AN UNCERTAINTY-AWARE OPTIMAL TRANSPORT FRAMEWORK FOR CROSS-POPULATION POLYGENIC RISK PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Polygenic risk scores (PRS) often lose predictive accuracy across ancestry groups because differences in allele frequencies, linkage disequilibrium (LD), and genetic architecture create distribution shifts between source and target populations. Existing cross-population PRS methods improve portability through Bayesian shrinkage and multi-ancestry modeling, but generally do not explicitly model structured correspondence between ancestry-specific genomic contexts. To address this limitation, we propose OT-PRS, an uncertainty-aware optimal transport framework for cross-population polygenic risk prediction. OT-PRS uses block-wise fused Gromov–Wasserstein transport to align source and target variants by jointly modeling SNP-level similarity and ancestry-specific LD geometry. The transported source effects define a target-aligned prior that is refined using target-population summary statistics, while transport entropy quantifies transfer uncertainty and enables uncertainty-aware score construction and individual-level reliability assessment. Experiments across multiple diseases and ancestry groups in UK Biobank demonstrate competitive performance against established PRS methods, and external validation in the All of Us Research Program further supports the portability of OT-PRS across cohorts. Our results highlight optimal transport as a flexible framework for improving the portability and reliability of polygenic risk prediction.

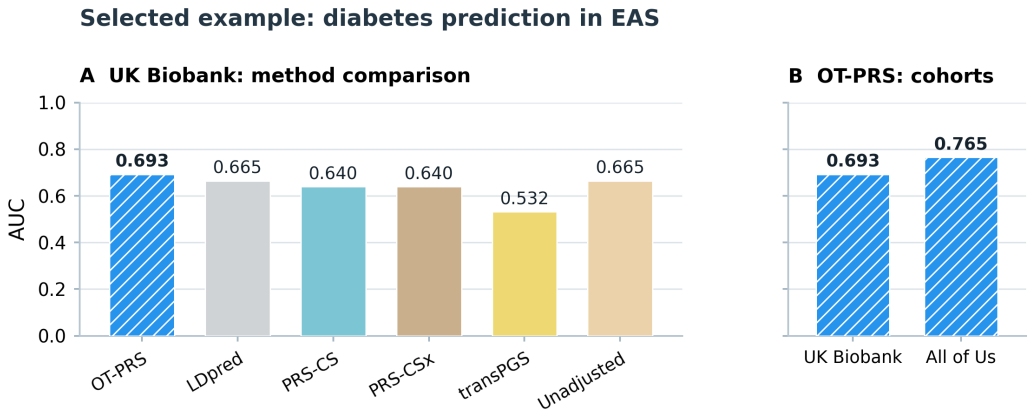


Figure 1: OT-PRS performance preview: diabetes prediction in East Asian (EAS) populations. (A) UK Biobank method comparison using five-fold mean AUC. (B) OT-PRS performance in UK Biobank and external evaluation in All of Us using fixed SNP effects. This selected example is complemented by the complete nine-task comparison in the Results section.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

1 INTRODUCTION

Genome-wide association studies (GWAS) have identified thousands of loci associated with complex human traits and diseases (Whalen et al., 2022; Khera et al., 2018). Because most complex traits are highly polygenic, however, individual variants typically have small marginal effects, making it difficult to assess genetic predisposition from single markers alone (Dudbridge, 2013). Polygenic risk scores (PRS) address this limitation by aggregating effects across many variants into a single measure of genetic liability (Choi et al., 2020; Torkamani et al., 2018; Lewis & Vassos, 2020). PRS has shown promise for risk stratification, early detection, and precision medicine (Lambert et al., 2019).

PRS often lose accuracy across ancestries because European-dominated GWAS do not fully capture differences in allele frequencies, linkage disequilibrium (LD), and genetic architecture (Duncan et al., 2019; Peterson et al., 2019), potentially widening disparities in genomic prediction (Martin et al., 2019). Existing methods address this problem through LD-aware shrinkage and multi-population information sharing (Vilhjálmsson et al., 2015; Ge et al., 2019; Ruan et al., 2022; Weissbrod et al., 2022). We investigate a complementary approach: explicitly aligning ancestry-specific genomic contexts to guide effect transfer. Matching variant identities does not guarantee equivalent tagging information, because ancestry-specific LD can change which SNPs capture an underlying association signal (Figure 2A). Moreover, multiple source variants may provide plausible contributions to a target SNP, creating ambiguity in the correspondence (Figure 2B). These challenges motivate probabilistic alignment within LD blocks and the use of transport entropy to inform SNP weighting and individual-level prediction.

Optimal transport provides a framework for learning soft correspondences between distributions and has been used to align source and target representations in domain adaptation (Peyré & Cuturi, 2019; Courty et al., 2017). Gromov–Wasserstein transport extends this idea to relational structure (Peyré et al., 2016), while fused Gromov–Wasserstein (FGW) transport jointly incorporates feature and structural similarity. We build on these ideas to introduce **OT-PRS**, an uncertainty-aware framework for cross-population polygenic risk prediction. Each LD block is represented as a structured distribution whose nodes are SNPs, whose features capture variant attributes, and whose geometry reflects ancestry-specific LD. An entropically regularized FGW objective learns a coupling between source and target SNPs using allele-frequency differences and LD-derived structural dissimilarities. The coupling defines a statistical correspondence for transferring association information; it does not assume that aligned variants have identical causal effects.

OT-PRS connects this alignment to target effect estimation by constructing a *transport-derived prior*. Source effect estimates are redistributed through the learned coupling to obtain a prior center for target effects. An LD-aware maximum a posteriori (MAP) update then combines this prior with target GWAS summary statistics. This separates the construction of cross-population correspondence from the incorporation of target association evidence. To account for ambiguity in the alignment, OT-PRS uses the normalized entropy of incoming transport weights to characterize SNP-level transfer uncertainty. These measures inform SNP weighting and are aggregated into individual-level scores that describe how strongly a PRS depends on ambiguous transfers. The resulting framework integrates structured alignment, target-informed estimation, and alignment-aware prediction. As a selected example, Figure 1 summarizes diabetes prediction in EAS participants, comparing OT-PRS with baseline methods in UK Biobank and showing its external evaluation in All of Us.

Our methodological contributions are as follows:

- **Structured genomic alignment for PRS transfer.** We formulate cross-population transfer within LD blocks as probabilistic alignment using both SNP-level features and ancestry-specific LD geometry, enabling source information to be shared beyond identical-SNP correspondence.
- **Transport-derived priors for target effect estimation.** We convert the learned coupling into a prior over target effects and integrate it with target summary statistics through LD-aware MAP estimation, linking genomic alignment to a regularized predictive estimator.
- **Alignment-aware weighting and individual-level uncertainty scores.** We use transport entropy to characterize ambiguity in variant correspondence, weight SNP contributions,

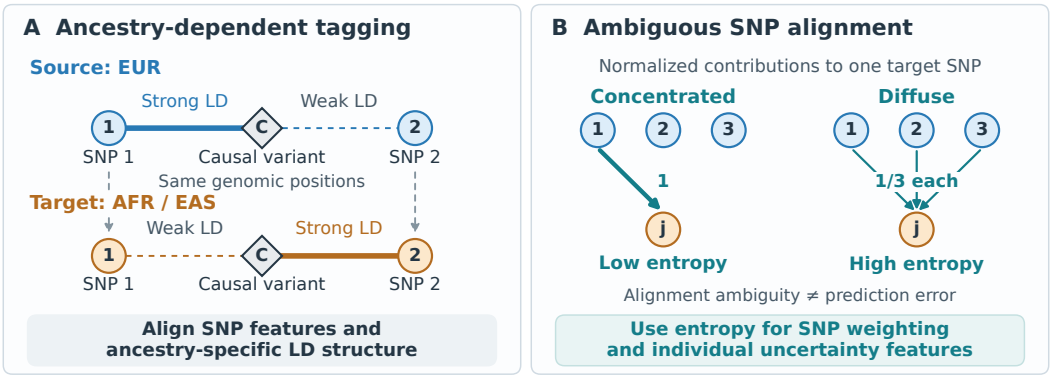


Figure 2: Challenges motivating uncertainty-aware PRS transfer. (A) Ancestry-specific LD changes SNP tagging relationships. (B) Concentrated and diffuse source contributions yield different transport entropies.

and summarize each individual's dependence on uncertain transfers, making alignment ambiguity available for downstream prediction and interpretation.

We evaluate OT-PRS across five diseases in African- and East Asian-ancestry target populations using European-ancestry source data, with independent validation in All of Us, controlled simulations, and ablations that assess the contributions of alignment, target evidence, and uncertainty modeling.

2 Related Work

Cross-population polygenic risk prediction. LD-aware methods, including LDpred, PRS-CS, and SBayesR, improve PRS estimation by regularizing SNP effects using LD reference panels (Vilhjálmsson et al., 2015; Ge et al., 2019; Lloyd-Jones et al., 2019). Cross-population approaches further address ancestry-related heterogeneity through multiethnic score combination, coupled Bayesian priors, fine-mapping, and transfer learning, as exemplified by PT-Multi, PRS-CSx, and transPGS (Márquez-Luna et al., 2017; Ruan et al., 2022; Weissbrod et al., 2022; Zhu et al., 2025). OT-PRS complements these approaches by learning cross-variant correspondence within LD blocks to construct a prior for target effects.

Optimal transport and structured alignment. Optimal transport learns couplings between distributions using feature-level discrepancies, while Gromov–Wasserstein transport aligns their internal relational structures (Peyré & Cuturi, 2019; Peyré et al., 2016). Fused Gromov–Wasserstein transport combines feature and structural similarity (Vayer et al., 2020). OT-PRS applies this formulation to ancestry-specific genomic contexts, jointly aligning SNP attributes and LD geometry.

Uncertainty quantification in PRS. Existing approaches quantify uncertainty in estimated genetic scores through posterior effect distributions and their propagation to individual-level PRS (Ding et al., 2022; Vsevolozhskaya & Zaykin, 2020; Timmins et al., 2024). Other methods address calibration and phenotype-level uncertainty through prediction intervals and conformal inference (Hou et al., 2024; Huang et al., 2025). OT-PRS focuses on a complementary source of uncertainty: ambiguity in cross-population variant alignment, summarized through transport entropy and incorporated into SNP weighting and individual-level modeling.

3 Methodology

OT-PRS combines block-wise genomic alignment, a transport-derived source prior, LD-aware refinement using target GWAS summary statistics, and uncertainty-aware prediction (Figure 3). Appendix A provides supporting formulas and implementation details.

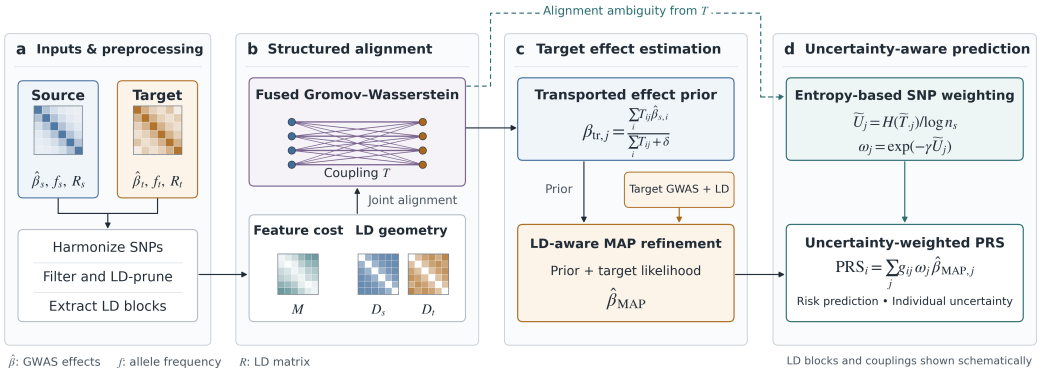


Figure 3: OT-PRS overview. (a) Data harmonization and LD-block preparation. (b) FGW alignment using allele frequencies and LD. (c) Transport-prior construction and target-informed refinement. (d) Entropy-based SNP weighting and individual uncertainty modeling.

3.1 Problem setup and genomic representation

We use source and target GWAS effects and standard errors, ancestry-specific allele frequencies, and LD reference matrices. Variants are harmonized and grouped into approximately independent LDetect blocks (Berisa & Pickrell, 2016). Block k contains n_k source and m_k target SNPs. The cross-population cost $M^{(k)}$ measures squared effect-allele frequency differences; within-population dissimilarities $A^{(k)}$ and $B^{(k)}$ are obtained from LD correlations as $1 - |r|^2$. Allele coding and variant order are aligned across all inputs. Filtering, pruning, and exact input definitions appear in Appendices A.1–A.2.

3.2 Block-wise structured alignment

For each block, a nonnegative coupling $T^{(k)}$ assigns mass from source to target SNPs. Its row and column marginals, $p^{(k)}$ and $q^{(k)}$, are based on normalized inverse-variance weights. We estimate the coupling through entropy-regularized fused Gromov–Wasserstein transport:

$$T^{(k),*} \in \arg\min_{T \in \Pi(p^{(k)}, q^{(k)})} (1-\alpha)\langle T, M^{(k)} \rangle + \alpha \mathcal{L}_{\mathrm{GW}}(T; A^{(k)}, B^{(k)}) - \varepsilon H(T). \tag{1}$$

Here, $\Pi(p, q)$ contains couplings with marginals p and q, $\mathcal{L}_{\mathrm{GW}}$ measures squared discrepancies between LD dissimilarities, and $H(T) = -\sum_{i,j} T_{ij} \log T_{ij}$. The parameters $\alpha \in [0, 1]$ and $\varepsilon > 0$ control feature–structure balance and entropic regularization. Unlike direct SNP matching, the coupling can redistribute source information across target variants within a block. Appendix A.2 gives the marginal and structural-loss formulas.

3.3 Transport-derived prior and target refinement

The incoming transport mass defines a target-aligned prior mean from source effects:

$$\beta_{\mathrm{tr},j}^{(k)} = \frac{\sum_{i=1}^{n_k} T_{ij}^{(k),*} \widehat{\beta}_{\mathrm{src},i}^{(k)}}{\sum_{i=1}^{n_k} T_{ij}^{(k),*} + \delta}, \qquad j = 1, \ldots, m_k, \tag{2}$$

where $\delta > 0$ ensures numerical stability. We refine these effects using target marginal estimates $\widehat{\beta}_{\mathrm{marg}}^{(k)}$ and target LD $R_{\mathrm{tgt}}^{(k)}$. With $\|v\|_Q^2 = v^\top Q v$, the regularized MAP estimator is

$$\widehat{\beta}_{\mathrm{MAP}}^{(k)} = \arg\min_{\beta} \frac{1}{2}\left\|\widehat{\beta}_{\mathrm{marg}}^{(k)} - R_{\mathrm{tgt}}^{(k)}\beta\right\|_{W^{(k)}}^2 + \frac{1}{2}\left\|\beta - \beta_{\mathrm{tr}}^{(k)}\right\|_{\Lambda^{(k)}}^2 + \frac{\lambda_{\mathrm{ridge}}}{2}\|\beta\|_2^2. \tag{3}$$

The diagonal matrix $W^{(k)}$ contains inverse squared target standard errors, $\Lambda^{(k)}$ is the positive-definite prior precision, and $\lambda_{\text{ridge}} \geq 0$ provides additional regularization. The first term fits target marginal associations through the LD-based mean model, the second shrinks target effects toward the transported source prior, and the third stabilizes estimation. This working Gaussian model incorporates LD through its mean and approximates marginal-estimate covariance by a diagonal matrix. The probabilistic formulation and linear-system solution appear in Appendix A.3.

3.4 Alignment uncertainty and risk prediction

We quantify alignment ambiguity using normalized entropy of each target SNP's incoming contributions:

$$\widetilde{T}_{ij}^{(k)} = \frac{T_{ij}^{(k),*}}{\sum_{i'=1}^{n_k} T_{i'j}^{(k),*}}, \qquad \widetilde{U}_j^{(k)} = -\frac{\sum_{i=1}^{n_k} \widetilde{T}_{ij}^{(k)} \log \widetilde{T}_{ij}^{(k)}}{\log n_k}. \tag{4}$$

For $n_k > 1$, $\widetilde{U}_j^{(k)} \in [0,1]$; singleton blocks receive zero uncertainty. Larger values indicate more diffuse source contributions. After concatenating blocks, we attenuate effects using $w_j = \exp(-\gamma_{\text{UQ}} \widetilde{U}_j)$ and construct

$$\widetilde{\beta}_{\text{tgt},j} = w_j \widehat{\beta}_{\text{MAP},j}, \qquad \text{PRS}_i = \sum_{j=1}^{m} g_{ij} \widetilde{\beta}_{\text{tgt},j}, \tag{5}$$

where g_{ij} is effect-allele dosage, $m = \sum_k m_k$, and $\gamma_{\text{UQ}} \geq 0$ controls attenuation. We also aggregate SNP entropy into an individual uncertainty score, weighted by squared genotype deviations and unweighted MAP effects. For binary outcomes, we use

$$\text{logit}\, \Pr(Y_i = 1) = \theta_0 + \theta_1 \text{PRS}_{i,Z} + \theta_2 UQ_{i,Z} + \theta_3 \text{PRS}_{i,Z} UQ_{i,Z} + \mathbf{c}_i^\top \boldsymbol{\theta}_c, \tag{6}$$

where Z denotes standardization and $\mathbf{c}_i$ contains adjustment covariates. The interaction allows the PRS–disease association to vary with alignment uncertainty. Standardization parameters and model coefficients are fitted on target training individuals and applied unchanged to held-out individuals. Individual outcomes support downstream model fitting, whereas target GWAS summary statistics support SNP-effect estimation. Entropy characterizes alignment ambiguity, not calibrated predictive variance. Appendix A.4 defines the individual score and predictor.

4 Experimental Setup

4.1 Datasets and prediction tasks.

We evaluated OT-PRS using two cohorts: UK Biobank for the primary cross-population prediction analysis and the All of Us Research Program for independent external validation. European-ancestry (EUR) GWAS summary statistics provided source association evidence, while African-ancestry (AFR) and East Asian-ancestry (EAS) populations served as targets. For each disease–population pair, OT-PRS combined source effects, target summary statistics, and ancestry-specific LD information to estimate target-population SNP effects.

The primary analysis used ancestry-specific GWAS summary statistics and individual-level UK Biobank genotype and phenotype data (Bycroft et al., 2018; Karczewski et al., 2025). We considered rheumatoid arthritis (RA), coronary artery disease (CAD), hypertension (HT), asthma (AM), and diabetes (DM). RA was evaluated only in AFR because EAS summary statistics were unavailable, yielding nine prediction tasks. Cohort characteristics and disease prevalences are provided in Appendix B.3, Table 2.

External validation used genotype and phenotype data from AFR and EAS participants in All of Us, with ancestry groups defined using the provided ancestry assignments. We evaluated the corresponding nine disease–population pairs using SNP effect estimates fixed before external evaluation. This analysis assessed whether the estimated genetic scores retained predictive discrimination in an independent cohort. All of Us cohort characteristics and disease prevalences are reported in Appendix B.3, Table 3.

4.2 PREPROCESSING.

For the primary analysis, we harmonized source and target summary statistics with ancestry-specific LD resources using chromosome, genomic position, and allele information. Variants were grouped into approximately independent LD blocks, filtered to remove missing or non-finite effect estimates and standard errors, and pruned using source-population LD. For large blocks, preliminary filtering by absolute source effect size reduced the number of variants entering pruning and transport. Details are provided in Appendices B.5 and B.6.

For All of Us, score variants were lifted from GRCh37 to GRCh38 and harmonized with genotype records by genomic position and alleles. Uniquely mapped autosomal variants with compatible alleles were retained. Chromosome-specific scores were computed using PLINK 2.0 and summed across chromosomes 1–22. Disease outcomes were defined using prespecified ICD-10 diagnosis codes.

4.3 BASELINES.

We compared OT-PRS with LDpred (Vilhjálmsson et al., 2015), PRS-CS (Ge et al., 2019), an unadjusted score based on target GWAS effects, and the multi-ancestry methods PRS-CSx (Ruan et al., 2022) and transPGS (Zhu et al., 2025). LDpred and PRS-CS used target-population summary statistics and LD references, whereas PRS-CSx and transPGS incorporated cross-population information.

4.4 IMPLEMENTATION.

OT-PRS was applied independently within each LD block. The learned FGW coupling defined a transported source-effect prior, which was refined using target summary statistics through the LD-aware MAP update. Refined effects were weighted by transport-derived reliability and concatenated across blocks to construct individual PRS. We used a fixed OT-PRS hyperparameter configuration across the evaluated disease–population pairs. Configurations for OT-PRS and all baseline methods are reported in Tables 5–6 in Appendix C.4.

4.5 PRIMARY EVALUATION.

We evaluated prediction in UK Biobank using five-fold cross-validation and report performance as mean and standard deviation across folds. Metrics included AUC, Nagelkerke's pseudo-R^2, and the odds ratio (OR), capturing discrimination, model fit, and the association between PRS and disease status, respectively. All methods were evaluated with adjustment for the top five target-population genetic principal components. Results are reported separately for each disease–population pair.

4.6 EXTERNAL VALIDATION.

We applied the frozen OT-PRS SNP effect estimates to ancestry-matched All of Us participants without re-estimating effects or tuning OT-PRS hyperparameters using external outcomes. Scores were standardized within each ancestry-specific external evaluation cohort. External discrimination was assessed using AUC from logistic regression models adjusted for age, sex, and the first five genetic principal components. This evaluation complements the UK Biobank analysis by testing the portability of the estimated scores across cohorts.

5 RESULTS

5.1 CROSS-POPULATION PREDICTION IN UK BIOBANK

We evaluated OT-PRS across nine disease–population tasks using EUR as the source population and AFR and EAS as targets. Figure 4 compares the complete framework, including target-summary-statistic refinement, with target-only and multi-ancestry baselines. OT-PRS achieved the highest mean AUC in six of nine tasks: RA and CAD in AFR, and CAD, hypertension, asthma, and diabetes in EAS. For example, OT-PRS achieved AUCs of 0.804 ± 0.024 for RA in AFR and 0.746 ± 0.034 for CAD in EAS, compared with 0.788 ± 0.028 and 0.716 ± 0.046, respectively, for PRS-CS.

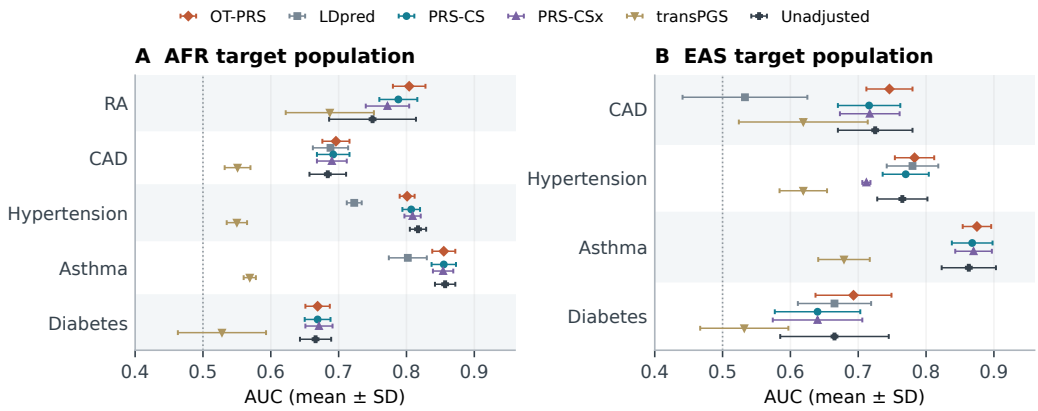


Figure 4: Cross-population prediction in UK Biobank. Points show mean AUC and error bars show standard deviations across five cross-validation folds for AFR (A) and EAS (B) targets.

Performance varied across tasks. In AFR, the unadjusted target-effect baseline achieved the highest mean AUC for hypertension (0.817 versus 0.801 for OT-PRS) and asthma (0.857 versus 0.855), while PRS-CSx performed best for diabetes (0.671 versus 0.669). Thus, OT-PRS was competitive across the evaluated tasks, with particularly consistent rankings in EAS, but did not improve every phenotype. These comparisons describe differences in mean performance; they do not establish statistical significance.

Nagelkerke's pseudo-R^2 provided a complementary assessment of model fit (Table 1). LDpred results were unavailable for RA–AFR, and asthma–EAS because its input-validation check flagged an excessively small mean GWAS chi-square statistic. OT-PRS had the highest value among the methods reported for this metric in six of nine tasks, including RA in AFR (0.428), CAD in EAS (0.370), and hypertension in EAS (0.461). PRS-CSx performed best for asthma in both populations, and PRS-CS performed best for diabetes in EAS. The rankings therefore depended on both the phenotype and the evaluation metric. Odds-ratio estimates further characterize PRS associations with disease status across eight available tasks (Appendix D.5, Figure 11).

Table 1: Nagelkerke's R^2 of baseline methods and OT-PRS across diseases and target populations in the label-informed setting.

Disease	Population	OT-PRS	LDpred	PRS-CS	PRS-CSx	transPGS
RA	AFR	**0.428**	–	0.321	0.323	0.001
CAD	AFR	**0.308**	0.254	0.261	0.288	0.005
	EAS	**0.370**	0.004	0.367	0.325	0.055
Hypertension	AFR	**0.373**	0.183	0.370	0.230	0.001
	EAS	**0.461**	0.354	0.448	0.396	0.055
Asthma	AFR	0.563	0.323	0.564	**0.567**	0.004
	EAS	0.653	–	0.663	**0.685**	0.077
Diabetes	AFR	**0.264**	–	0.259	0.259	0.005
	EAS	0.319	0.212	**0.323**	0.315	0.008

5.2 External validation in All of Us

We assessed cross-cohort portability by applying the fixed OT-PRS SNP effects to ancestry-matched participants in All of Us, without effect-size re-estimation or hyperparameter tuning using external outcomes. Figure 5 reports the AUCs of the external evaluation models, adjusted for age, sex, and the first five genetic principal components.

External AUCs ranged from 0.587 to 0.708 in AFR and from 0.615 to 0.820 in EAS. The highest values were observed for CAD in EAS (0.820), hypertension in EAS (0.789), and diabetes in EAS (0.765). For the four diseases evaluated in both ancestry groups, AUC was numerically higher in EAS than in AFR. Asthma showed weaker external discrimination than the other endpoints in both

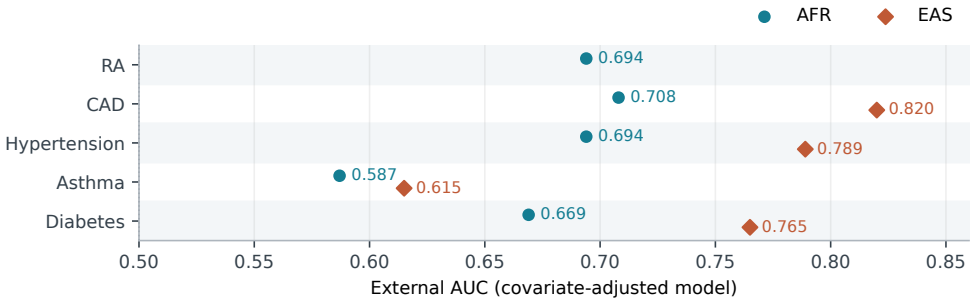


Figure 5: External validation of OT-PRS in All of Us. Points show AUCs of models containing the transferred score and adjustment covariates. SNP effects were fixed before external evaluation.

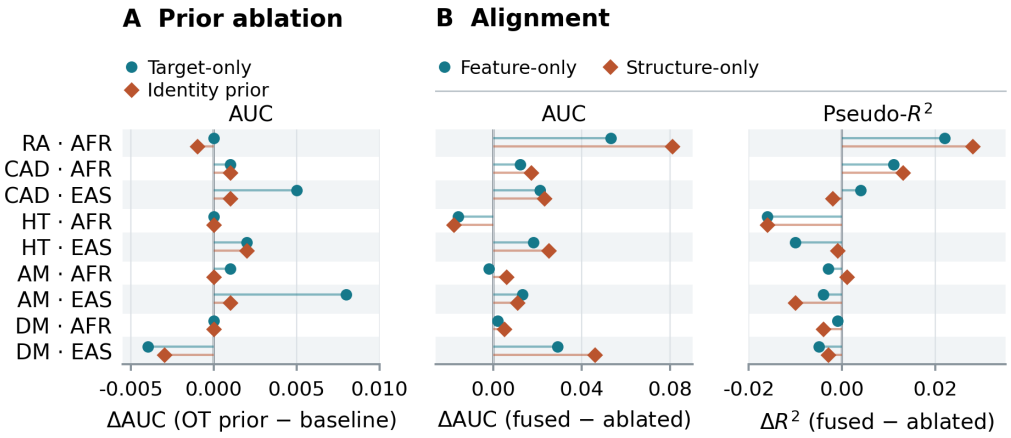


Figure 6: Ablation studies of OT-PRS. (A) Mean AUC differences between the OT-prior estimator and target-only or identity-prior estimators. (B) Differences in mean AUC (left) and Nagelkerke's pseudo-R^2 (right) between fused transport and feature-only or structure-only alignment.

groups. These findings support the portability of the estimated scores to an independent cohort, while showing substantial variation across diseases. The external analysis evaluates the transferred scores within covariate-adjusted models; it does not isolate their incremental discrimination over covariates alone.

5.3 Ablation studies

5.3.1 Contribution of the transport-derived prior.

We compared target-only, identity-prior, and OT-prior MAP estimators under identical preprocessing and evaluation conditions, excluding uncertainty components. The OT prior matched or exceeded target-only estimation in eight of nine tasks and the identity prior in seven (Figure 6A). Mean AUCs were 0.7691, 0.7677, and 0.7690 for OT-prior, target-only, and identity-prior estimation, respectively. Thus, OT redistribution provided modest, task-dependent gains beyond target evidence and direct source borrowing.

5.3.2 Feature and LD-structure alignment.

With target-effect refinement retained, fused transport achieved the highest mean AUC in seven of nine tasks (Figure 6B). For RA in AFR, AUC was 0.804, compared with 0.751 for feature-only and 0.723 for structure-only alignment. Exceptions were hypertension and asthma in AFR, while pseudo-R^2 rankings were less consistent.

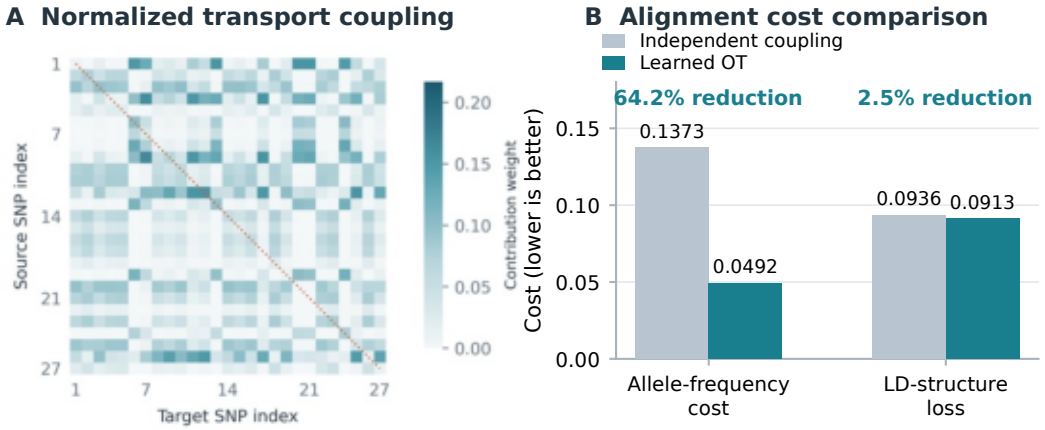


Figure 7: Illustration of learned transport for a CAD EUR-to-AFR LD block with 27 retained SNPs. (A) Normalized transport coupling showing source-SNP contributions to each target SNP. (B) Allele-frequency matching cost and LD-structure loss under an independent coupling with the same marginals and the learned OT coupling.

5.3.3 Uncertainty modeling and preprocessing.

Including the uncertainty–PRS interaction generally improved discrimination, without establishing calibration of transport entropy as prediction uncertainty. For RA in AFR, AUC increased from 0.782 with PRS alone to 0.804 with uncertainty and its interaction. Replacing inverse-variance marginals with uniform marginals changed mean AUC by at most 0.005, while pruning improved AUC in all nine tasks. Details appear in Appendices D.1–D.3.

5.4 Interpretation of learned transport

We examined a CAD EUR-to-AFR LD block containing 27 retained SNPs to illustrate the learned alignment. The normalized coupling assigns contributions from multiple source variants to each target SNP, showing how OT redistributes source information beyond direct SNP matching (Figure 7A). Compared with an independent coupling with the same marginals, the learned plan reduced allele-frequency matching cost from 0.1373 to 0.0492 (64.2%) and LD-structure mismatch from 0.0936 to 0.0913 (2.5%) (Figure 7B). Ancestry-specific LD matrices and effective contribution weights are provided in Figure 8. This example illustrates alignment under the transport objective, without establishing causal correspondence or predictive improvement; the predictive contribution of redistribution is evaluated separately through the matched-prior ablation (Figure 6A). Additional block-level details and genome-wide analyses of effect transformations and alignment entropy are provided in Appendix D.4.

6 Discussion and Conclusion

We introduced OT-PRS, an uncertainty-aware framework for cross-population polygenic risk prediction that combines block-wise fused Gromov–Wasserstein alignment, LD-aware refinement using target GWAS summary statistics, and transport-derived uncertainty modeling. Evaluations across nine disease–population tasks in UK Biobank showed competitive predictive performance, complemented by external validation in All of Us. Ablation analyses indicated that target evidence and regularized source borrowing account for much of the performance, with transport-based redistribution providing modest, task-dependent incremental gains. These findings support structured genomic alignment as a useful component of cross-population PRS estimation, while motivating further evaluation in diverse cohorts and calibration of transport-derived uncertainty.

AI USE STATEMENT

We used AI for language polishing, and manuscript organization. The authors take responsibility for the accuracy and integrity of the final manuscript.

ETHICS STATEMENT

This study uses publicly available GWAS summary statistics and controlled-access individual-level genotype and phenotype data from UK Biobank and the All of Us Research Program. UK Biobank participants provided informed consent, and the resource received ethical approval from the North West Multi-centre Research Ethics Committee. Our UK Biobank analyses were conducted under an approved application; identifying application details are withheld for anonymous review and will be disclosed in the final version. Individual-level data were accessed and analyzed in accordance with the respective resources' data-use agreements and governance requirements.

REPRODUCIBILITY STATEMENT

Section 3 describes the OT-PRS formulation, and Section 4 specifies the experimental setup. Appendix A provides implementation details and algorithmic procedures. Appendix B documents data sources, phenotype definitions, and preprocessing. Appendix C describes evaluation protocols, baseline configurations, OT-PRS hyperparameters, and computational settings, while Appendix D reports additional sensitivity analyses. External validation in All of Us uses fixed SNP effect estimates without effect re-estimation or hyperparameter tuning using external outcomes. Access to individual-level UK Biobank and All of Us data is governed by the respective resources and requires authorization.

REFERENCES

Tomaz Berisa and Joseph K. Pickrell. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics (Oxford, England)*, 32(2):283–285, January 2016. ISSN 1367-4811 1367-4803. doi: 10.1093/bioinformatics/btv546.

Clare Bycroft, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T Elliott, Kevin Sharp, Allan Motyer, Damjan Vukcevic, Olivier Delaneau, Jared O'Connell, et al. The uk biobank resource with deep phenotyping and genomic data. *Nature*, 562(7726):203–209, 2018.

Shing Wan Choi, Timothy Shin-Heng Mak, and Paul F. O'Reilly. Tutorial: a guide to performing polygenic risk score analyses. *Nature protocols*, 15(9):2759–2772, September 2020. ISSN 1750-2799 1754-2189. doi: 10.1038/s41596-020-0353-1.

Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1853–1865, 2017. doi: 10.1109/TPAMI.2016.2615921.

Yi Ding, Kangcheng Hou, Kathryn S Burch, Sandra Lapinska, Florian Privé, Bjarni Vilhjálmsson, Sriram Sankararaman, and Bogdan Pasaniuc. Large uncertainty in individual polygenic risk score estimation impacts prs-based risk stratification. *Nature genetics*, 54(1):30–39, 2022.

Frank Dudbridge. Power and predictive accuracy of polygenic risk scores. *PLoS genetics*, 9(3): e1003348, March 2013. ISSN 1553-7404 1553-7390. doi: 10.1371/journal.pgen.1003348.

Laramie Duncan, Hanyang Shen, Bizu Gelaye, J Meijsen, K Ressler, M Feldman, R Peterson, and Ben Domingue. Analysis of polygenic risk score usage and performance in diverse human populations. *Nature communications*, 10(1):3328, 2019.

Tian Ge, Chia-Yen Chen, Yang Ni, Yen-Chen Anne Feng, and Jordan W. Smoller. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature communications*, 10(1): 1776, April 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-09718-5.

Kangcheng Hou, Ziqi Xu, Yi Ding, Ravi Mandla, Zhuozheng Shi, Kristin Boulier, Arbel Harpak, and Bogdan Pasaniuc. Calibrated prediction intervals for polygenic scores across diverse contexts. *Nature Genetics*, 56(7):1386–1396, 2024.

Jiawei Huang, Yuxuan Du, Kevin R Kelly, Jinchi Lv, Yingying Fan, Jiang F Zhong, and Fengzhu Sun. Deepdeconuq estimates malignant cell fraction prediction intervals in bulk rna-seq tissue. *PLoS Computational Biology*, 21(6):e1013133, 2025.

Konrad J Karczewski, Rahul Gupta, Masahiro Kanai, Wenhan Lu, Kristin Tsuo, Ying Wang, Raymond K Walters, Patrick Turley, Shawneequa Callier, Nirav N Shah, et al. Pan-uk biobank genome-wide association analyses enhance discovery and resolution of ancestry-enriched effects. *Nature genetics*, 57(10):2408–2417, 2025.

Amit V. Khera, Mark Chaffin, Krishna G. Aragam, Mary E. Haas, Carolina Roselli, Seung Hoan Choi, Pradeep Natarajan, Eric S. Lander, Steven A. Lubitz, Patrick T. Ellinor, and Sekar Kathiresan. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature genetics*, 50(9):1219–1224, September 2018. ISSN 1546-1718 1061-4036. doi: 10.1038/s41588-018-0183-z.

Samuel A. Lambert, Gad Abraham, and Michael Inouye. Towards clinical utility of polygenic risk scores. *Human molecular genetics*, 28(R2):R133–R142, November 2019. ISSN 1460-2083 0964-6906. doi: 10.1093/hmg/ddz187.

Cathryn M. Lewis and Evangelos Vassos. Polygenic risk scores: from research tools to clinical instruments. *Genome medicine*, 12(1):44, May 2020. ISSN 1756-994X. doi: 10.1186/s13073-020-00742-5.

Luke R. Lloyd-Jones, Jian Zeng, Julia Sidorenko, Loïc Yengo, Gerhard Moser, Kathryn E. Kemper, Huanwei Wang, Zhili Zheng, Reedik Magi, Tõnu Esko, Andres Metspalu, Naomi R. Wray, Michael E. Goddard, Jian Yang, and Peter M. Visscher. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nature communications*, 10(1):5086, November 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-12653-0.

Carla Márquez-Luna, Po-Ru Loh, South Asian Type 2 Diabetes (SAT2D) Consortium, SIGMA Type 2 Diabetes Consortium, and Alkes L Price. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genetic epidemiology*, 41(8):811–823, 2017.

Alicia R. Martin, Masahiro Kanai, Yoichiro Kamatani, Yukinori Okada, Benjamin M. Neale, and Mark J. Daly. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nature genetics*, 51(4):584–591, April 2019. ISSN 1546-1718 1061-4036. doi: 10.1038/s41588-019-0379-x.

Roseann E. Peterson, Karoline Kuchenbaecker, Raymond K. Walters, Chia-Yen Chen, Alice B. Popejoy, Sathish Periyasamy, Max Lam, Conrad Iyegbe, Rona J. Strawbridge, Leslie Brick, Caitlin E. Carey, Alicia R. Martin, Jacquelyn L. Meyers, Jinni Su, Junfang Chen, Alexis C. Edwards, Allan Kalungi, Nastassja Koen, Lerato Majara, Emanuel Schwarz, Jordan W. Smoller, Eli A. Stahl, Patrick F. Sullivan, Evangelos Vassos, Bryan Mowry, Miguel L. Prieto, Alfredo Cuellar-Barboza, Tim B. Bigdeli, Howard J. Edenberg, Hailiang Huang, and Laramie E. Duncan. Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell*, 179(3):589–603, October 2019. ISSN 1097-4172 0092-8674. doi: 10.1016/j.cell.2019.08.051.

Gabriel Peyré and Marco Cuturi. Computational optimal transport with applications to data sciences. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

Gabriel Peyré, Marco Cuturi, and Justin Solomon. Gromov-wasserstein averaging of kernel and distance matrices. In *International conference on machine learning*, pp. 2664–2672. PMLR, 2016.

Yunfeng Ruan, Yen-Feng Lin, Yen-Chen Anne Feng, Chia-Yen Chen, Max Lam, Zhenglin Guo, Lin He, Akira Sawa, Alicia R. Martin, Shengying Qin, Hailiang Huang, and Tian Ge. Improving polygenic prediction in ancestrally diverse populations. *Nature genetics*, 54(5):573–580, May 2022. ISSN 1546-1718 1061-4036. doi: 10.1038/s41588-022-01054-7.

Iain R Timmins, PRACTICAL Consortium, and Frank Dudbridge. Bayesian approach to assessing population differences in genetic risk of disease with application to prostate cancer. *PLoS genetics*, 20(4):e1011212, 2024.

Ali Torkamani, Nathan E. Wineinger, and Eric J. Topol. The personal and clinical utility of polygenic risk scores. *Nature reviews. Genetics*, 19(9):581–590, September 2018. ISSN 1471-0064 1471-0056. doi: 10.1038/s41576-018-0018-x.

Titouan Vayer, Laetitia Chapel, Rémi Flamary, Romain Tavenard, and Nicolas Courty. Fused gromov-wasserstein distance for structured objects. *Algorithms*, 13(9):212, 2020.

Bjarni J. Vilhjálmsson, Jian Yang, Hilary K. Finucane, Alexander Gusev, Sara Lindström, Stephan Ripke, Giulio Genovese, Po-Ru Loh, Gaurav Bhatia, Ron Do, Tristan Hayeck, Hong-Hee Won, Sekar Kathiresan, Michele Pato, Carlos Pato, Rulla Tamimi, Eli Stahl, Noah Zaitlen, Bogdan Pasaniuc, Gillian Belbin, Eimear E. Kenny, Mikkel H. Schierup, Philip De Jager, Nikolaos A. Patsopoulos, Steve McCarroll, Mark Daly, Shaun Purcell, Daniel Chasman, Benjamin Neale, Michael Goddard, Peter M. Visscher, Peter Kraft, Nick Patterson, and Alkes L. Price. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *American journal of human genetics*, 97(4):576–592, October 2015. ISSN 1537-6605 0002-9297. doi: 10.1016/j.ajhg.2015.09.001.

Olga A Vsevolozhskaya and Dmitri V Zaykin. Quantifying posterior effect size distribution of susceptibility loci by common summary statistics. *Genetic epidemiology*, 44(4):339–351, 2020.

Omer Weissbrod, Masahiro Kanai, Huwenbo Shi, Steven Gazal, Wouter J. Peyrot, Amit V. Khera, Yukinori Okada, Alicia R. Martin, Hilary K. Finucane, and Alkes L. Price. Leveraging fine-mapping and multipopulation training data to improve cross-population polygenic risk scores. *Nature genetics*, 54(4):450–458, April 2022. ISSN 1546-1718 1061-4036. doi: 10.1038/s41588-022-01036-9.

Sean Whalen, Jacob Schreiber, William S. Noble, and Katherine S. Pollard. Navigating the pitfalls of applying machine learning in genomics. *Nature Reviews Genetics*, 23(3):169–181, March 2022. ISSN 1471-0056, 1471-0064. doi: 10.1038/s41576-021-00434-9. URL `https://www.nature.com/articles/s41576-021-00434-9`.

Yiyang Zhu, Wenying Chen, Kexuan Zhu, Yuxin Liu, Shuiping Huang, and Ping Zeng. Polygenic prediction for underrepresented populations through transfer learning by utilizing genetic similarity shared with european populations. *Briefings in Bioinformatics*, 26(1):bbaf048, 2025.

CONTENTS

A Methodological and implementation details **14**

A.1 LD blocks and variant retention 14

A.2 Genomic representation and transport inputs 14

A.3 Probabilistic refinement and MAP solution 14

A.4 Individual-level uncertainty and prediction 15

A.5 Algorithmic summary 15

B Data preparation and phenotype definitions **16**

B.1 Data access 16

B.2 Datasets and preprocessing 16

B.3 Cohort characteristics 16

B.4 UK Biobank phenotype definitions 16

B.5 Summary-statistic alignment 17

B.6 Genotype filters 17

B.7 All of Us score harmonization 17

C Evaluation and reproducibility **18**

C.1 Cross-validation and summary-statistic overlap 18

C.2 Evaluation model specification 18

C.3 Baseline implementation 18

C.4 OT-PRS parameters and software 18

C.5 Runtime and block-size sensitivity 19

D Additional results and sensitivity analyses **19**

D.1 Transport marginals 20

D.2 LD pruning 20

D.3 Individual-level uncertainty 20

D.4 Additional transport interpretation 21

D.5 PRS associations with disease 21

E Limitations and future work **21**

A METHODOLOGICAL AND IMPLEMENTATION DETAILS

This appendix expands the core model in Section 3, specifying transport inputs, the MAP solution, individual-level uncertainty, and the effect-estimation algorithm.

A.1 LD BLOCKS AND VARIANT RETENTION

We used EUR LDetect block boundaries in GRCh37 coordinates (Berisa & Pickrell, 2016), with ancestry-specific LD matrices computed within the same genomic intervals. For blocks containing at least 2,500 merged variants, we first retained up to 2,000 variants with the largest absolute source effect estimates. We then applied source-population LD pruning at $r^2 = 0.2$ and capped the retained block size at 200 SNPs. Table 6 reports the preprocessing parameters.

A.2 GENOMIC REPRESENTATION AND TRANSPORT INPUTS

For block k, allele frequencies, GWAS effects, and LD matrices are aligned to retained source and target variant orders, with consistent effect-allele coding. Let $f^{(k)}_{\mathrm{src},i}$ and $f^{(k)}_{\mathrm{tgt},j}$ denote effect-allele frequencies. The feature cost is

$$M^{(k)}_{ij} = \left(f^{(k)}_{\mathrm{src},i} - f^{(k)}_{\mathrm{tgt},j}\right)^2. \tag{7}$$

For the ancestry-specific LD correlation matrices $R^{(k)}_{\mathrm{src}}$ and $R^{(k)}_{\mathrm{tgt}}$, define

$$A^{(k)}_{ii'} = 1 - \left|R^{(k)}_{\mathrm{src},ii'}\right|^2, \qquad B^{(k)}_{jj'} = 1 - \left|R^{(k)}_{\mathrm{tgt},jj'}\right|^2. \tag{8}$$

These matrices encode pairwise LD dissimilarities within each population. The inverse-variance marginals, before any implementation-level clipping, are

$$p^{(k)}_i = \frac{\left(\mathrm{se}^{(k)}_{\mathrm{src},i}\right)^{-2}}{\sum_{i'=1}^{n_k}\left(\mathrm{se}^{(k)}_{\mathrm{src},i'}\right)^{-2}}, \qquad q^{(k)}_j = \frac{\left(\mathrm{se}^{(k)}_{\mathrm{tgt},j}\right)^{-2}}{\sum_{j'=1}^{m_k}\left(\mathrm{se}^{(k)}_{\mathrm{tgt},j'}\right)^{-2}}. \tag{9}$$

Retained standard errors must be finite and strictly positive. The implementation configuration specifies an inverse-variance clipping quantile of 0.99 (Table 6); uniform marginals are evaluated only as an ablation. The admissible transport plans are

$$\Pi(p^{(k)}, q^{(k)}) = \left\{T \geq 0 : T\mathbf{1}_{m_k} = p^{(k)},\ T^\top \mathbf{1}_{n_k} = q^{(k)}\right\}. \tag{10}$$

The squared-loss structural term in Eq. equation 1 is

$$\mathcal{L}_{\mathrm{GW}}\left(T; A^{(k)}, B^{(k)}\right) = \sum_{i,i'=1}^{n_k}\sum_{j,j'=1}^{m_k}\left(A^{(k)}_{ii'} - B^{(k)}_{jj'}\right)^2 T_{ij}T_{i'j'}. \tag{11}$$

The entropy regularizer uses $H(T) = -\sum_{i,j} T_{ij}\log T_{ij}$, with $0\log 0 = 0$. Physical SNP matching establishes compatible coordinates and allele coding; FGW additionally aligns local genomic relationships.

A.3 PROBABILISTIC REFINEMENT AND MAP SOLUTION

Let $\widehat{\beta}^{(k)}_{\mathrm{marg}}$ denote target marginal GWAS effects. Define their working covariance and precision by

$$\Sigma^{(k)} = \mathrm{diag}\left((\mathrm{se}^{(k)}_{\mathrm{tgt},1})^2, \ldots, (\mathrm{se}^{(k)}_{\mathrm{tgt},m_k})^2\right), \qquad W^{(k)} = (\Sigma^{(k)})^{-1}. \tag{12}$$

We use the Gaussian pseudo-likelihood and transported-effect prior

$$\widehat{\beta}^{(k)}_{\mathrm{marg}} \mid \beta \sim \mathcal{N}(R^{(k)}_{\mathrm{tgt}}\beta, \Sigma^{(k)}), \qquad \beta \sim \mathcal{N}(\beta^{(k)}_{\mathrm{tr}}, (\Lambda^{(k)})^{-1}). \tag{13}$$

Here, $R_{\mathrm{tgt}}^{(k)}\beta$ approximates the relationship between joint and marginal effects. LD enters through the forward mean model; the diagonal covariance is a computational approximation to the full LD-dependent covariance of marginal estimates. The positive-definite matrix $\Lambda^{(k)}$ controls shrinkage toward the transported prior from Eq. equation 2.

Taking the negative log pseudo-likelihood and log prior, then adding the ridge penalty, gives Eq. equation 3. Setting its gradient to zero yields

$$\Big[\Big(R_{\mathrm{tgt}}^{(k)}\Big)^{\top} W^{(k)} R_{\mathrm{tgt}}^{(k)} + \Lambda^{(k)} + \lambda_{\mathrm{ridge}} I\Big]\widehat{\beta}_{\mathrm{MAP}}^{(k)} = \Big(R_{\mathrm{tgt}}^{(k)}\Big)^{\top} W^{(k)} \widehat{\beta}_{\mathrm{marg}}^{(k)} + \Lambda^{(k)}\beta_{\mathrm{tr}}^{(k)}. \tag{14}$$

The system combines target marginal evidence and LD with the transport-derived prior. Its coefficient matrix is positive definite when $\Lambda^{(k)}$ is positive definite, so the regularized MAP estimate is unique. The estimate can be computed by solving this system without explicitly forming a matrix inverse.

A.4 Individual-level uncertainty and prediction

The normalized entropy in Eq. equation 4 uses $0 \log 0 = 0$. It is set to zero for a singleton source block to avoid division by $\log 1$. With $\gamma_{\mathrm{UQ}} = 1$, the reliability weight $w_j = \exp(-\gamma_{\mathrm{UQ}}\widetilde{U}_j)$ lies in $[e^{-1}, 1]$. This attenuation rule depends on the transport marginals and regularization; it is not a posterior precision or a calibrated probability of correct transfer.

The block-wise reliability weights and attenuated effects are

$$w_j^{(k)} = \exp\Big(-\gamma_{\mathrm{UQ}}\widetilde{U}_j^{(k)}\Big), \qquad \widetilde{\beta}_{\mathrm{tgt},j}^{(k)} = w_j^{(k)}\widehat{\beta}_{\mathrm{MAP},j}^{(k)}, \tag{15}$$

After concatenating blocks, let f_j denote the target effect-allele frequency. We aggregate alignment uncertainty as

$$\mathrm{UQ}_i = \frac{\sum_{j=1}^{m}(g_{ij} - 2f_j)^2 \left(\widehat{\beta}_{\mathrm{MAP},j}\right)^2 \widetilde{U}_j}{\sum_{j=1}^{m}(g_{ij} - 2f_j)^2 \left(\widehat{\beta}_{\mathrm{MAP},j}\right)^2 + \eta}, \tag{16}$$

The constant $\eta > 0$ ensures numerical stability. Variants contribute more when squared genotype deviations and squared refined effects are larger. The formula uses unweighted MAP effects, rather than the attenuated effects used for PRS. The resulting score summarizes individual exposure to alignment ambiguity; it is not a predictive variance or confidence interval.

Let $\mathrm{PRS}_{i,Z}$ and $\mathrm{UQ}_{i,Z}$ denote standardized scores. Their contribution to the downstream predictor is

$$S_i = \theta_1 \mathrm{PRS}_{i,Z} + \theta_2 \mathrm{UQ}_{i,Z} + \theta_3 \mathrm{PRS}_{i,Z}\mathrm{UQ}_{i,Z}. \tag{17}$$

An intercept and the adjustment covariates specified in Appendix C.2 complete the phenotype model. The interaction allows the association between PRS and phenotype to vary with alignment uncertainty. Standardization parameters and downstream coefficients are fitted using labeled target training individuals and applied unchanged to held-out individuals. These individual labels are distinct from the target GWAS summary statistics used for SNP-effect estimation.

A.5 Algorithmic summary

Algorithm 1 summarizes effect estimation and scoring. Downstream phenotype-model fitting uses the resulting PRS and uncertainty features under the evaluation protocol in Appendix C.1.

Table 2: Characteristics of the ancestry-specific UK Biobank cohorts. Disease prevalence and female participants are reported as percentages; age is reported as mean (SD). RA prediction was not evaluated in EAS.

Characteristic	EUR	AFR	EAS
Participants	472,463	8,052	1,573
Age, years	56.76 (8.03)	51.94 (8.08)	52.48 (7.71)
Female (%)	54.46	57.71	62.87
CAD (%)	7.41	4.55	3.06
RA (%)	1.47	1.56	–
AM (%)	11.62	12.00	8.79
HT (%)	26.00	38.03	20.80
DM (%)	3.94	10.08	4.69

Algorithm 1: OT-PRS effect estimation and scoring

Inputs: Source and target GWAS effects, standard errors, allele frequencies, ancestry-specific LD, LD blocks, target genotypes, and parameters including $\Lambda^{(k)}$.

1. Harmonize alleles and assign variants to LD blocks.
2. For each block, filter variants, apply LD pruning, and enforce the size cap.
3. Construct feature and structural costs and inverse-variance marginals.
4. Estimate the FGW coupling using Eq. equation 1.
5. Construct the prior mean using Eq. equation 2.
6. Refine target effects using Eq. equation 14.
7. Compute entropy using Eq. equation 4, then reliability weights and weighted effects as in Eq. equation 15.
8. After processing all blocks, concatenate effects and uncertainty quantities.
9. Compute individual scores using Eqs. equation 5 and equation 16.

Outputs: Target SNP weights, individual PRS, and uncertainty scores.

B Data preparation and phenotype definitions

B.1 Data access

UK Biobank participants provided informed consent, and the resource has ethical approval from the North West Multi-centre Research Ethics Committee. Individual-level analyses were conducted under an approved UK Biobank application; the application number is withheld for anonymous review. Ancestry-specific summary statistics were obtained from Pan-UK Biobank (Bycroft et al., 2018; Karczewski et al., 2025). External validation used controlled-access genotype and phenotype data from All of Us.

B.2 Datasets and preprocessing

B.3 Cohort characteristics

Table 2 summarizes the UK Biobank cohorts used in this study.

Table 3 summarizes the All of Us cohorts used for external validation in this study.

B.4 UK Biobank phenotype definitions

Case status was defined using the ICD-10 and self-reported codes in Table 4. Participants required genotype data, a corresponding ancestry assignment, and nonmissing disease status. Exclusions were applied separately for each disease. Diabetes combined the disease categories represented by the listed codes.

Table 3: Characteristics of the ancestry-specific All of Us external-validation cohorts. Disease prevalence and female participants are reported as percentages; age is reported as mean (SD). RA prediction was not evaluated in EAS.

Characteristic	AFR	EAS
Participants	78,673	12,034
Age, years	55.21 (14.92)	48.95 (16.84)
Female (%)	60.53	63.21
CAD (%)	12.51	6.57
RA (%)	2.51	–
AM (%)	16.01	8.63
HT (%)	42.10	21.33
DM (%)	22.34	10.61

Table 4: UK Biobank phenotype definitions. Diagnosis and self-reported codes were drawn from fields 41270 and 20002, respectively.

Disease	ICD-10 codes	Self-reported codes
CAD	I20, I21, I22, I23, I24, I25	1074, 1075
RA	M05, M06	1046
Hypertension	I10, I11, I12, I13, I14, I15	1065, 1072
Diabetes	E10, E11, E12, E13, E14	1220, 1222, 1223
Asthma	J45, J46	1111

B.5 Summary-statistic alignment

UK Biobank summary statistics and LD resources were represented in GRCh37. We standardized chromosome, position, reference and alternate alleles, effect estimates, and standard errors, and matched records using $(\mathrm{chr}, \mathrm{pos}, \mathrm{ref}, \mathrm{alt})$. Variants with invalid effect estimates or standard errors were excluded. The retained variants were aligned with the genotype and LD resources before applying the block-level retention procedure in Appendix A.1.

B.6 Genotype filters

Genotype preprocessing retained autosomal variants and removed strand-ambiguous SNPs (A/T and C/G) and duplicate SNP identifiers. Where minor-allele frequencies were available, variants with MAF below 0.01 were excluded. Some analyses additionally restricted variants to HapMap3 membership; this was not a universal filter. Ancestry-specific LD matrices were stored in HDF5 format for block-wise access.

B.7 All of Us score harmonization

Frozen SNP weights were lifted from GRCh37 to GRCh38 using the hg19-to-hg38 chain file. Only uniquely mapped autosomal variants were retained. Lifted variants were aligned to All of Us ACAF genotype records using position and REF/ALT alleles. At multiallelic sites, the corresponding effect allele was matched against the original chromosome-specific variant records. Negative-strand mappings were handled by complementing alleles while retaining the corresponding effect estimate. Scores were computed using PLINK 2.0 for chromosomes 1–22 and summed within each participant. Standardization was performed after restriction to the final ancestry-specific evaluation cohort. Disease phenotypes were defined using prespecified OMOP concept definitions.

C Evaluation and reproducibility

C.1 Cross-validation and summary-statistic overlap

Five-fold cross-validation was applied to downstream phenotype-model fitting within each UK Biobank target cohort. Models were fitted in the training portion and evaluated in the held-out fold; reported standard deviations describe variation across the five test folds. This split does not itself separate participants used to estimate the GWAS summary statistics. Because Pan-UK Biobank statistics and the evaluation genotypes originate from UK Biobank, potential discovery–evaluation overlap limits interpretation of these results as independent out-of-sample prediction. All of Us provides a separate external cohort for evaluating the transferred SNP effects.

C.2 Evaluation model specification

UK Biobank models adjusted for the first five genetic principal components. All of Us models additionally adjusted for age and sex. For a PRS association model without the uncertainty interaction, the OR per standard-deviation increase in PRS is $\exp(\theta_{\mathrm{PRS}})$, where θ_{PRS} is the coefficient of the standardized score in the covariate-adjusted logistic regression. This coefficient is distinct from a constant PRS effect in a model containing a PRS–uncertainty interaction.

For the likelihood-based comparison, let ℓ_0 denote the log likelihood of the covariate-only null model and ℓ_1 that of the model including the genetic predictors under evaluation. With n observations, the reported pseudo-R^2 is

$$R^2_{\mathrm{Nag}} = \frac{1 - \exp[2(\ell_0 - \ell_1)/n]}{1 - \exp(2\ell_0/n)}. \tag{18}$$

This statistic measures improvement in model fit relative to the specified null model; it is not a literal proportion of variance in the binary outcome explained by PRS.

C.3 Baseline implementation

LDpred. We used LDpred 1.0.11 with the `coord`, `gibbs`, and `score` steps, target summary statistics, and ancestry-matched LD references. Custom input fields were `rsid`, `chr`, `pos`, `a1`, `a0`, `beta`, and `p`. Each run used the corresponding GWAS sample size. A runtime compatibility patch supported the legacy release without changing its inference procedure. For asthma–EAS, and diabetes–AFR, LDpred terminated with an input-validation error indicating an excessively small mean GWAS chi-square statistic. The diagnostic identified the summary statistics, their parsing, or the supplied GWAS sample size as possible causes. These runs are reported as unavailable rather than assigned numerical performance values.

PRS-CS and PRS-CSx. PRS-CS used target summary statistics and ancestry-matched UK Biobank LD blocks. PRS-CSx jointly used EUR and target summary statistics, ancestry-specific sample sizes, and the PRS-CSx LD reference resources. Both used an ancestry-matched target `bim` file for SNP alignment. The unadjusted baseline used target marginal GWAS effects directly.

transPGS. We used an adapted, chunk-wise version of the authors' summary-level implementation. EUR variants were pruned with PLINK using `--indep-pairwise 200 50 0.1` and intersected with target summary statistics. Source and target genotype matrices and the precomputed transport matrix were aligned by exact `snp_id`, with `rsid` as a fallback. Chunks contained up to 5,000 SNPs. Variants with more than 5% missingness in either ancestry were excluded; remaining missing genotypes were mean-imputed, and zero-variance or non-finite-variance SNPs were removed. Chunk-wise effect estimates were concatenated for scoring. The comparison therefore reflects this adapted implementation.

C.4 OT-PRS parameters and software

We used the same OT-PRS parameter configuration across diseases and target populations (Table 6). OT-PRS was implemented in Python using NumPy, pandas, h5py, and the Python Optimal Transport

Table 5: Baseline implementation settings. GWAS sample sizes were ancestry- and phenotype-specific.

Method	Parameter	Value
LDpred	LD radius	1000
	Causal-fraction grid	$\{1.0, 0.5, 0.1, 0.01, 0.001\}$
	Effect / input format	LINREG / CUSTOM
PRS-CS	Global shrinkage ϕ	0.01
	Iterations	1000
	LD reference	Ancestry-matched UK Biobank blocks
PRS-CSx	Populations	EUR and target ancestry
	Global shrinkage ϕ	0.01
	Iterations	1000
	LD reference	Public PRS-CSx reference resources
transPGS	Chunk size	5,000 SNPs
	EUR LD pruning	`--indep-pairwise 200 50 0.1`
	Missingness threshold	0.05 in either ancestry
	Missing-data handling	Mean imputation within chunk

library. PLINK was used for genotype preprocessing and PRS scoring, and transPGS was run in R. OT-PRS used four CPU threads unless otherwise specified.

Table 6: OT-PRS implementation parameters.

Component	Parameter	Value
Transport	Feature/structure tradeoff α	0.5
	Entropic regularization ε	0.05
	Marginals	Inverse-variance
	Inverse-variance clipping quantile	0.99
Refinement	Prior-strength parameter λ_{OT}	0.2
	Ridge penalty λ_{ridge}	0.01
Uncertainty	Entropy attenuation γ_{UQ}	1.0
Preprocessing	Large-block prefilter threshold	2,500 SNPs
	Variants retained before pruning	2,000 SNPs
	LD pruning threshold r^2	0.2
	Maximum retained block size	200 SNPs
Numerical stability	Transport denominator δ	10^{-12}

C.5 Runtime and block-size sensitivity

Table 7 reports the supplied CAD–AFR wall-clock measurements under the same CPU environment. OT-PRS timing includes block-wise preprocessing, transport, MAP refinement, and output generation; baseline timings include model fitting and score generation. These are illustrative measurements for one task and the stated implementations.

A separate block-size sensitivity analysis varied the maximum retained block size from 200 to 800 SNPs and included a no-pruning configuration. Runtimes ranged from 399.84 to 443.55 seconds, with reported peak memory of 1.42 GB in every run. This benchmark is reported separately from the CAD–AFR comparison because the supplied measurements differ; it does not establish a general scaling law.

D Additional results and sensitivity analyses

These tables provide the detailed comparisons underlying the sensitivity summaries in the main results. Numerical entries retain the supplied results; bold identifies the larger reported value within each comparison, including ties.

Table 7: Illustrative CAD–AFR runtime comparison. OT-PRS used four CPU threads.

Method	Wall-clock time
OT-PRS	5 min 34 s
LDpred	10 min 58 s
PRS-CS	35 min 41 s
PRS-CSx	1 h 44 min 04 s
transPGS (chunk-wise)	7 h 19 min 54 s

D.1 TRANSPORT MARGINALS

We replaced the default inverse-variance marginals with uniform weights, $p_i^{(k)} = 1/n_k$ and $q_j^{(k)} = 1/m_k$, keeping the remaining pipeline fixed. Table 8 reports the complete comparison. Absolute differences in mean AUC were at most 0.005, and pseudo-R^2 differences were at most 0.002.

Table 8: Transport-marginal sensitivity. AUC variability is the standard deviation across five folds.

		AUC (mean ± SD)		Pseudo-R^2	
Disease	Population	Inv.-variance	Uniform	Inv.-variance	Uniform
RA	AFR	0.804 ± 0.024	$\mathbf{0.809 \pm 0.037}$	**0.428**	**0.428**
CAD	AFR	$\mathbf{0.696 \pm 0.020}$	0.695 ± 0.022	**0.308**	**0.308**
CAD	EAS	$\mathbf{0.746 \pm 0.034}$	0.741 ± 0.035	**0.370**	0.369
Hypertension	AFR	$\mathbf{0.801 \pm 0.011}$	$\mathbf{0.801 \pm 0.011}$	**0.373**	0.371
Hypertension	EAS	$\mathbf{0.783 \pm 0.029}$	0.781 ± 0.029	**0.461**	0.460
Asthma	AFR	$\mathbf{0.855 \pm 0.018}$	0.854 ± 0.019	**0.563**	0.561
Asthma	EAS	$\mathbf{0.875 \pm 0.021}$	0.874 ± 0.017	0.653	**0.655**
Diabetes	AFR	0.669 ± 0.019	$\mathbf{0.670 \pm 0.018}$	**0.264**	0.263
Diabetes	EAS	$\mathbf{0.693 \pm 0.056}$	0.691 ± 0.056	**0.319**	**0.319**

D.2 LD PRUNING

We compared the default preprocessing with a no-pruning variant (Table 9). The default configuration had higher mean AUC in every task; pseudo-R^2 was higher in eight tasks and unchanged for asthma in EAS. This comparison concerns the preprocessing configurations as a whole and does not separately identify the effect of every filtering or block-size decision.

Table 9: LD-pruning sensitivity. AUC variability is the standard deviation across five folds.

		AUC (mean ± SD)		Pseudo-R^2	
Disease	Population	Pruned	No pruning	Pruned	No pruning
RA	AFR	$\mathbf{0.804 \pm 0.024}$	0.766 ± 0.042	**0.428**	0.417
CAD	AFR	$\mathbf{0.696 \pm 0.020}$	0.690 ± 0.028	**0.308**	0.292
CAD	EAS	$\mathbf{0.746 \pm 0.034}$	0.726 ± 0.019	**0.370**	0.365
Hypertension	AFR	$\mathbf{0.801 \pm 0.011}$	0.773 ± 0.014	**0.373**	0.289
Hypertension	EAS	$\mathbf{0.783 \pm 0.029}$	0.774 ± 0.033	**0.461**	0.454
Asthma	AFR	$\mathbf{0.855 \pm 0.018}$	0.838 ± 0.021	**0.563**	0.516
Asthma	EAS	$\mathbf{0.875 \pm 0.021}$	0.866 ± 0.038	**0.653**	**0.653**
Diabetes	AFR	$\mathbf{0.669 \pm 0.019}$	0.667 ± 0.012	**0.264**	0.234
Diabetes	EAS	$\mathbf{0.693 \pm 0.056}$	0.660 ± 0.042	**0.319**	0.316

D.3 INDIVIDUAL-LEVEL UNCERTAINTY

We compared three downstream models using the same PRS: PRS alone, PRS with individual uncertainty (UQ), and PRS with UQ and their interaction (Table 10). This ablation concerns downstream

uncertainty features; it does not remove entropy-based SNP weighting from the PRS. The interaction model matched or exceeded the other models' reported AUC in all nine tasks. These comparisons assess predictive contribution, not uncertainty calibration.

Table 10: Contribution of individual uncertainty to downstream prediction. Columns report AUC and Nagelkerke's pseudo-R^2 for the three models.

		PRS		PRS + UQ		With interaction	
Disease	Population	AUC	R^2	AUC	R^2	AUC	R^2
RA	AFR	0.782	0.420	0.800	0.423	**0.804**	**0.427**
CAD	AFR	0.692	0.302	0.694	0.303	**0.696**	**0.308**
CAD	EAS	0.742	0.366	**0.746**	0.367	**0.746**	**0.370**
Hypertension	AFR	0.798	0.359	0.798	0.359	**0.801**	**0.373**
Hypertension	EAS	0.781	0.459	0.781	0.459	**0.783**	**0.461**
Asthma	AFR	0.848	0.552	0.848	0.552	**0.855**	**0.563**
Asthma	EAS	0.872	0.651	0.873	0.651	**0.875**	**0.653**
Diabetes	AFR	**0.669**	0.258	**0.669**	0.258	**0.669**	**0.264**
Diabetes	EAS	0.686	0.317	**0.693**	**0.319**	**0.693**	**0.319**

D.4 Additional transport interpretation

Block-level genomic context. Figure 8 provides ancestry-specific LD matrices and effective contribution weights for the CAD EUR-to-AFR block examined in Section 5.4.

Genome-wide effect transformations. We additionally examined 84,211 SNPs across chromosomes 1–22 in the CAD EUR-to-AFR weight output (Figure 9). The same-SNP Spearman correlation between source and transported effects was 0.143, while target marginal and refined effects had a correlation of 0.876. Because transport pools information across source SNPs, the same-SNP comparison does not directly measure alignment quality. Uncertainty weighting attenuated refined effects with a median weight of 0.397, while largely preserving their ranks ($\rho = 0.999$). These comparisons describe coefficient transformations; they do not isolate the predictive contribution of each component.

Genome-wide alignment ambiguity. Median normalized transport entropy was 0.924, and 68.6% of SNPs had entropy above 0.90, indicating predominantly diffuse alignments under the fitted transport configuration. Figure 10 shows their distribution across chromosomes. The 0.90 cutoff is descriptive, not a validated reliability threshold. High entropy reflects alignment ambiguity under the chosen marginals and regularization; it does not establish poor predictive performance or calibrated prediction uncertainty.

D.5 PRS associations with disease

Figure 11 summarizes the reported odds ratios across eight disease–population pairs. Estimates ranged from 2.687 for RA in AFR to 8.012 for asthma in EAS, with all reported intervals excluding one. These results complement the discrimination analyses by characterizing associations between PRS and disease status. RA was evaluated only in AFR, and the diabetes–EAS estimate was unavailable.

E Limitations and future work

Evaluation scope. Our evaluation covers five diseases and two non-European target ancestry groups. Future work should assess OT-PRS across broader phenotypes, ancestries, and cohorts, particularly populations underrepresented in current GWAS resources.

Covariance approximation. MAP refinement uses a diagonal working covariance for target marginal GWAS estimates while incorporating LD through the mean model. Future extensions

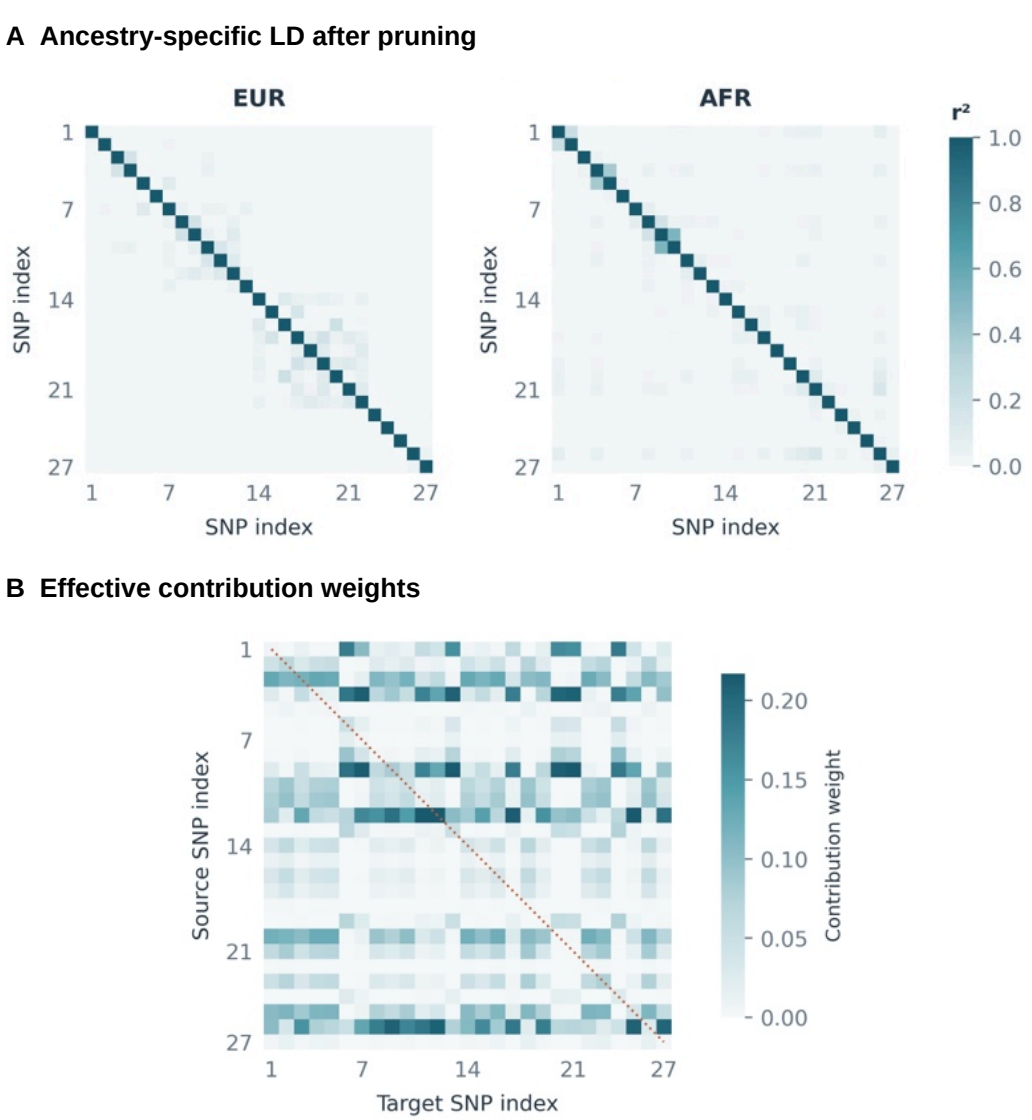


Figure 8: Additional details for the CAD EUR-to-AFR LD block in UK Biobank shown in Figure 7. (A) Ancestry-specific LD matrices after pruning, shown as pairwise r^2 for EUR and AFR. (B) Effective source-to-target contribution weights for the same block. Rows and columns index source and target SNPs, respectively.

could incorporate richer covariance models and scalable approximations to better capture dependence among association estimates while maintaining computational efficiency.

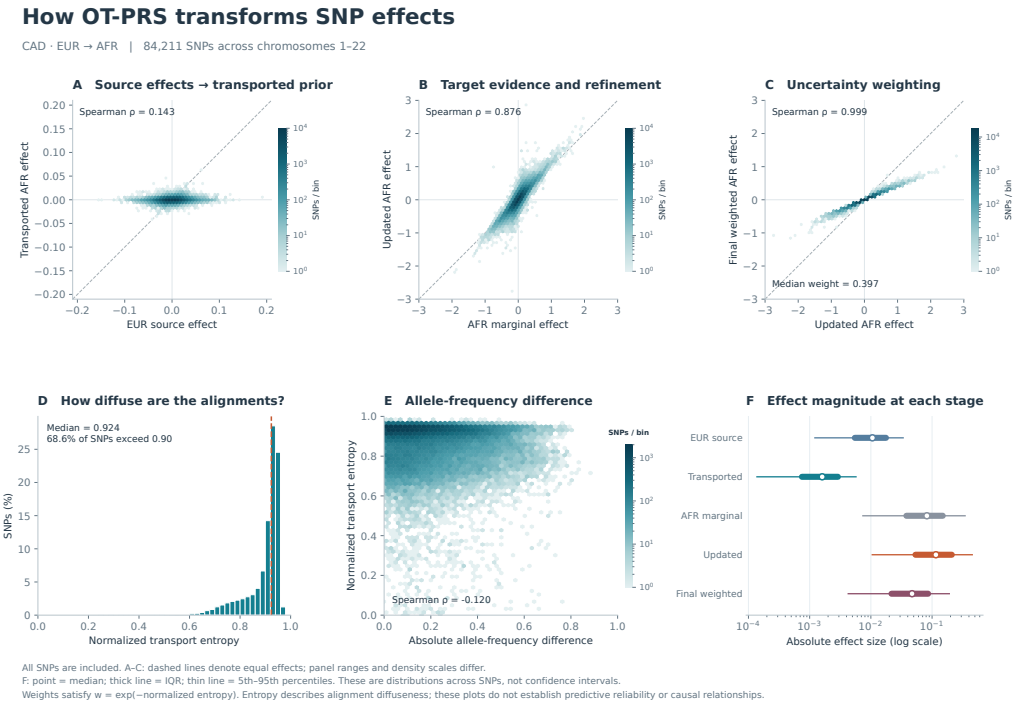


Figure 9: Effect transformations and alignment uncertainty across 84,211 CAD EUR-to-AFR SNPs. (A–C) Source versus transported, target marginal versus refined, and refined versus final weighted effects. Dashed lines indicate equality; colors indicate SNP counts on logarithmic scales. (D) Normalized transport entropy. (E) Absolute allele-frequency differences versus entropy. (F) Absolute effect magnitudes: points show medians, thick lines interquartile ranges, and thin lines fifth–ninety-fifth percentiles. These intervals describe variation across SNPs, not confidence intervals.

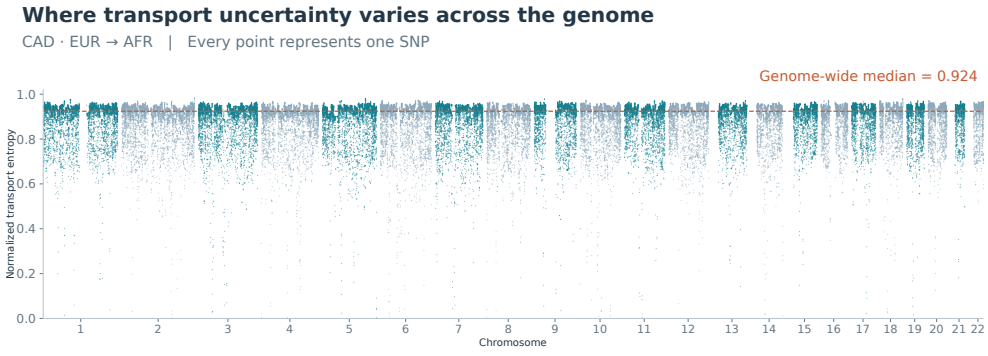


Figure 10: Genome-wide normalized transport entropy for CAD EUR-to-AFR. Each point represents a SNP ordered by genomic position within chromosome; alternating colors distinguish chromosomes. The dashed line marks the median entropy (0.924). Chromosome widths reflect observed positions rather than exact reference lengths. This is an alignment-uncertainty profile, not a GWAS association plot.

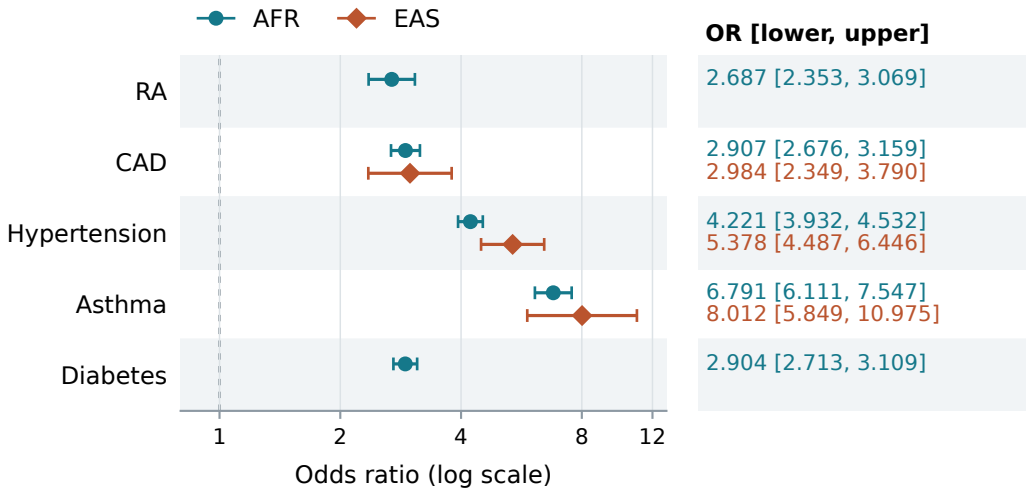


Figure 11: Odds ratios for OT-PRS across eight disease–population pairs in UK Biobank. Points show estimates and horizontal bars indicate the reported lower and upper interval limits. Colors and marker shapes distinguish AFR and EAS. The horizontal axis is logarithmic; the dashed line indicates an odds ratio of one. RA was evaluated only in AFR, and the diabetes–EAS estimate was unavailable.