Supplementary Information

Ancestry-enriched and shared polygenic signals associated with high IQ in a Japanese extreme-phenotype cohort

Jun Kawasaki, Shun Nogawa, Kazuki Tainaka

Supplementary Note

Supplementary Table S1 provides the integrated retained-variant annotations and is supplied as a separate CSV file; **Supplementary Table S2** summarizes the sample and variant quality-control flow; and **Supplementary Table S3** reports the primary and sensitivity LD-specification results.

Recruitment context and ascertainment instrument

Participants were recruited through a high-cognitive-ability community coordinated by HIQA that included many members of Mensa. Mensa membership generally requires qualifying performance within the upper 2% of the general population on an approved, supervised intelligence test. This recruitment context provided supporting evidence that the pool was enriched for individuals with previously established high-range cognitive performance. However, individual Mensa membership status was not systematically recorded and was not used to define the study phenotype. Eligibility was determined uniformly and independently by a CAMS score of at least 140. The Cognitive Ability Measurement Scale (CAMS) is a high-range matrix-reasoning assessment developed and administered by the Japan High IQ Certification Support Organization (HIQA). It is designed to assess fluid reasoning in the upper ability range. The available HIQA instrument-construction documentation describes 36 matrix-reasoning items

administered with limited time pressure and scoring based on a two-parameter logistic item-response model. CAMS scores are expressed on an IQ-type scale with a mean of 100 and an SD of 15; study eligibility required a CAMS score of at least 140. Under a normal distribution with this scaling, a score of 140 corresponds to $z \approx 2.67$ (approximately the upper 0.4%). In an available calibration subset with Cattell Culture Fair Intelligence Test (CFIT) Scale 3 scores ($n = 63$), the CAMS–CFIT correlation was $r = 0.45$. This correlation provides limited convergent evidence for nonverbal reasoning but does not establish interchangeability with CFIT or equivalence to a comprehensive clinical IQ battery such as the Wechsler Adult Intelligence Scale.

Quality-control procedures

Sample QC comprised genotype call rate > 95%, reported/genetic sex concordance, relatedness filtering at $PI_{HAT} > 0.1875$, and ancestry assessment. Variant QC comprised call rate > 95%, minor-allele frequency > 1%, and control Hardy-Weinberg equilibrium $P > 1 \times 10^{-6}$. The association analysis was restricted to directly genotyped autosomal variants.

The complete sample and variant quality-control flow is provided in **Supplementary Table S2**.

Population structure among the 91 cases was assessed by principal-component analysis using 78,393 LD-pruned variants. Principal-component calculation used PLINK v1.9 [1].

Effective sample size and qualitative power interpretation

For an unbalanced case-control study, the conventional effective sample size is:

$N_{\text{eff}} = \frac{4N_{\text{case}}N_{\text{control}}}{N_{\text{case}}+N_{\text{control}}}$.

With 91 cases and 41,528 controls, N_{eff} is approximately 363. Selecting individuals from an extreme tail can enrich trait-associated alleles under suitable genetic and sampling models, but the magnitude of any efficiency gain depends on allele frequency, effect size, phenotype measurement, and the relationship between the selected tail and unselected controls. No adjusted effective sample size is used for inference here. The design is suited to detecting unusually strong signals and to aggregate comparisons with external GWAS data, but it does not provide stable heritability, formal genetic-correlation, or pathway-enrichment estimates [2,3].

Cross-population LD analysis

Method

The cross-population direction-of-effect analysis used the Coleman et al. European intelligence GWAS as the direction-of-effect reference [4]. Summary-level variants were harmonized to GRCh37 by chromosome and position, and Japanese and European effect alleles were aligned before signed-*Z* reconstruction. Japanese signed *Z*-statistics were calculated directly as $Z_{\text{JP}} = \frac{\beta_{\text{JP}}}{\text{SE}_{\text{JP}}}$. European signed *Z*-statistics were reconstructed from the Coleman et al. two-sided *P* values and reported *β* directions as $Z_{\text{EUR}} = \text{sign}(\beta_{\text{EUR}}) \times \Phi^{-1}\left(1 - \frac{P_{\text{EUR}}}{2}\right)$. European *β* supplies direction only, and its magnitude is reconstructed from the two-sided *P* value. LD clumping used PLINK v2.0.0-a.7.1 [5] against a 1000 Genomes Phase 3 East Asian reference [6] of *N* = 504 samples (v2 subset, 355,879 variants), ranking candidates by European *P* value with no *P* ceiling

and merging correlated variants at $r^2 \geq 0.10$ within a 500 kb window under the primary specification. Only variants compatible with reference alleles were retained. After clumping, 86,681 LD-clumped lead variants were retained for primary analysis. Before clumping, the final Japanese-Coleman harmonized candidate set contained 307,822 variants meeting signed-*Z* eligibility, LD eligibility, and finite European-*P* criteria.

For each European *P*-value rank *K* in {100, 500, 1,000, 2,000}, we tested the excess of directional concordance over 50% with a one-sided exact binomial test and reported Clopper-Pearson 95% confidence intervals. For the four prespecified *K* thresholds, $P < 0.0125$ (0.05/4) was used as a conservative family-wise benchmark; the exact one-sided *P* values remain the reported inferential results. Signed *Z*-statistics were summarized across LD-clumped lead variants by Spearman rank correlation with a 1,000-replicate deterministic variant bootstrap (master seed 42; per-setting seed 42 + setting index). Pearson correlation was reported as a complementary summary. The weighted directional aggregate statistic was defined as

$$Z = \frac{\sum_{i=1}^{m} s_i \, w_i}{\sqrt{\sum_{i=1}^{m} w_i^2}}$$

where $s_i \in \{+1, -1\}$ indicated per-lead concordance, $w_i = -\log_{10}\left[\max\left(P_{\mathrm{EUR},i}, 10^{-300}\right)\right]$ was the European-association weight for lead variant i, and m was the number of LD-clumped lead variants. This quantity is not a conventional effect-size Stouffer meta-analysis or a formal estimate of trans-ancestry genetic correlation. The signed-*Z* Spearman rank correlation was unweighted across LD-clumped leads, whereas the

European-association-weighted directional *Z* statistic weighted each lead's concordance sign by $-\log_{10}$(European *P*).

Sensitivity analyses used r^2 thresholds of 0.05 and 0.20 with 500 kb, and $r^2 = 0.10$ with 250 kb and 1 Mb. Under the primary and three of the four sensitivity settings, 17 East Asian LD-defined lead loci emerged from the 26 retained $P < 1 \times 10^{-5}$ variants; under $r^2 = 0.20$ with 500 kb the *FAM230I* cluster on chromosome 22 splits and the count becomes 18. rs146572333 was too infrequent in the 1000 Genomes Phase 3 EAS reference (MAC = 7) for reliable tag-SNP inference; this was classified as inconclusive rather than unsupported, and no proxy-support claim or regional association plot was made.

Primary aggregate result

After clumping with the 1000 Genomes Phase 3 East Asian reference ($r^2 = 0.10$, 500 kb), directional concordance exceeded 50% at all tested European-GWAS rank thresholds. Concordance was significant for the top 500 (56.0%, 280/500; one-sided $P = 4.13 \times 10^{-3}$; 95% CI 0.515–0.604), top 1,000 (54.5%, 545/1,000; $P = 2.43 \times 10^{-3}$; 95% CI 0.514–0.576), and top 2,000 (53.2%, 1,064/2,000; $P = 2.25 \times 10^{-3}$; 95% CI 0.510–0.554) variants, while the top-100 estimate was not significant (56.0%, 56/100; $P = 0.136$; 95% CI 0.457–0.659). The top-500, top-1,000, and top-2,000 results also remained below the conservative Bonferroni benchmark of $P < 0.0125$ for the four tested *K* thresholds. The European-association-weighted directional *Z* statistic was positive ($Z = 3.01$; one-sided $P = 1.32 \times 10^{-3}$), and the four sensitivity LD specifications all preserved the positive direction.

Sensitivity results

Numerical results for the primary and sensitivity LD specifications are provided in **Supplementary Table S3**.

Signed-*Z* rank correlation

Across all 86,681 primary LD-clumped lead variants, the signed-*Z* Spearman rank correlation was positive but small (ρ = 0.0058; two-sided P = 0.0899; bootstrap 95% CI, −0.0013 to 0.0120). The Pearson correlation across the same leads was r = 0.0089 (two-sided $P = 8.83 \times 10^{-3}$), reported as a complementary summary only. The primary Spearman result is a positive, small, and non-significant signed-*Z* rank correlation.

These statistics summarize weighted directional sharing and are not formal estimates of trans-ancestry genetic correlation.

Winner's curse and effect-size interpretation

Winner's curse describes upward bias in effect estimates selected for exceeding a significance threshold, especially in a study with limited power [7]. The rs146572333 estimate (β = 1.666; OR = 5.29) is therefore not assumed to be an unbiased estimate of the underlying effect. Firth sensitivity [8] reduces concern about separation but does not correct selection bias. Independent data are required to estimate the effect magnitude reliably. For rs146572333, the Firth sensitivity output was β = 1.66589, SE = 0.292035, and $P = 1.1673 \times 10^{-8}$, identical to the reported primary values at this precision.

Ethical considerations

The protocol, consent materials, and data-management plan were reviewed before enrollment. Genetic and phenotypic records were de-identified for analysis, with linking information held separately under restricted access. Public outputs are limited to privacy-filtered aggregate data. The study concerns a complex phenotype shaped by many genetic and environmental influences and does not support genetic determinism, discrimination, selection, or ranking of human worth.

Polygenic score transferability analysis

Calculation and models

PRSice-2 [9] used Savage et al. [10] as the European intelligence base GWAS. Directly genotyped target variants were LD-clumped at $r^2 < 0.1$ within 250 kb. Scores were standardized to the control mean and standard deviation, and logistic regression included sex and PC1–PC10. Six prespecified thresholds formed the primary analysis. Because the six scores were nested and strongly correlated, nominal covariate-adjusted *P* values were reported for the prespecified grid, with $P < 0.0083$ (0.05/6) used only as a conservative Bonferroni benchmark rather than treating the thresholds as six independent discoveries. The within-sample optimized threshold is a separate secondary analysis.

Fixed thresholds (primary analysis)

The full fixed-threshold results are reported in Table 2 of the main manuscript. All six prespecified fixed thresholds showed nominal covariate-adjusted associations with case status. Because the six scores were nested and strongly correlated, the threshold

results were interpreted as a prespecified pattern rather than as six independent discoveries. Four of the six thresholds met the conservative Bonferroni benchmark of $P < 0.0083$. Incremental Nagelkerke R^2 ranged from 0.0047 to 0.0082, and nominal covariate-adjusted P values ranged from 1.23×10^{-3} to 1.36×10^{-2}.

Optimized threshold (secondary analysis)

The full optimized-threshold model metrics are reported in Table 3 of the main manuscript. At $P = 0.0064$, the score included 6,518 SNPs. OR per SD was 1.51 (95% CI 1.23–1.85; $P = 9.60 \times 10^{-5}$), incremental Nagelkerke R^2 was 0.0119, and the PGS-only, covariates-only, and full-model AUCs were 0.612, 0.758, and 0.793. The likelihood-ratio P value was 9.38×10^{-5}. Because the threshold was selected in the same sample, these optimized metrics are secondary and should not replace the fixed-grid evidence.

Male-only sensitivity analysis

Using the full-sample optimized threshold, the male-only model included 85 cases and 20,300 controls and adjusted for PC1–PC10. The association estimate was $\beta = 0.380$ (SE = 0.109; $P = 5.06 \times 10^{-4}$), corresponding to OR per SD = 1.46 (95% CI 1.18–1.81); AUC was 0.634 for the PGS + PC1–PC10 model.

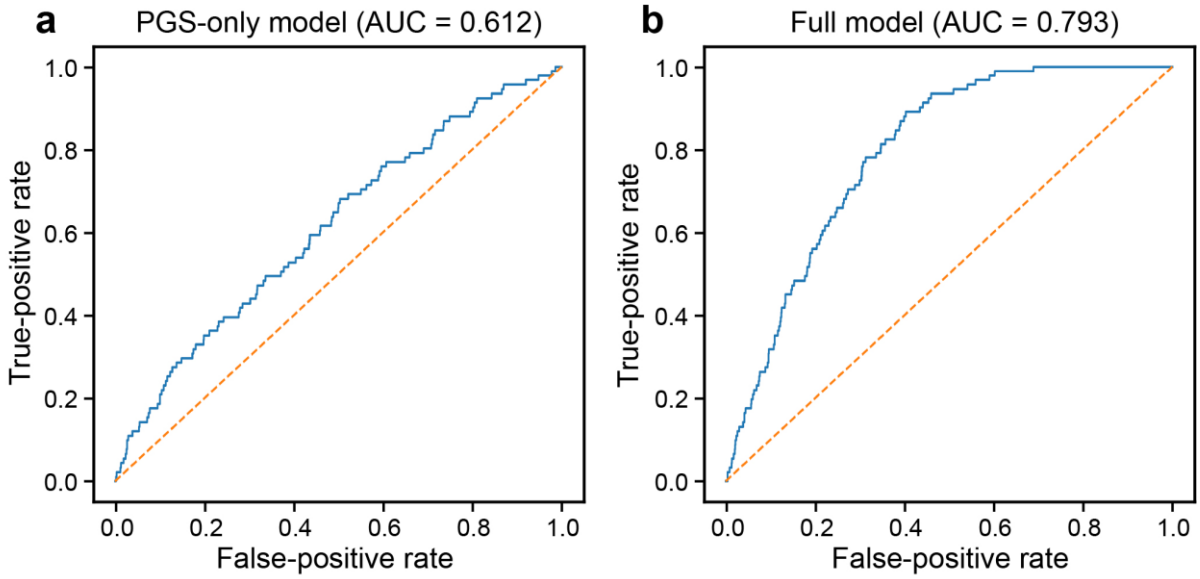
Population-frequency analysis

Top-variant effect-allele frequencies were compared with the ToMMo 38KJPN-SNV/INDEL Allele Frequency Panel [11] and gnomAD v4 genome populations [12]. The European proxy pooled non-Finnish European and Finnish groups using summed allele counts and allele numbers, and frequency ratios were calculated before display

rounding. In gnomAD v4.1.1 genomes, rs146572333 had AC/AN = 84/5,180 in East Asian genomes (AF = 0.0162162) and AC/AN = 33/78,656 in the pooled European proxy (AF = 0.000419548), corresponding to a 38.7-fold ratio. The displayed frequencies 0.0162 and 0.0004 are rounded values. After primary East Asian LD clumping (r^2 = 0.10, 500 kb) the 26 retained $P < 1 \times 10^{-5}$ variants resolved to 17 lead loci, distributed as 5 East Asian enriched, 5 similar frequency, and 7 European enriched (**Figure 2** and **Supplementary Table S1**). Categories were defined as East Asian enriched for EAS/EUR > 2.0, European enriched for EAS/EUR < 0.5, and similar frequency for 0.5 ≤ EAS/EUR ≤ 2.0.

rs376128944 was excluded by the post hoc array-probe QC decision and is not a retained lead. **Supplementary Table S1** contains the submission-facing aggregate variant annotation and reference-frequency fields.

Supplementary figure



Supplementary Fig. S1. Polygenic-score classification curves at the secondary optimized threshold. (**a**) PGS-only model (AUC = 0.612). (**b**) Full model including PGS, sex, and PC1–PC10 (AUC = 0.793). The covariates-only AUC = 0.758 is provided as a reference value; that curve is not shown.

Supplementary tables

Supplementary Table S1. Integrated annotation and population frequencies for retained top variants. This table is supplied as a separate CSV file. All 26 retained $P < 1 \times 10^{-5}$ variants with East Asian LD lead flags and locus IDs from the primary East Asian LD analysis (1000 Genomes Phase 3 EAS reference, $N = 504$; PLINK2 v2.0.0-a.7.1; $r^2 = 0.10$; 500 kb; unphased). Seventeen variants are flagged as LD leads; nine are correlated non-lead members of their respective LD-defined loci. The 17 lead categories comprise 5 East Asian enriched, 5 similar frequency, and 7 European enriched loci. **Figure 2** gene labels reproduce the Gene field in **Supplementary Table**

S1. For rs146572333, gnomAD v4.1.1 genome frequencies were reconstructed from official AC/AN fields; the European proxy pools the non-Finnish European and Finnish groups, and the fold difference was calculated before rounding the displayed allele frequencies.

Supplementary Table S2. Sample and variant quality-control flow.

QC stage	Samples	Variants
Pre-QC	41,650	654,027
Sample call rate > 95%	41,632	-
Sex check passed	41,625	-
Relatedness filter	41,619	-
Variant call rate > 95%	-	612,453
Minor-allele frequency > 1%	-	423,891
Control HWE $P > 1 \times 10^{-6}$	-	398,234
Final directly genotyped autosomal analysis set	-	362,797
Final	**41,619**	**362,797**

Note. Counts after the individual components of the final variant-filtering step were not separately retained in the available analysis record.

Supplementary Table S3. Sensitivity analyses of the European-association-weighted directional *Z* statistic across LD specifications.

Setting	r^2	Window (kb)	*N* leads	Weighted *Z*	One-sided *P*
Primary	0.10	500	86,681	3.007	1.32×10^{-3}
Sensitivity 1	0.05	500	65,916	2.871	2.05×10^{-3}
Sensitivity 2	0.20	500	116,706	3.827	6.48×10^{-5}
Sensitivity 3	0.10	250	88,430	3.237	6.04×10^{-4}
Sensitivity 4	0.10	1000	86,340	3.011	1.30×10^{-3}

Note. LD, linkage disequilibrium; N leads, number of LD-clumped lead variants. P values are from one-sided normal tests.

References

1 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575, doi:10.1086/519795 (2007).

2 Li, Y. *et al.* Extreme sampling design in genetic association mapping of quantitative trait loci using balanced and unbalanced case-control samples. *Sci. Rep.* **9**, 15504, doi:10.1038/s41598-019-51790-w (2019).

3 Zabaneh, D. *et al.* A genome-wide association study for extremely high intelligence. *Mol. Psychiatry* **23**, 1226-1232, doi:10.1038/mp.2017.121 (2018).

4 Coleman, J. R. I. *et al.* Biological annotation of genetic loci associated with intelligence in a meta-analysis of 87,740 individuals. *Mol. Psychiatry* **24**, 182-197, doi:10.1038/s41380-018-0040-6 (2019).

5 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).

6 The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).

7 Xiao, R. & Boehnke, M. Quantifying and correcting for the winner's curse in genetic association studies. *Genet. Epidemiol.* **33**, 453-462, doi:10.1002/gepi.20398 (2009).

8 Firth, D. Bias reduction of maximum likelihood estimates. *Biometrika* **80**, 27-38, doi:10.1093/biomet/80.1.27 (1993).

9 Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience* **8**, giz082, doi:10.1093/gigascience/giz082 (2019).

10 Savage, J. E. *et al.* Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912-919, doi:10.1038/s41588-018-0152-6 (2018).

11 Tohoku Medical Megabank Organization. ToMMo 38KJPN-SNV/INDEL Allele Frequency Panel. *jMorp* https://jmorp.megabank.tohoku.ac.jp/downloads/tommo-38kjpn-20220929-af_snvindelall (2022).

12 Chen, S. *et al.* A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625**, 92-100, doi:10.1038/s41586-023-06045-0 (2024).