly from the publicly-released sumstats download.
- [scripts/build_pharmacogenomics.py](#) – per-variant pharmacogenomic lookup (Section 3.5).
- [scripts/build_genomics_tophits.py](#) – per-disorder GWS SNP extraction with unified column schema.

The PGC3 wave 3 EUR sumstats (Trubetsky 2022) and the Lam 2019 EAS sumstats are publicly available from PGC and Figshare (URLs in references). The EUR-cohort PRS numbers reported in Section 3.2 (+60.33, –0.67) are reproducible by downloading the EUR sumstats and re-running `build_genomics_chip_prs_mhc.py`; we do not re-distribute the sumstats files. Pharmacogenomic claims and per-locus subject-genotype intersections are reproducible from chip-typed sites only and require no additional downloads.

3 Results

The headline contrast: cross-referencing the same chip-typed genotypes against PGC3 wave 3 EUR-cohort and Lam 2019 EAS-cohort schizophrenia GWAS yields qualitatively different LD landscapes. In EUR, ~66% of typed GWS SNPs cluster in the chr 6 MHC. In EAS, none do. The remaining subsections quantify this contrast and characterize the EAS-cohort per-locus signal that becomes visible once the MHC dominance is removed.

3.1 23andMe v5 coverage of psychiatric GWS SNPs

Across the two schizophrenia GWAS analyzed:

GWAS	GWS SNPs	23andMe v5-typed	Coverage
PGC3 wave 3 EUR (Trubetskoy 2022)	20,457	1,491	7.29%
Lam 2019 EAS	1,730	54	3.12%

The lower absolute number of EAS-cohort GWS SNPs (1,730 vs 20,457) reflects the smaller case sample size (22,778 vs 76,755 cases). The coverage rate (3.12% vs 7.29%) reflects the chip’s known bias toward EUR-imputation tag SNPs, particularly in the MHC region.

3.2 Chromosome 6 MHC dominance in the EUR-cohort analysis

Of the 1,491 typed GWS SNPs from PGC3 wave 3 EUR, **990 (66.4%) fall within the chr 6: 25–34 Mb extended MHC region**. The naïvely computed centered PRS for this subject from the full 1,491-SNP set is +60.33. Stratifying by MHC region:

Partition	Typed SNPs	Centered PRS
All typed GWS	1,491	+60.33
MHC only (chr 6: 25–34 Mb)	990	+60.99
Non-MHC	501	−0.67

The non-MHC partition centered PRS (−0.67 over 501 SNPs) is statistically indistinguishable from the heterozygous-everywhere baseline. The full +60.33 score is therefore essentially a homozygosity readout of the subject’s chr 6 MHC tag SNP dosage, which is in long-range LD across the entire extended MHC region. We extended the partition to three additional well-characterized long-range LD regions in EUR populations — the chromosome 8p23 inversion, the chromosome 17q21.31 H1/H2 inversion, and the lactase region (chr 2: 134–137 Mb) — and found that of these three, only chr 17q21.31 contributes meaningfully to the SCZ score (+0.97 over 16 SNPs, ~1.6% of the magnitude). The chr 8p23 and lactase regions are essentially silent for psychiatric chip-PRS.

The MHC dominance is not unique to schizophrenia. In exploratory analyses across 28 EUR-cohort psychiatric and substance-use GWAS datasets (exploratory data not publicly archived), the same pattern held: MHC contributed $\geq 95\%$ of the centered-PRS magnitude for cross-disorder meta-analysis, MDD, bipolar, and anxiety, sometimes exceeding 100% (i.e., the non-MHC partition contributed in the opposite sign, slightly reducing the magnitude relative to MHC alone). The chr 6 MHC homozygosity was the dominant driver of the chip-derived weighted sum across most EUR-cohort psychiatric disorder PRS in this subject’s exploratory data.

3.3 Chromosome 6 MHC absence in the EAS-cohort analysis

The same partition applied to the Lam 2019 East Asian schizophrenia GWAS yields a strikingly different result. Of the 1,730 GWS SNPs in the EAS-cohort sumstats, **zero fall within the chr 6: 25–34 Mb MHC region**. Of the 54 typed by 23andMe v5, none are in the MHC. The centered PRS for the typed EAS-SCZ SNPs is +1.17, and *all* of this signal is non-MHC.

We considered three potential explanations for the EAS MHC absence:

- *Statistical power*. The Lam 2019 case sample size (22,778) is roughly 30% of PGC3 wave 3’s (76,755). The MHC effect sizes might be similar across populations but only reach $p < 5 \times 10^{-8}$ in the larger study.

- *HLA haplotype divergence.* HLA allele frequencies and haplotype block structures differ substantially between EUR and EAS populations, so the specific tag SNPs reaching GWS in EUR may not be the same SNPs (or even on the same haplotypes) reaching GWS in EAS.
- *Imputation reference panel differences.* The Lam 2019 study used different imputation reference panels (1000G Phase 3 + Asian-specific reference) than PGC3 (HRC). Some MHC tag SNPs may be present in one panel but not the other.

We do not adjudicate between these explanations from the present data. The empirical observation, however, is robust: **for an East Asian subject computing a single-subject chip-PRS against an ancestry-matched GWAS, the chr 6 MHC LD inflation that dominates EUR-cohort chip-PRS is not present.**

3.4 Per-locus EAS-cohort schizophrenia signal landscape

The dominant single-Mb genomic windows in the Lam 2019 EAS-SCZ GWS hit landscape are listed below. *Note:* the GWS-SNP counts in each window largely reflect LD-correlated tag SNPs of a small number of underlying causal variants; they should not be interpreted as the count of independent associations. Without locus-level clumping, this is a per-Mb tag-SNP density rather than an effective-locus measurement.

Genomic window (1 Mb)	GWS SNPs	23andMe-typed (subject genotype)
chr12:123–124 Mb	510	11 typed (5 hom-risk, 1 het, 5 hom-other)
chr10:104–105 Mb	285	13 typed (8 hom-risk, 1 het, 4 hom-other)
chr 2: 58–59 Mb	216	10 typed (0 hom-risk, 9 het, 1 hom-other)
chr 3:161–162 Mb	141	3 typed (0 hom-risk, 3 het)
chr 3: 50–51 Mb	89	3 typed (1 hom-risk, 1 het, 1 hom-other)
chr 3:180–181 Mb	81	0 typed
chr 2:201–202 Mb	70	5 typed (3 hom-risk, 0 het, 2 hom-other)

The chr10q24 cluster (104–105 Mb) corresponds to the BORCS7/AS3MT/NT5C2/CNNM2 region, a well-characterized trans-ancestry replicated psychiatric locus where brain expression and methylation studies have implicated BORCS7, AS3MT, and NT5C2 [11, 12]. The subject is homozygous for the risk-direction allele at 8 of the 13 typed tag SNPs in this region, including high-effect-size hits (rs10786701: $p = 2.17 \times 10^{-14}$, $OR = 0.897$, subject hom for non-T allele; rs7917772: $p = 1.14 \times 10^{-10}$, $OR = 1.091$, subject hom for A risk allele).

The chr 2: 58–59 Mb cluster (VRK2/ZFP36L2 region) presents a different pattern: the subject is heterozygous at 9 of 10 typed tag SNPs, suggesting one risk-haplotype and one non-risk-haplotype inheritance configuration at the locus.

The chr 12q24 cluster (123–124 Mb) is the largest single-Mb GWS hot-spot in the Lam 2019 sumstats but the underlying causal gene is less established than at 10q24. Nearby genes outside the 123–124 Mb window include *NCOR2* (~124.8 Mb) and *SCARB1* (~125.3 Mb); the most credible psychiatric mapping in the broader 12q24 neighborhood involves the *ATXN2/SH2B3/CUX2* cluster well upstream (chr 12: 111–112 Mb), also outside this window. We do not attempt a definitive top-SNP-to-gene mapping for the 123–124 Mb hot-spot here.

3.5 Per-variant pharmacogenomic profile

The subject's CPIC-aligned pharmacogenomic profile from the 19-variant lookup is summarized below. **These are research-grade single-subject genotype lookups, not clinical**

pharmacogenomic test results; clinical action requires a CLIA/CAP-certified panel covering the full set of relevant alleles per gene. “Normal” or “no elevated risk” calls below are scoped to the typed variant only — multi-allele genes (CYP2D6, CYP2C19, DPYD, UGT1A1, CYP2B6) have additional clinically-actionable variants that may not be on this chip.

Gene · variant	Genotype	Functional read
CYP2C19 *2 (rs4244285)	GG (*1/*1)	Extensive metabolizer at *2 locus. <i>Caveat:</i> *17 (increased function), *4, *6, *8 not typed; full-gene phenotype could differ
CYP2C19 *3 (rs4986893)	GG	No *3 allele; combined *2/*3-determined extensive-metabolizer status confirmed
VKORC1 -1639 (rs9923231)	CT	Heterozygous at this locus; EAS T allele freq ~90% (most EAS hom-T). <i>Caveat:</i> warfarin dose decisions require IWPC algorithm with CYP2C9, age, BSA, concomitant medications — single-variant interpretation insufficient
CYP2C9 *2 (rs1799853)	CC	No *2; normal warfarin / phenytoin clearance at this locus
CYP2C9 *3 (rs1057910)	AA	No *3
SLCO1B1 *5 (rs4149056)	TT (*1A/*1A)	No elevated myopathy risk attributable to SLCO1B1 rs4149056. <i>Caveat:</i> ABCG2 rs2231142 not assessed; CPIC 2022 statin guideline incorporates both genes
DPYD *2A (rs3918290)	CC (*1/*1)	Normal DPD activity at the *2A locus only. <i>Caveat:</i> *13 (rs55886062), c.2846A>T (rs67376798), and HapB3 (rs56038477) not typed — additional sources of DPD deficiency relevant for fluoropyrimidine dosing
CYP2B6 *6 (rs3745274)	GG	No *6 at this locus. <i>Caveat:</i> *4, *18, and rs2279343 (the second *6 haplotype variant) not typed
UGT1A1 *6 (rs4148323)	GG	No EAS-specific Gilbert variant from this position. <i>Caveat:</i> UGT1A1 *28 (TA repeat polymorphism) is the dominant Gilbert / irinotecan-toxicity variant globally and is not assessable from chip SNP data
CYP3A5 *3 (rs776746)	CC [†]	Likely *3/*3 nonexpresser (typical EAS, ~75%)
IFNL3 / IL28B (rs12979860)	CC	Favorable HCV interferon response genotype (when applicable; less practically relevant in the era of direct-acting HCV antivirals)
ALDH2 *504Lys (rs671)	AG	Glu/Lys het; mild flush, slower acetaldehyde clearance. *504Lys allele freq ~20–25% in East Asian subpopulations; by HWE, ~30–40% of Han Chinese are AG carriers, ~5–10% AA homozygotes
CYP1A2 *1F (rs762551)	AA (*1F/*1F)	Rapid-inducer phenotype (most evident in induced states; baseline differential is contested in the literature; CYP1A2 currently has no CPIC Level A guideline)
ADORA2A (rs5751876)	CC	Reduced caffeine-induced anxiety
APOE rs429358 + rs7412	TT + CC	ε3/ε3 reference genotype; neither elevated nor reduced AD risk relative to population baseline. APOE is one of many AD risk factors and is not deterministic

[†] Strand-flip ambiguity: 23andMe v5 reports rs776746 inconsistently across versions; clinical panel verification recommended before action.

Pharmacogenomic findings interact: single-variant calls do not capture compound-gene metabolizer phenotypes (e.g., CYP2D6+CYP3A4 for tamoxifen, CYP2C19+CYP3A4 for clopidogrel response).

3.6 East Asian variants the chip does not type

Several CPIC-actionable East Asian pharmacogenomic variants are absent from the 23andMe v5 chip:

- **ADH1B *2** (rs1229984) — 1000 Genomes Phase 3 EAS *2 allele frequency ~30% (HWE carrier rate ~50%; 1000G CHB 29%, CHS 24%, JPT 27%, KHV 35%); some specific Han Chinese subpopulations have been reported with carrier rates approaching 70%. EUR *2 allele frequency ~3–5%. The *2 allele accelerates ethanol → acetaldehyde conversion. In combination with ALDH2 *504Lys (which dominates flush severity by reducing acetaldehyde clearance), ADH1B *2 modulates flush intensity, alcohol metabolism speed, and alcohol-dependence vulnerability [22].
- **CYP2D6 *10** (rs1065852) — *10 allele frequency ~40–50% in East Asians. By Hardy-Weinberg, ~70–85% of East Asians are *10 carriers. Affects codeine activation, atomoxetine clearance, and many tricyclic antidepressants and antipsychotics.
- **NUDT15 *3** (rs116855232) — East Asian-enriched; allele frequency ~10% in East Asian populations vs < 1% in Europeans. CPIC Level A guideline: NUDT15 *3 carriers face elevated risk of severe myelosuppression on standard thiopurine doses (azathioprine, 6-mercaptopurine, thioguanine), and CPIC recommends substantial dose reduction or alternative therapy depending on the genotype. Arguably the highest-impact untyped EAS variant after ADH1B/ALDH2 [21].
- **HLA-B*15:02** — a multi-SNP haplotype, not directly typed; ~5–15% of Han Chinese, Thai, Malay, and Vietnamese carriers. FDA black-box warning. CPIC strongly recommends *against* carbamazepine and oxcarbazepine in HLA-B*15:02-positive carbamazepine/oxcarbazepine naïve patients [19].
- **HLA-B*58:01** — ~5–10% of Han Chinese carriers; allopurinol severe cutaneous adverse reaction (SCAR) risk. CPIC recommends an alternative urate-lowering agent in *58:01-positive patients facing first-time allopurinol exposure [20].

These five variants are sufficient to motivate against using the 23andMe v5 chip as a substitute for a CPIC-aligned clinical pharmacogenomic panel in East Asian patients, particularly when thiopurines, carbamazepine, or allopurinol prescription is contemplated.

4 Discussion

4.1 Why MHC dominance differs across cohorts

The 66% MHC share of EUR-cohort GWS SNPs in the typed subset is consistent with previously published observations on chip-PRS and HLA [1, 5]: the MHC is one of the strongest reproducible psychiatric loci, has dense long-range LD, and is intensively typed for HLA imputation purposes on chips designed for European-ancestry users. The 0% MHC share in the Lam 2019 EAS-cohort GWS SNP set is consistent with the Lam et al. 2019 authors' own reported observation [2]: several MHC variants reach genome-wide significance in EUR but not in EAS due to allele-frequency differences — specifically rs13194504 with EAS minor allele frequency < 1% vs 9% in EUR, and the C4-BS allele being uncommon in Han Chinese and Korean populations. The smaller EAS case sample size (22,778 vs PGC3's 76,755) further reduces statistical power at MHC tag SNPs.

We emphasize that this is a sample-size and allele-frequency threshold phenomenon, not a structural absence of MHC psychiatric signal in East Asians. Larger East-Asian-cohort psychiatric GWAS releases at higher case counts can recover MHC GWS hits at lower-frequency tag

SNPs; the 0/1,730 finding here is specific to the Lam 2019 22,778-case scale and would shift at, e.g., a 50,000+ EAS case meta-analysis. The methodological consequence for chip-PRS users at *currently-published* EAS-cohort GWAS sample sizes is robust: the MHC LD-inflation source that dominates EUR-cohort chip-PRS does not dominate EAS-cohort chip-PRS at the Lam 2019 scale.

4.2 Implications for chip-PRS interpretation

The practical takeaway for chip-PRS users is asymmetric across populations:

- For European-ancestry subjects: the chip-PRS for psychiatric disorders is likely to be MHC-dominated. Reporting an unstratified centered PRS without an MHC partition gives a misleading impression of polygenic accumulation. The MHC partition should be reported separately.
- For East Asian subjects (using ancestry-matched GWAS): the MHC LD inflation source is largely absent. The non-MHC PRS landscape is more representative of distributed polygenic signal. The chr 10q24, chr 12q24, chr 2p16, and chr 3q26 clusters become the dominant per-locus contributors.

This is a counterintuitive practical implication: *the per-locus chip-PRS signal landscape is less LD-block-dominated for East Asian subjects than for European subjects when ancestry-matched GWAS is used* — a statement about which loci dominate the weighted sum, not about the predictive accuracy of either score. We emphasize that the larger ancestry-cohort-mismatch problem (effect-size weight transfer, allele-frequency differences, transferability of GWS-SNP identity across populations) still applies; what we have removed is one specific LD-inflation artifact, not the underlying inferential challenge. We also note that the EAS-vs-EUR LD-dominance contrast is sample-size dependent: at larger EAS-SCZ case counts, MHC tag SNPs would presumably reach GWS and re-introduce some of the LD-block dominance, though likely at a different set of tag SNPs than in EUR cohorts due to the rs13194504 / C4 allele-frequency differences.

4.3 The 23andMe v5 East Asian gap

A complementary observation is that the 23andMe v5 chip, while functional for capturing ALDH2 *504Lys, CYP2C19 *2/*3, the EDAR selection-sweep signal, and several other East Asian-relevant variants, misses a small number of clinically actionable variants for East Asian patients. ADH1B *2 (rs1229984) is the most consequential single variant for understanding East Asian alcohol pharmacology and is not typed. CYP2D6 *10 (rs1065852) affects pharmacology of codeine, atomoxetine, and many psychiatric medications and is not typed. NUDT15 *3 (rs116855232; CPIC Level A) drives thiopurine-induced myelosuppression risk; its allele frequency in East Asians (~10%) is more than ten-fold higher than in Europeans (< 1%) and the variant is not typed by the chip — arguably the most consequential pharmacogenomic gap for East Asian patients facing thiopurine-class drugs (azathioprine, 6-mercaptopurine). The HLA-B*15:02 and HLA-B*58:01 multi-SNP haplotypes are not directly tagged; CPIC strongly recommends against carbamazepine and oxcarbazepine in HLA-B*15:02-positive carbamazepine/oxcarbazepine-naïve patients (Phillips et al. 2018 [19]), and recommends an alternative urate-lowering agent in HLA-B*58:01-positive patients facing first-time allopurinol exposure (Saito et al. 2016 [20]). Each of these gaps reduces the chip's substitutability for a CPIC-aligned clinical pharmacogenomic panel in East Asian patients.

5 Limitations

1. **Single subject.** All findings are anecdotal at the individual level. The MHC dominance observation in EUR-cohort chip-PRS, however, has been replicated across 28 disorder GWAS in this subject's data and is consistent with the structural properties of EUR-cohort GWAS and chip design; we expect it to generalize. The MHC absence in EAS-cohort GWS hits is a property of the GWAS itself, not of the subject; we expect it to generalize across all subjects analyzed against the Lam 2019 GWAS.
2. **No within-locus LD clumping.** We partition by long-range LD regions (MHC, chr 8p23, chr 17q21, lactase) but do not perform standard PLINK clumping at non-LD-region loci. Within CACNA1C, within DRD2, within BORCS7/AS3MT/NT5C2, multiple typed tag SNPs in the same haplotype block contribute correlated information that further inflates the PRS magnitude. A full clumped pipeline would attenuate non-MHC PRS magnitudes by an unknown but probably modest factor.
3. **Uncalibrated.** No reference-population PRS distribution is available for any of the disorders in our panel using exactly the 23andMe-v5-restricted SNP set. Centered PRS magnitudes are raw weighted sums, not percentiles or quantiles.
4. **No imputation.** All analyses use chip-typed sites only. Imputation against 1000 Genomes Phase 3 EAS would substantially expand coverage and partially close the EAS pharmacogenomic gaps (HLA-B haplotypes in particular). We chose to remain at the chip-typed level to characterize the chip's direct utility.
5. **Strand encoding caveats.** 23andMe v5 reports most SNPs on the dbSNP forward strand, but inconsistencies have been observed at some pharmacogenomically-relevant positions (notably CYP3A5 rs776746). Where ambiguity exists, this is flagged.
6. **No covariate adjustment.** Population genetic PRS pipelines adjust for principal components, age, sex, and study-specific covariates. We do none of these.
7. **Pharmacogenomic claims are functional, not predictive.** Per-variant pharmacogenomic findings describe the molecular and population-genetic literature on each variant; they are not a clinical risk prediction. Several CPIC guidelines recommend genotype-guided dosing or alternative-drug selection at specific variant-genotype combinations, but adherence to those guidelines requires a clinical panel rather than a consumer chip.

6 Conclusion

The chr 6 MHC tag-SNP dominance that drives single-subject consumer-chip psychiatric PRS is specific to European-cohort GWAS at currently-published sample sizes. Lam et al. 2019's allele-frequency observations (rs13194504 EAS MAF < 1% vs 9% EUR; C4-BS rare in Han Chinese) provide the locus-level explanation; this paper documents the chip-PRS-level consequence. Practical recommendation for chip-PRS users: report MHC and non-MHC partitions separately, particularly when computing on EUR-cohort GWAS; for East Asian subjects, prefer ancestry-matched GWAS where the LD-inflation source does not dominate at present sample-size scales. Five untyped East-Asian-relevant pharmacogenomic variants (ADH1B *2, CYP2D6 *10, NUDT15 *3, HLA-B*15:02, HLA-B*58:01) limit the chip's use as a CPIC-aligned clinical pharmacogenomic substitute in this population.

Code, data, and contact

- Code: <https://github.com/daliu/daliu.github.io/tree/master/scripts>
- Vault: per-subject pharmacogenomic findings, retained privately
- GWAS summary statistics: PGC3 wave 3 SCZ [1], Lam 2019 EAS-SCZ [2] — both publicly available from the Psychiatric Genomics Consortium and Figshare collections
- Subject genotype: 23andMe v5 raw download, kept private

References

- [1] Trubetskoy V, Pardiñas AF, Qi T, et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* 604, 502–508 (2022). PMID: 35396580.
- [2] Lam M, Chen CY, Li Z, et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nature Genetics* 51, 1670–1678 (2019). PMID: 31740837.
- [3] O’Connell KS, Koromina M, van der Veen T, et al. Genomics yields biological and phenotypic insights into bipolar disorder. *Nature* 639, 968–975 (2025). PMID: 39843750.
- [4] Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. Trans-ancestry genome-wide study of depression identifies 697 associations implicating cell types and pharmacotherapies. *Cell* (2025). PMID: 39814019.
- [5] Sekar A, Bialas AR, de Rivera H, et al. Schizophrenia risk from complex variation of complement component 4. *Nature* 530, 177–183 (2016). PMID: 26814963.
- [6] Martin AR, Kanai M, Kamatani Y, et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature Genetics* 51, 584–591 (2019). PMID: 30926966.
- [7] Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications* 10, 1776 (2019). PMID: 30992449.
- [8] Vilhjálmsson BJ, Yang J, Finucane HK, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores (LDpred). *Am J Hum Genet* 97, 576–592 (2015). PMID: 26430803.
- [9] Privé F, Arbel J, Vilhjálmsson BJ. LDpred2: better, faster, stronger. *Bioinformatics* 36, 5424–5431 (2021). PMID: 33326037.
- [10] Choi SW, O’Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Giga-Science* 8, giz082 (2019). PMID: 31307061.
- [11] Li M, Jaffe AE, Straub RE, et al. A human-specific AS3MT isoform and BORCS7 are molecular risk factors in the 10q24.32 schizophrenia-associated locus. *Nature Medicine* 22, 649–656 (2016). PMID: 27158905.
- [12] Duarte RRR, Troakes C, Nolan M, Srivastava DP, Murray RM, Bray NJ. Genome-wide significant schizophrenia risk variation on chromosome 10q24 is associated with altered cis-regulation of BORCS7, AS3MT, and NT5C2 in the human brain. *Am J Med Genet B* 171B, 806–814 (2016).
- [13] Lee CR, Luzum JA, Sangkuhl K, et al. CPIC guideline for CYP2C19 and clopidogrel. *Clin Pharmacol Ther* (2022). PMID: 35034351.
- [14] Lima JJ, Thomas CD, Barbarino J, et al. CPIC guideline for CYP2C19 and PPI dosing. *Clin Pharmacol Ther* (2021). PMID: 32770672.
- [15] Cooper-DeHoff RM, Niemi M, Ramsey LB, et al. CPIC guideline for SLCO1B1, ABCG2, and CYP2C9 and statin-associated musculoskeletal symptoms. *Clin Pharmacol Ther* (2022). PMID: 35152405.
- [16] Amstutz U, Henricks LM, Offer SM, et al. CPIC guideline for DPYD and fluoropyrimidine dosing. *Clin Pharmacol Ther* (2018). PMID: 29152729.

- [17] Johnson JA, Caudle KE, Gong L, et al. CPIC guideline for pharmacogenetics-guided warfarin dosing. *Clin Pharmacol Ther* (2017). PMID: 28198005.
- [18] Crews KR, Monte AA, Huddart R, et al. CPIC guideline for OPRM1, CYP2B6, and select opioids. *Clin Pharmacol Ther* (2021). PMID: 33387367.
- [19] Phillips EJ, Sukasem C, Whirl-Carrillo M, et al. CPIC guideline for HLA genotype and use of carbamazepine and oxcarbazepine. *Clin Pharmacol Ther* (2018). PMID: 29392710.
- [20] Saito Y, Stamp LK, Caudle KE, et al. CPIC guideline for HLA-B*58:01 and allopurinol. *Clin Pharmacol Ther* (2016). PMID: 26094938.
- [21] Relling MV, Schwab M, Whirl-Carrillo M, et al. CPIC guideline for thiopurine dosing based on TPMT and NUDT15 genotypes: 2018 update. *Clin Pharmacol Ther* (2019). PMID: 30447069.
- [22] Edenberg HJ. The genetics of alcohol metabolism: role of alcohol dehydrogenase and aldehyde dehydrogenase variants. *Alcohol Res Health* 30, 5–13 (2007). PMID: 17718394.
- [23] Kim M, Haney JR, Zhang P, et al. Brain gene co-expression networks link complement signaling with convergent synaptic pathology in schizophrenia. *Nature Neuroscience* 24, 799–809 (2021). PMID: 33958802.
- [24] Wray NR, Lee SH, Mehta D, Vinkhuyzen AAE, Dudbridge F, Middeldorp CM. Research review: polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry* 55, 1068–1087 (2014). PMID: 25132410.