ImputePGTA: accurate embryo genotyping and polygenic scoring from ultra-low-pass sequencing

Jeremiah H. Li[1], Tobias Wolfram[1], Ivan Davidson[1], Justin Schleede[1], Jennifer Swift[1], Spencer Moore[1], David Stern[1], Michael Christensen[1], Alexander Strudwick Young[1,2]

[1]*Herasight Research, USA*
[2]*Human Genetics Department, UCLA David Geffen School of Medicine, Los Angeles, CA, USA*

Abstract

Preimplantation genetic testing for polygenic risk (PGT-P) holds great promise for reducing lifetime disease burden, but has been held back by the difficulty of genotyping embryos. Preimplantation genetic testing for aneuploidy (PGT-A) is a standard-of-care technology used in over half of in vitro fertilization (IVF) cycles in the United States. PGT-A is used to detect chromosomal abnormalities using ultra-low-pass (ULP) sequencing data (typically 0.002x to 0.006x) or, less commonly, genotyping array-based data. Here we describe ImputePGTA, a Hidden Markov Model-based algorithm that enables accurate reconstruction of embryo genomes from array or ULP sequencing data from embryos and parental genome data. A key innovation of our algorithm is its ability to provide accurate embryo genotypes and polygenic scores (PGSs) along with posterior distributions given limited embryo data and imperfectly phased parental haplotypes, as encountered in real-world applications. The accuracy of the embryo genome reconstruction increases with that of the phasing quality of parental haplotypes. We describe a method, phaseGrafter, that improves parental phasing by combining statistical phasing from short-reads with read-backed phasing from long-reads, which further enable phasing of rare pathogenic variants. We validate our results through simulations, downsampled gold standard data, and comparison of six reconstructed embryo genomes from real PGT-A data to high-coverage, post-birth whole genome sequencing data. Our imputed embryo genotypes have a dosage correlation of 0.961 with high-quality post-birth genotypes (0.998 when using embryo array data). The imputed embryo polygenic scores for 17 diseases have a mean absolute difference of 0.16 standard deviations (0.023 when using embryo array data) with PGSs calculated from high-quality post-birth genotypes, lower than from imputation of array data from reference panels. We show that the attenuation in expected gains from embryo selection due to posterior uncertainty is only ~5-10% for typical PGT-A data. Our approach removes an important technological barrier to using PGT-P and will facilitate more widespread adoption.

Introduction

In vitro fertilization (IVF) has expanded rapidly, reaching 432,641 cycles in the United States alone in 2023 and accounting for 2.6% of all births that year[1]. This rise is driven in part by trends in elective oocyte cryopreservation[2] and increasing infertility[3], as well as the growing demand for pre-implantation genetic testing (PGT). Initially, PGT addressed monogenic disorders (PGT-M), enabling at-risk parents to select embryos free of severe genetic diseases such as cystic fibrosis or Huntington's disease[4]. More recently, clinical focus has broadened to PGT for structural rearrangements (PGT-SR) and aneuploidy (PGT-A). PGT-A is a standard offering that increases IVF success rates by screening embryos for chromosomal abnormalities[5]. In the United States, PGT-A was performed in 14% of IVF cycles in 2014, rising to 59% in 2022[6,7].

Most diseases and traits are influenced by many variants across the genome — i.e., they are polygenic and can be predicted using polygenic scores (PGSs) that weight genotypes at different variants based on estimates of the variants' effects. Polygenic embryo screening (PGT-P) uses polygenic scores to predict the traits and disease risks of the embryos if implanted, giving information that can be used to choose which embryos to implant. PGT-P has the potential to reduce lifetime disease burden[8] and is currently offered by multiple companies in the United States[9–14]. Despite ongoing ethical and scientific debates[15–18], public support for PGT-P is high[19–21], and the utility of PGT-P is likely to improve with the growth of biobank data and the development of new methods for polygenic prediction[22]. PGT-P, however, requires comprehensive genomic data — typically whole-genome sequencing or genotyping array data — derived from limited embryo DNA, which necessitates costly and specialized wet-lab techniques and analytical methods that are not yet widely available[23]. One approach is to genotype embryos at the set of markers on an array and then to impute missing genotypes from a reference panel. While this approach can capture most common genetic variation used in PGSs, they miss high-effect rare variants that can substantially contribute to disease risk[24,25] and can introduce imputation errors, especially for ancestries not well-represented in reference panels[26].

Enabling analyses of the embryo's entire inherited genome with the data generated during standard PGT-A could dramatically lower the barriers to PGT-P. Standard PGT-A protocols, however, generate ultra-low-pass (ULP) sequencing data, resulting in coverages as low as 70,000 100bp single-end reads[27] ($\sim$0.002x). While this is sufficient for detecting large-scale chromosomal abnormalities, obtaining reliable genotype calls at individual variants using standard methods requires significantly higher coverage. Though several recent studies[28,29] have repurposed historical PGT-A data to perform genome-wide association studies (GWASs), these studies aggregate sparse information across thousands of individuals, and neither rely on nor produce accurate individual genotype calls[30,31]. Thus far, sequence-based PGT-A data alone have not been used for PGT-P[32].

In other contexts, low-pass sequencing combined with reference-panel-based genotype imputation is routinely used to impute individual genomes from depths as low as $\sim$0.1x[33–36]. However, accuracy deteriorates rapidly with increasingly rare variants and for genetic ancestries poorly represented in the reference panel. Thus, these methods are useful primarily for cohort-level analyses where accurate individual genotype calls are not required and are not suitable for clinical applications.

In contrast to population studies, embryo genotyping in the context of IVF benefits from the fact that parental genomes are routinely available for sequencing. Each embryo's genome is a mosaic of maternal and paternal haplotypes, implying that comprehensive reconstruction could be feasible via identification of the inheritance vectors describing the parental haplotypes inherited by individual embryo at each genomic position. Leveraging parental haplotypes therefore offers a solution to the sparse data problem posed by ULP embryo sequencing. Existing methods such as genotyping-array-based reconstruction[10] and karyomapping[37] (both based on a 300k SNP chip), hybrid schemes such as MARSALA[38] (> 1x whole genome plus targeted regions), and low-coverage pipelines like (S)Haploseek[39,40] ($0.2 - 5$x) and haplarithmisis[41] (10x) can reconstruct embryo haplotypes on at least portions of the genome (though the breadth of variation that can be reconstructed differs significantly depending on the assay). However, these methods all assume the parents are phased perfectly — a prohibitively strong assumption — and operate on embryo data with orders of magnitude higher coverage than standard ULP PGT-A sequencing.

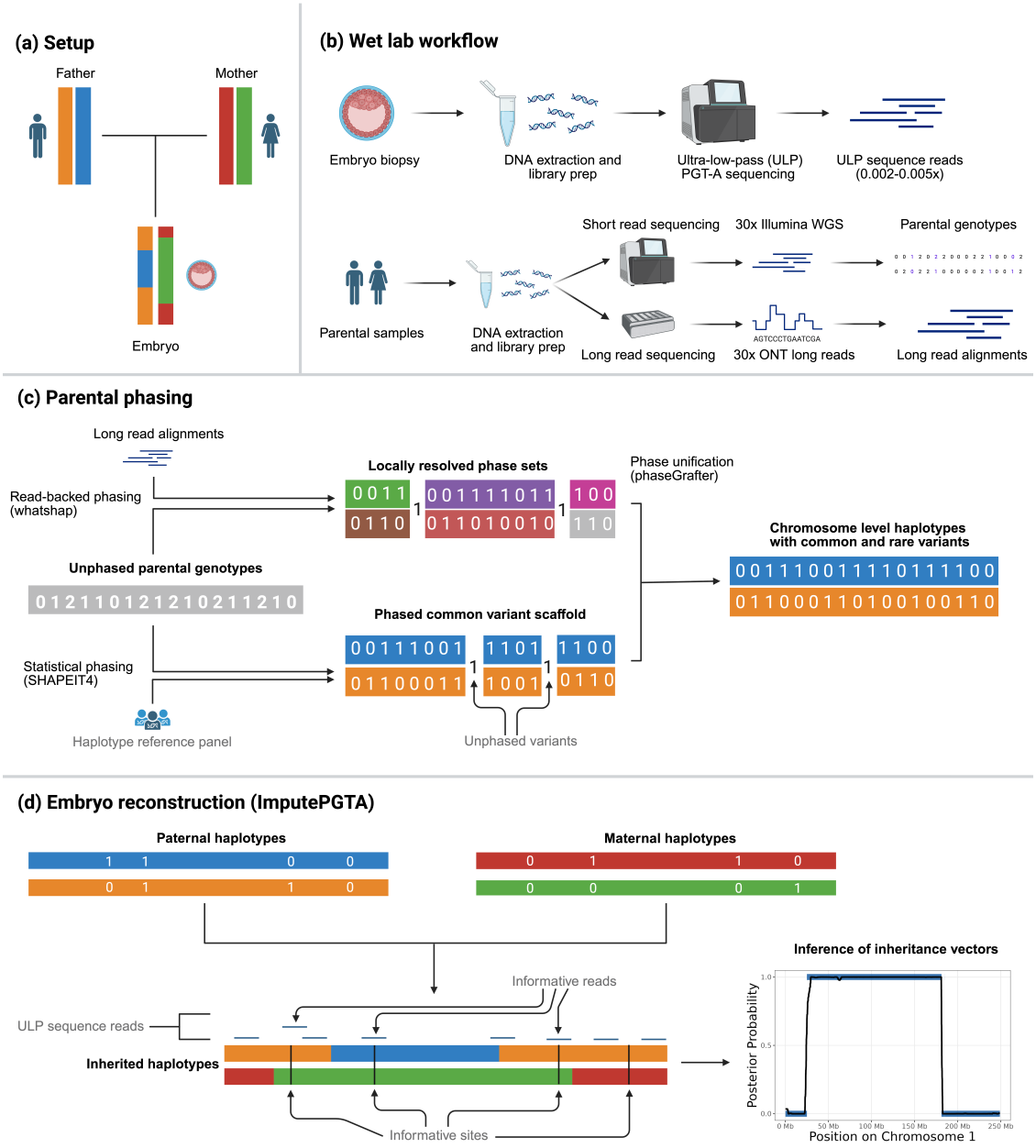

Figure 1: Graphical Abstract: Graphical representation of the workflow used to reconstruct the embryo genotype given sequence data on the parents and ultra-low-pass (ULP) PGT-A data on the embryo. **(a)** The experimental setup where we have two parents and an embryo. Each parental haplotype is given its own color; the embryo inherits a mosaic of the parental haplotypes. **(b)** The wet lab workflow for the embryo and the parents. A trophectoderm biopsy is taken from the embryo and sequenced at ultra low depth (down to 0.002x). Blood or saliva samples are collected from the parents and sequenced at high depth using both Illumina short reads and Oxford Nanopore long reads. **(c)** Parental haplotypes are estimated using phaseGrafter, our novel dynamic programming algorithm for unifying phase estimates from a common variant scaffold generated from statistical phasing (*e.g.*, SHAPEIT4) and independent phase sets generated by whatshap during read-backed phasing using Oxford Nanopore long reads. **(d)** ImputePGTA, our embryo genome reconstruction method, takes as inputs the inferred parental haplotypes and the embryo reads overlapping trio-segregating sites (*i.e.*, sites where at least one parent is heterozygous). Inheritance vectors are then inferred, producing posterior distributions over inheritance vectors and offspring genotypes. Posteriors from one of the real PGT-A cases are shown in black; in dark blue are the "true" inheritance vectors inferred from the high-coverage data from the born child.

It is not typically possible to produce accurate haplotypes at the chromosome-level from the most common types of genome-wide data: short-read sequencing and genotyping arrays. While sophisticated statistical phasing methods have been developed[42], the resulting haplotypes are imperfect and contain switch errors (consecutive heterozygotes that are incorrectly phased relative to each other), resulting in estimated haplotypes that *themselves* are mosaics of the true haplotypes. This makes inferring inheritance patterns from estimated parental haplotypes more challenging than from true parental haplotypes because changes in which *estimated* parental haplotype an offspring inherits from can be due to either recombination or switch error, with switch errors often far more numerous. To accurately impute embryo genotypes from coverage levels typical of PGT-A assays ($\sim$0.002-0.006x), these limitations become intractable with existing methods.

Here we describe our combined wet-lab and computational pipeline for embryo whole-genome reconstruction (**Figure 1**). To achieve clinical-grade accuracy for embryo genotypes at both common and rare variants, we sequence parents with both short and long reads. We then merge statistical phasing information from short reads with read-backed phasing from long reads using a novel method called phaseGrafter, which unifies the results into consistent parental haplotypes. When available, grandparental data are also incorporated, providing additional phase resolution at heterozygous sites through Mendelian inheritance rules. While one could use parental haplotypes estimated from short reads and statistical phasing — which we show can give sufficient accuracy for embryo selection based on common variant PGSs — this approach often fails to produce embryo genotypes with sufficient accuracy for clinical use, especially at rare variants.

Following parental phasing, we use ImputePGTA — a novel Hidden Markov Model (HMM) based algorithm — to infer posterior distributions over the inheritance vectors (*i.e.* which estimated parental haplotype the embryo inherited at each position), embryo genotypes, and PGSs. A key innovation is that ImputePGTA treats the parental switch-error rates (λ) as an explicit parameter inferred from the data, enabling accurate inference of embryo inheritance patterns from ULP PGT-A data with coverage as low as 0.002x. ImputePGTA also obtains highly accurate results when applied to embryo array data, though we focus on the sequencing setup due to its ubiquity in PGT-A testing.

We benchmark the model *in silico* using the Platinum Pedigree[43], a publicly available gold-standard dataset, by both simulating offspring with PGT-A data and downsampling reads from the real offspring data. We perform real-world validation of our approach using data from six embryos across four families that underwent PGT-A testing (2 with ULP data and 2 with array data), subsequent implantation, and live birth. We examine the performance of our method for estimation of embryo PGSs for 17 diseases by comparing to PGSs computed from high quality, post-birth genotypes, obtaining a mean-absolute difference of 0.157 SDs for embryos with ULP sequencing data and 0.0298 SDs for embryos with genotyping array data. Our results show that accurate genome reconstruction and polygenic scoring, along with posterior uncertainties, can be achieved from routine PGT-A data, removing an important barrier to wider adoption of PGT-P.

Results

ImputePGTA for embryo genotyping and polygenic scoring

ImputePGTA (**Figure 1D**) is a new hidden Markov model (HMM) for imputation of offspring genotypes from ULP sequence data or array data given estimated parental haplotypes. Here we briefly describe ImputePGTA at a high level, with more details in **Methods**. We focus on the ULP sequence data case as it is the more prevalent and challenging scenario.

Given estimated parental haplotypes, the offspring genotype is determined by the haplotypes it copies from (inherits) from each of the parents at each position, which can be represented by a binary inheritance vector of length 2, with possible values in $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$. The first element is the paternal haplotype the embryo copies from at that position (0 or 1), and the second element is the maternal haplotype it copies. The inheritance vector changes when there is a cross-over or a switch error in the parental haplotypes. ImputePGTA infers a posterior distribution over each embryo's inheritance vectors given the embryo's

sequencing reads and the estimated parental haplotypes. The key feature that makes ImputePGTA more effective than previous approaches[10,41] is that switch errors in the parental haplotypes are explicitly modeled, with the switch error rate for each parent inferred by maximum likelihood, enabling the inheritance vectors to switch at the appropriate rate for the phasing quality of each parent.

ImputePGTA also enables computationally efficient sampling from the full joint posterior distribution of the inheritance vectors, which can be used to compute posterior distributions of PGSs for embryos. This enables the propagation of uncertainty to predictions of disease risks (**Methods**) — including those incorporating family history — important for giving properly calibrated absolute and relative risk predictions in PGT-P. To illustrate PGS inference using ImputePGTA in a realistic setting, we used PGSs for 17 diseases from a recently published manuscript by Herasight, covering several cancers, metabolic and cardiovascular diseases, Alzheimer's disease, multiple sclerosis, inflammatory and autoimmune diseases, glaucoma, and osteoporosis[9] (**Methods**).

We tested the performance of ImputePGTA in simulations and on downsampled real data from the Platinum Pedigree[43]. Two key aspects we aimed to characterize were (1) the effect of switch errors in the parental haplotypes as described by parameter λ, which describes the rate of single switch errors per cM (**Methods**), and (2) the effect of embryo sequencing data coverage. We describe genotype imputation accuracy in terms of genome-wide dosage correlation and genotype concordance, and the accuracy of PGS estimated from the imputed genotypes in terms of mean absolute error from the true PGS.

Simulated offspring

We took gold-standard parental phased haplotypes for two parents in the Platinum Pedigree (NA12878 and NA12877) and simulated genetic recombination using sex-specific recombination rates drawn from the literature[44] to produce five simulated offspring (**Methods**).

To characterize the effect of λ and embryo coverage, we induced switch errors in both sets of parental haplotypes with λ values ranging from 0cM^{-1}to 1cM^{-1}(recombinations happen with rate 0.01cM^{-1}, so a switch-error rate of 1cM^{-1}implies switch-errors are 100x as frequent as recombinations), and simulated sequencing reads on the offspring at a range of coverages spanning those typical for a PGT-A assay, ranging from 0.002x to 1x. For intuition, ImputePGTA's performance is constrained by whether, and how quickly, a switch in the inheritance vector due to a switch error or recombination can be detected given the read data, both of which are influenced by λ and coverage.

ImputePGTA first infers λ. We observed that our estimates of λ were generally well-calibrated, with a slight upward bias at higher coverages (**Supplementary Figure S1**). It then imputes embryo genotypes from the marginal posterior distributions. We compared the imputed genotypes to the known genotypes of the simulated offspring at trio-segregating (TS) sites, defined as sites where at least one parent is heterozygous.

Genotype concordance and dosage correlation

Overall, we observed very high accuracies across the range of coverages when λ is low, with performance degrading as λ increases (**Figure 2A, Supplementary Table S1.1**). At 0.002x, concordance between the most likely imputed genotype and the true genotype at TS sites ranged from $97.83 \pm 0.11\%$ (mean $\pm$ standard error) at $\lambda = 0$ to $68.607 \pm 0.196\%$ at $\lambda = 1$. Similarly, at 0.002x, the genome-wide correlation between the posterior mean genotype (dosage) and true genotypes at the same coverage ranged from $0.9814 \pm 1.152 \times 10^{-3}$ at $\lambda = 0$ to $0.7412 \pm 1.086 \times 10^{-3}$ at $\lambda = 1$ (**Supplementary Table S1.1**).

Since the parent of origin of an allele in the offspring is ambiguous when both parents are heterozygous, we observed slightly worse accuracy at TS sites where both parents were heterozygous vs. where only one parent was heterozygous (**Supplementary Table S1.2**). An additional benefit of ImputePGTA is that posterior probabilities for each imputed variant are calculated, enabling one to optionally filter on the posterior at the variant level to obtain high-confidence subsets of calls (**Supplementary Table S1.3**).

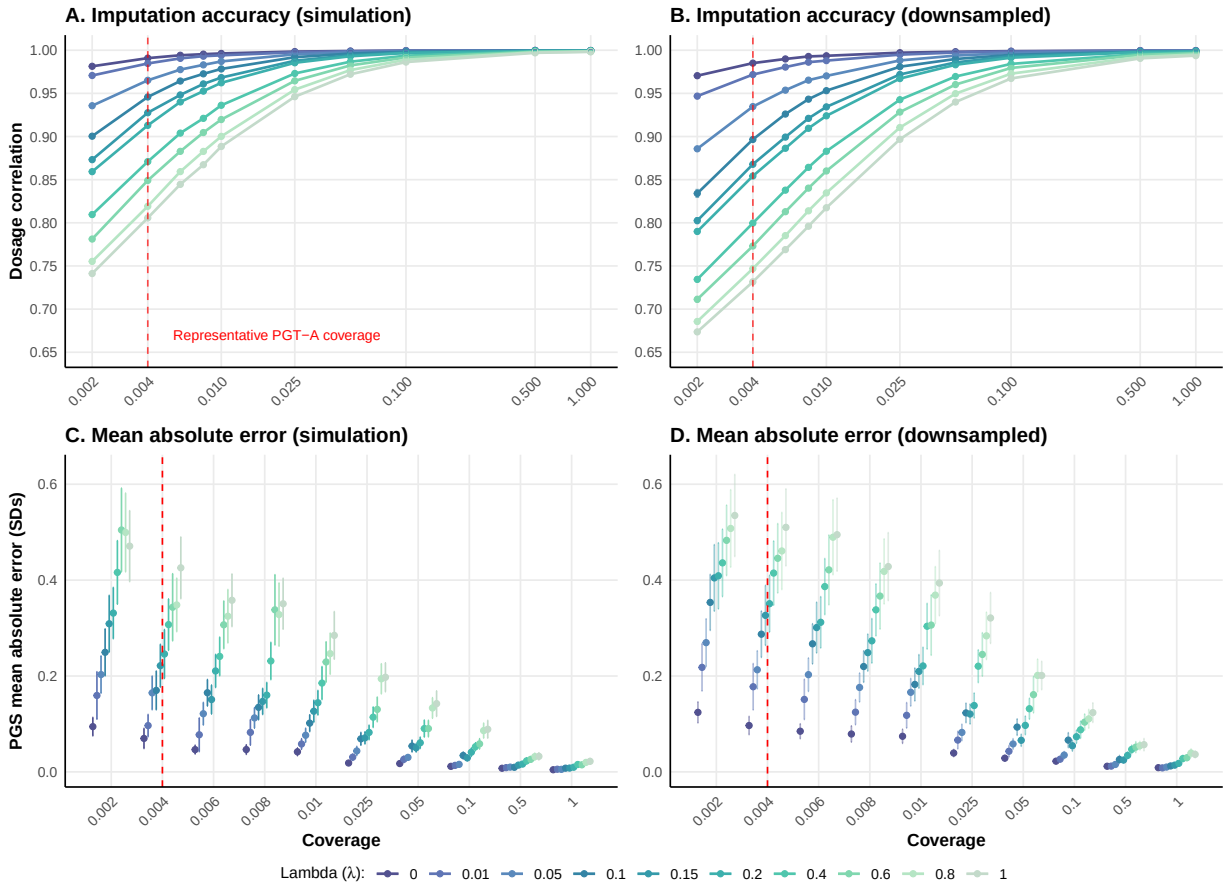

Figure 2: *In silico* experiments. (a) Imputation accuracy as measured by mean genome-wide correlation between imputed genotype dosages (dosage correlation) and true genotypes across a range of sequence depth (coverage) and a subset of switch-error rates (λ) in the parental haplotypes for five simulated offspring of the Platinum Pedigree parents. The red dotted line marks 0.004x, a representative PGT-A coverage. The variously colored points for a given coverage are jittered for visual clarity. **(b)** The same for the five real offspring downsampled to various coverages. Standard errors are small enough to not be visible in these plots. **(c)** Mean absolute error between the posterior mean PGS and true PGS in units of standard deviations of the PGS. 95% confidence intervals are shown as whiskers. Averages are across 5 simulated offspring from the platinum pedigree and 17 disease PGSs at each (λ, coverage) combination. **(d)** the same as for (c) but for the five real offspring downsampled to various coverages.

Polygenic score accuracy

For each sample, we calculated the posterior mean for 17 disease PGSs and obtained 500 posterior samples of the PGSs. **Figure 2C** shows the mean absolute error (MAE) (in units of PGS standard deviations) of the posterior mean PGS estimate relative to the true PGS. At a typical coverage for ULP PGT-A data (0.004x), the MAE$\pm$SE was 0.09659 ± 0.01174 when $\lambda = 0.01$, rising to 0.4257 ± 0.03274 when $\lambda = 1$.

We computed 95% equal-tailed credible intervals (CIs) of the PGS distributions, which we found to be reasonably well calibrated, with an average empirical coverage of 89.60% for the 95% CIs across the parameter space (**Supplementary Figure S2**). Notably, the CIs were better calibrated at the lower end of coverage, when uncertainty is greater (the average empirical coverage at read depth 0.01x or less was 92.6% vs 87.5% for read depth greater than 0.01x). When coverage is high and λ is very low, the CI coverage decreases somewhat, but with negligible practical consequence as the MAE is extremely low (**Figure 2C**).

These results indicate that parental phasing quality and embryo sequencing coverage are the two key parameters that control the expected imputation performance and quality of PGS estimates — if phasing

performance is expected to be low, then increasing embryo coverage can compensate for this limitation; if it is expected to be high, then achieving high embryo coverage is less important.

Validation on downsampled gold-standard data

The Platinum Pedigree dataset contains a variant "truth" set including eight offspring of the two parents, five of which have publicly available high-coverage short read data. We downsampled the short reads for these offspring to the same coverages as in our simulations, imputed them using ImputePGTA, drew posterior samples, and computed the same metrics as before (**Methods, Supplementary Note**).

Since we used the same parental haplotypes for these experiments as we did for the simulated offspring, we should obtain similar results in terms of imputation accuracy. However, some differences may arise due to differences from real sequencing data: the masking of reads in technically difficult regions (**Methods**), the fact that the simulated offspring were created using the exact genetic map used for inference, and any remaining genotyping or phase errors in the "truth" set.

We observed results that were qualitatively consistent with results from simulated offspring, with similarly high imputation accuracy and low PGS MAE, especially when λ was low (**Figure 2B, 2D**; **Supplementary Tables S2.1, S2.2, S2.3**). At 0.004x, genotype concordance at TS sites ranged from $96.45\% \pm 0.3052\%$ at $\lambda = 0$ to $61.30\% \pm 0.2771\%$ at $\lambda = 1$; the dosage correlation ranged from $0.9705 \pm 2.672 \times 10^{-3}$ at $\lambda = 0$ to $0.6736 \pm 1.51 \times 10^{-3}$ at $\lambda = 1$. (**Supplementary Table S2.1**). The MAE in the posterior mean PGS was 0.1776 ± 0.0246 when $\lambda = 0.01$, rising to 0.5100 ± 0.0409 when $\lambda = 1$. The 95% equal-tailed CIs were equally well-calibrated with an average empirical coverage of $89.9 \pm 0.454\%$, with the same qualitative variation with changing λ and embryo coverage.

Effect of imputation uncertainty on selection efficacy

The expected gain from embryo selection when the goal is to maximize a quantitative phenotype is the expected difference between the maximum PGS among a number of embryos compared to the average when selecting randomly, scaled by the correlation between PGS and phenotype[45,46]. When the goal is minimization of disease risk, the expected gain is more complicated to calculate because disease risk is a non-linear function of PGS and can include complex modeling of family history[46] (see **Methods** for details on producing posterior distributions of disease risk from posterior PGS distributions). However, the expected reduction in PGS value when selecting the embryo with the minimum polygenic risk compared to the average provides a tractable metric that can be interpreted as the reduction in disease liability (as in the standard liability threshold model) when multiplied by the square root of the liability-scale R^2.

The more uncertainty there is in the posterior inheritance vectors, the closer the posterior probability is to 0.5, the unconditional transmission probability based on Mendelian Laws. Thus, posterior uncertainty results in shrinkage of the posterior mean towards the expectation given the parents, which does not vary between embryos and is thus non-informative for embryo selection. We therefore sought to characterize the impact of posterior uncertainty in embryo PGS values on the expected change in true PGS when selecting the embryo with the minimum posterior mean PGS compared to selecting a random embryo.

We show that the attenuation in the expected gain for selection on a given trait using imputed PGSs is equal to the within-family correlation between the imputed and true PGS, which can be estimated from the posterior and within-family PGS variances (**Methods**): *i.e.*, the *attenuation factor* due to imputation is

$$A_{\text{imputed}} = \frac{\mathbb{E}[\text{gain}_{\text{imputed}}]}{\mathbb{E}[\text{gain}]} = \sqrt{1 - \frac{\mathbb{E}[\nu]}{\sigma_w^2}} = \text{Corr}(\text{PGS}, \widehat{\text{PGS}})$$

Where $\widehat{\text{PGS}}$ is the posterior mean PGS, σ_w^2 is the within-family PGS variance, and $\mathbb{E}[\nu]$ is the mean PGS posterior variance in the family. The equality follows because for a fixed number of embryos, the expected gain scales linearly with the within-family PGS standard deviation[46]: $\mathbb{E}[\text{gain}] \propto \sigma_w$ for the true PGS and $\mathbb{E}[\text{gain}_{\text{imputed}}] \propto \sqrt{\sigma_w^2 - \mathbb{E}[\nu]}$ for the imputed PGS, as $\sigma_w^2 - \mathbb{E}[\nu]$ is the within-family variance of the imputed mean PGS values (**Methods**) — under unbiased imputation, the ratio of these is $\text{Corr}(\text{PGS}, \widehat{\text{PGS}})$.

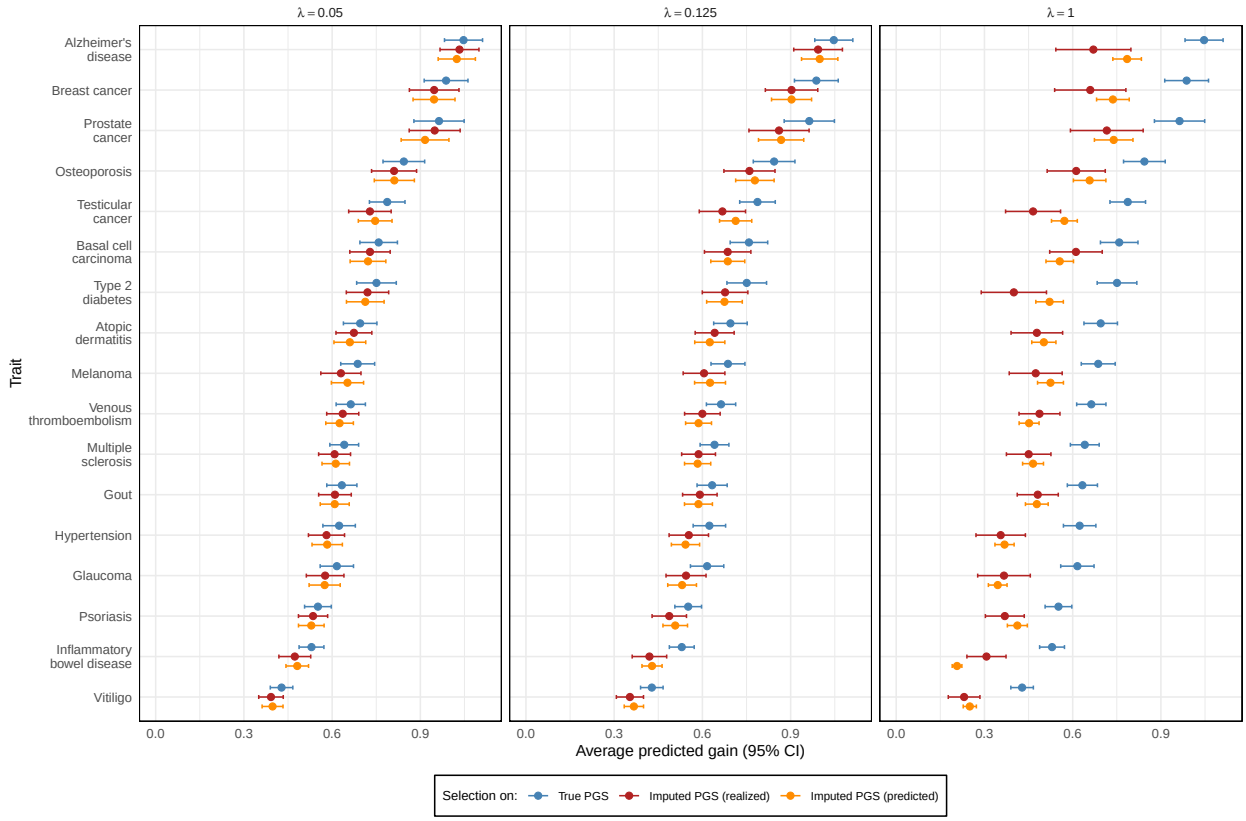

Figure 3: Attenuation of predicted gains due to selection on imputed PGS. The x-axis shows the average gain (95% CIs), defined as the absolute difference between the selected embryo's PGS and average PGS in units of PGS SDs from 100 instances of 5-embryo families with the same parents. The panels show results for three values of λ, the switch-error rate in parental haplotypes: the value based on empirical results for parents phased with short reads and long reads, ($\lambda = 0.05\text{cM}^{-1}$), short reads only ($\lambda = 0.125\text{cM}^{-1}$), and the highest value analyzed in the *in silico* experiments ($\lambda = 1\text{cM}^{-1}$). For each panel, three results are shown: the first two are the mean gain from two selection strategies: (a) selection based on true PGS values (blue), and (b) selection based on the imputed PGS values (red). The third is the theoretical gain predicted by the posterior variances (orange). The difference between the red and blue data points reflects the actual decrease in expected gains (based on true PGS values) resulting from imputation uncertainty.

While the average within-family PGS variance, $\mathbb{E}[\sigma_w^2]$, is half the PGS variance in the population under random mating[47], the within-family variance for a *particular* family can significantly deviate from this, and is affected by non-random mating.

The within-family variance for a particular family can be estimated by simulation of meiosis. For example, for the Platinum Pedigree, within-family variance calculated from 500 simulated offspring (below) ranged from 92.4% of the reference population variance for Alzheimer's disease to 13.6% for Vitiligo (**Supplementary Table S5.1**). The PGS for Alzheimer's disease has an outsized contribution from the APOE locus, implying that there can be substantially more within-family variation than the average for families with one or more parents that are heterozygous for a APOE risk allele (as is the case for the father here). The PGS for vitiligo, in contrast, shows a substantial correlation between parents[9] (0.085, S.E. 0.008) — indicative of assortative mating and/or population structure — which is expected to reduce within-family variance as a fraction of total variance. These results show that while theoretical results such as those in Karavani et al.[46] give useful estimates of the expected utility of embryo selection at the population level, they rely on assumptions such as random-mating and normality of PGSs that often do not hold in practice, with actual gains for specific diseases in individual families often substantially more or less than expected based on such results.

To validate our theoretical result, we simulated 100 batches of 5 offspring genomes from the Platinum Pedigree parents, each of which represents a possible realization of an IVF cycle (**Methods**). For each batch, we induced switch errors in the parents at a rate of $\lambda = 0.05\text{cM}^{-1}$, 0.125cM^{-1}, or 1cM^{-1}, simulated reads from the five offspring at 0.004x coverage, imputed them using ImputePGTA, and drew 100 posterior samples for each of the 17 PGSs. These parameters were chosen as they are representative of the empirical results from the real PGT-A cases with ULP data later described. The lower value of λ corresponds to results when both long and short reads are used for parental phasing, the second to when only short reads are used, and the third to the highest value of λ used in our *in silico* experiments.

For each trait, we then computed the average gain in true PGS across 100 batches when selecting based on the true PGS and on the imputed mean PGS. We also computed the theoretical expected gain as calculated from the posterior variances (**Figure 3**). The translation of these PGSs to absolute and relative risk reductions under a standard liability threshold model are shown in **Supplementary Figures S3 – S4** (see **Methods, Supplementary Tables S5.2-3**). We found strong agreement between the empirical gain and theoretical gain when selecting on imputed PGS values, indicating that the theoretical result is valid and our posterior variances are well-calibrated. Overall, the attenuation is minor at λ values of $0.05\text{cM}^{-1}$and 0.125cM^{-1}, with $A_{\text{imputed}} = 0.9494$ and 0.891 respectively, while at $\lambda = 1\text{cM}^{-1}$, A_{imputed} decreased to 0.661, highlighting the importance of accurate parental phasing. The empirical gain when selecting on the true PGS ranged from 0.428 for vitiligo to 1.047 SDs for Alzheimer's disease, reflecting differences in within-family variance for these PGSs (**Supplementary Table S5.1**).

Application to real PGT-A data in four families

Family ID	Children	Data type	Data source	Grandparental data	Genetic ancestry
FAM1	1	ULP sequencing (∼0.004x)	IGENOMIX	Both grandmothers	European
FAM2	1	ULP sequencing (∼0.004x)	IGENOMIX	None	European
FAM3	2	Genotyping array with ∼800k markers (Affymetrix Axiom UKB)	Genomic Prediction	None	European
FAM4	2	Genotyping array with ∼300k markers (Illumina HumanCytoSNP-12)	Natera	None	European

Table 1: Details of the families for which we obtained both PGT-A data from implanted embryos and high-coverage WGS data on the born child. The data modality of the PGT-A assay used during the IVF process (either sequencing or genotyping array) is noted in the "Data type" column — if ULP sequencing, the approximate coverage is noted in parentheses. The "data source" column indicates the provider that originally performed and delivered the results for the PGT-A assay. The "Grandparental data" columns indicate which grandparent(s) were sequenced in this study and used for parental phasing. The genetic ancestries of the parents are noted in the last column.

To validate the performance of our methods (**Figure 1**) in real world settings, we conducted a study of four families who underwent IVF where real PGT-A data was generated (**Methods**), followed by successful implantation, pregnancy, and live birth (**Table 1**). This enables us to compare embryo genotypes and PGSs from applying our pipeline to real PGT-A data to genotypes and PGSs computed from high-quality whole-genome sequence data from the born children. We analyzed two born children originally assayed using an ULP sequence-based PGT-A test and four originally assayed using genotyping arrays.

For the two embryos with sequence data, the reads were single-end IonTorrent short reads with an average read length of 112bp and an average coverage of 0.0046x (**Methods**). The embryos with array data were genotyped using two different chips (**Table 1**).

Each pair of parents was sequenced with both Illumina short reads and Oxford Nanopore long reads at a target coverage of 30x in order to genotype and phase all detectable SNPs and small indels in the parents,

including rare variants, using our pipeline (**Figure 1**) (**Methods**). The born children were sequenced at high depth (average 36.6x after deduplication) using Illumina short reads to determine their "true" genotype.

For FAM1, we also sequenced both grandmothers of the born child using short reads (**Table 1**). Grandparental data enables phase resolution at a subset of the heterozygotes in each parent by applying Mendelian inheritance rules, providing a scaffold atop which statistical phasing and long-read-backed phasing can be applied. For each embryo, we used ImputePGTA to impute genotypes, PGSs, and to draw 1000 posterior samples for each PGS.

Whole-genome parental phasing using phaseGrafter

		Heterozygotes phased		Inferred λ (switch errors/cM)	
Family	**Parental data type**	**Father**	**Mother**	**Father**	**Mother**
FAM1	Short reads, long reads, and both grandmothers	99.95%	99.95%	5.663×10^{-4}	1.494×10^{-4}
FAM1	Short reads & long reads	99.93%	99.92%	0.03999	0.02835
FAM2	Short reads & long reads	99.96%	99.95%	0.04090	0.03597
FAM3	Short reads & long reads	99.98%	99.97%	0.04560	0.03797
FAM4	Short reads & long reads	99.94%	99.94%	0.05724	0.02545
FAM1	Short reads	89.00%	89.13%	0.1039	0.05459
FAM2	Short reads	92.74%	91.54%	0.1332	0.08604
FAM3	Short reads	95.65%	95.59%	0.1795	0.1870
FAM4	Short reads	95.54%	94.97%	0.1809	0.07500

Table 2: Phasing results. Results of phasing using the described procedure (**Figure 1C**) on each parent in the set of families analyzed. We phased each pair of parents once using both long and short reads (the "base case") and once using only short reads (denoted in the Family column as "SR only"). Results for FAM1 are shown for when grandparental data was used to assist in phase estimation (GP) and when omitting it (no GPs). The proportion of successfully phased heterozygous genotype calls for each individual are shown for each case, as well as the inferred values of λ from ImputePGTA.

The accuracy of ImputePGTA depends significantly on the phasing quality of the parental haplotypes. While statistical phasing can be very accurate over short genetic distances, only variants also present in the reference panel can be phased. Sequencing the parents with long reads enables the phasing of these omitted rare variants[48], in addition to dramatically improving statistical phasing by using long-read-derived phase sets as prior information[49] (**Table 2**). To unify phase information from these orthogonal sources, we developed a phasing pipeline that is able to integrate information from statistical phasing and read-backed phasing from long reads into consistent haplotypes using a novel HMM-inspired dynamic programming algorithm called phaseGrafter (**Methods**). This pipeline is also able to integrate grandparental data. The results of this pipeline are accurate chromosome-level haplotypes for the parents that span both common and rare variants (**Methods**). To highlight the added value of long reads, we phased all four sets of parents once using both short and long read data (the "base case"), and once using only short read (SR) data (*i.e.*, using only statistical phasing).

Table 2 shows the percentage of parental heterozygotes phased and the inferred λ for each parent using each approach. In the base case, we phased an average of 99.95% of heterozygotes across all parents, while only 93.02% were phaseable using only SRs. The inclusion of long reads also resulted in an average 3.143-fold decrease in λ compared to using SRs only.

While grandparental data is not essential to achieve acceptably low levels of phasing error, it can significantly improve the results (**Methods**). For FAM1, where we had sequencing data on both grandmothers, using grandparental data decreased the inferred switch error rate to $\sim$0 (**Table 2, Figure 4**). For this particular European-ancestry family, we obtained nearly perfect phasing with only one grandparent from each side of the family — however, for families with ancestries not well represented in the reference panel, having both pairs of grandparents could be necessary to achieve near perfect phasing.

Estimation of embryo genotypes

Embryo	Parental data	Embryo data type	Genotype concordance	Dosage correlation	PGS mean absolute error	PGS posterior variance
FAM1-1	Short & long reads & both grandmothers	ULP	0.9882	0.9903	0.06407	0.01129
FAM1-1	Short & long reads	ULP	0.9478	0.9603	0.1800	0.04962
FAM1-1	Short reads	ULP	0.896	0.9234	0.3133	0.08374
FAM2-1	Short & long reads	ULP	0.9515	0.9614	0.1332	0.04016
FAM2-1	Short reads	ULP	0.9025	0.9183	0.2210	0.07399
FAM3-1	Short & long reads	array	0.998	0.9981	0.03695	1.896×10^{-4}
FAM3-1	Short reads	array	0.9955	0.9959	0.07323	0.01429
FAM3-2	Short & long reads	array	0.9987	0.9987	0.01821	2.859×10^{-4}
FAM3-2	Short reads	array	0.9974	0.9977	0.07074	0.01503
FAM4-1	Short & long reads	array	0.9975	0.9978	0.04028	6.785×10^{-4}
FAM4-1	Short reads	array	0.9957	0.9964	0.06601	0.002162
FAM4-2	Short & long reads	array	0.9976	0.9979	0.02366	0.001824
FAM4-2	Short reads	array	0.996	0.9966	0.06670	0.004637

Table 3: Genotype and PGS imputation performance Sample level metrics for imputation and PGS estimation accuracy for all samples across the families with real PGT-A data. Embryo data type was either ULP sequence data (∼0.004x) or genotyping array data (Table 1). Imputation accuracy is measured by the genotype concordance and dosage correlation between the imputed results and the true genotypes at trio-segregating (TS) sites. PGS accuracy is described by the mean absolute error — defined as the absolute difference from the PGS calculated from high-quality, post-birth genotypes in units of SDs of the PGS — across 17 disease PGSs. The mean posterior variance on the z-score scale across traits for each sample is also provided.

Using the same metrics as the *in silico* experiments, we give results comparing posterior embryo genotype estimates to high-coverage whole genome sequencing genotypes from the born child (**Table 3**). For embryos with ULP sequence data and the base case for parental phasing (short and long reads), the dosage correlation at all sites where at least one parent was variant from the reference genome was on average $0.9751 \pm 8.711 \times 10^{-5}$, and $0.9608 \pm 5.231 \times 10^{-4}$ when restricting to TS sites. When using only short reads on the parents, this was reduced to $0.9508 \pm 2.123 \times 10^{-3}$ and $0.9208 \pm 2.520 \times 10^{-3}$ for all variant sites and TS sites respectively. These numbers are consistent with results from the downsampled gold standard experiment (**Figure 2B**). For embryos with array data, we observed very high imputation accuracy across all embryos in the base case and when using only short reads — at TS sites, the average dosage correlation was $0.9982 \pm 2.021 \times 10^{-4}$ and $0.9966 \pm 3.854 \times 10^{-4}$, respectively. For FAM1, adding maternal grandparents increased the dosage correlation at TS sites from 0.9603 to 0.9903 (**Table 3**).

Figure 4 compares ImputePGTA marginal posterior inheritance vector probabilities from ULP PGT-A data to those inferred from the high-coverage post-birth data. (**Supplementary Figures S5 – S7** shows all autosomes.) The effect of phasing quality is readily apparent: higher values of λ correspond to a more frequent switching of inheritance vectors. The more frequently the inheritance vector switches, the harder it becomes to infer the switches from sparse embryo sequence data, resulting in larger regions where the posterior probabilities are intermediate, visible in some transitions in the top panel of **Figure 4**. The closer the probability is to 0.5, the larger the posterior variance and lower the efficacy of selection (**Methods and Figure 3**).

Inheritance of a rare pathogenic variant

As an illustration of the utility of long-read sequencing of parents, which enables imputation of whole-genome data on embryos — including rare variants — we identified that the father and child in FAM1 carry a rare pathogenic frameshift variant in *VPS13C* (NG_027782.1:g.145304dup, rs1315150327), resulting from a single thymine insertion at chr15:61920162 (GRCh38) (**Methods**). This variant leads to a frameshift at codon 2461, introducing a premature termination codon and is associated with autosomal recessive early-onset

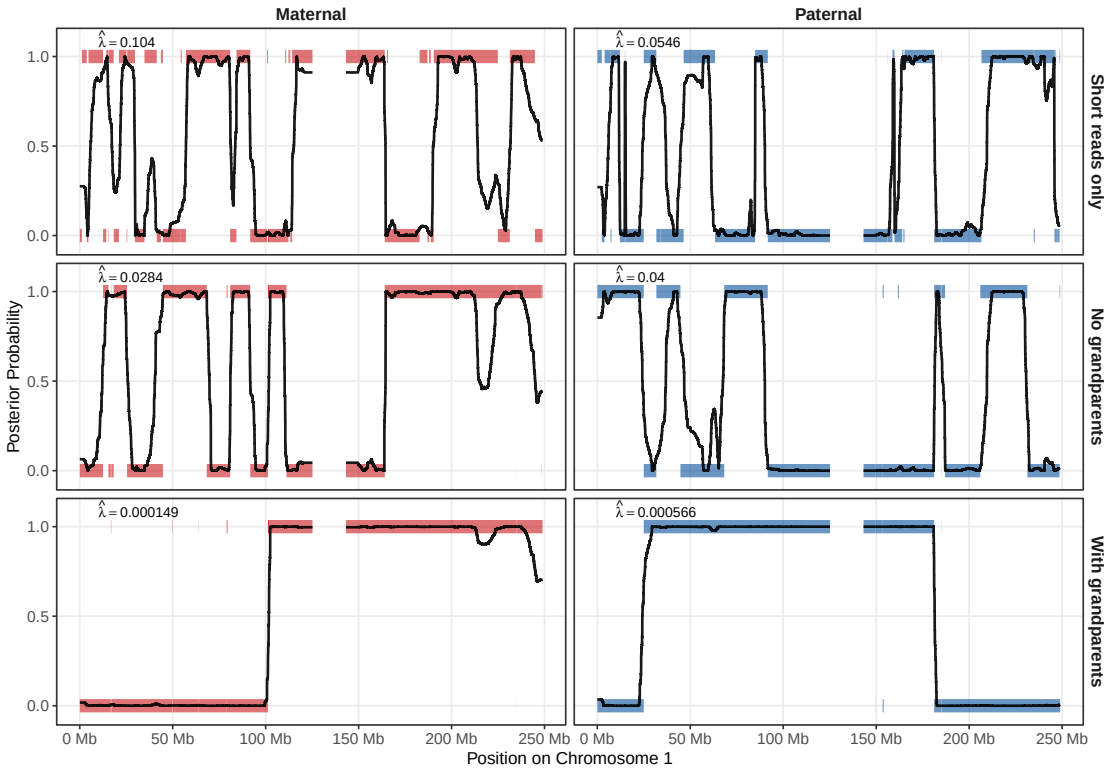

Figure 4: Marginal posterior inheritance vectors for FAM1 given different parental phasing strategies. For the offspring from FAM1 with ULP PGT-A data, marginal posterior probabilities referencing the *estimated* parental haplotypes on chromosome 1 are displayed in black. Probabilities at 0 and 1 indicate inheritance of one or the other of the parents' estimated haplotypes. A transition from 0 to 1 (or vice versa) indicates an inferred crossover event or switch error. The thick colored bars show the rounded posterior probabilities obtained from running ImputePGTA on the born child's genotypes obtained from high-coverage WGS and reflect a "best-case" estimate for the true inheritance vectors. Top to bottom: phasing parents using short reads only; short and long reads, but without grandparental data; short reads, long reads, and grandparental data. The inferred values of λ are shown in each pane. The effect of lower values of λ (more accurate phasing) can be seen as fewer switches in inheritance vectors need to be inferred accurately from sparse PGT-A data. Gaps on the x-axis reflect regions of at least 1Mb where no variants were genotyped in the parents.

Parkinson disease[50]. The variant has an allele frequency of 4.4×10^{-6} in gnomAD v4.1[51]. In the UKB 200k WGS phased release, only two copies of this allele are observed — at this allele count, statistical phasing performs extremely poorly[52]. Our inclusion of long reads on the parents, however, enables us to correctly impute this variant in the child with posterior probability > 0.999.

Polygenic scores

For 17 disease PGSs, we sampled 1,000 posterior inheritance vectors for each embryo to obtain posterior distributions (**Methods**, **Figures 5–6**). We computed the MAE as the mean absolute difference between the posterior mean PGS and born child's PGS (**Table 3**). Generally, the born child's PGS overlapped with regions of high posterior density, with some exceptions when posterior uncertainty was very high or very low — in low uncertainty cases, absolute error was also very low (*e.g.* psoriasis in the array cases). This likely reflects aspects of the model that differ from reality as well as limitations due to low read density when parental phasing quality is lower (*e.g.* when using short reads only). Genotyping errors in the parents and/or post-birth offspring could also contribute.

Posterior densities are generally non-Normal, sometimes due to posterior uncertainty at variants with large weights in the PGS. For example, the mother in FAM4 is a carrier for the APOE-ε4 allele — the long

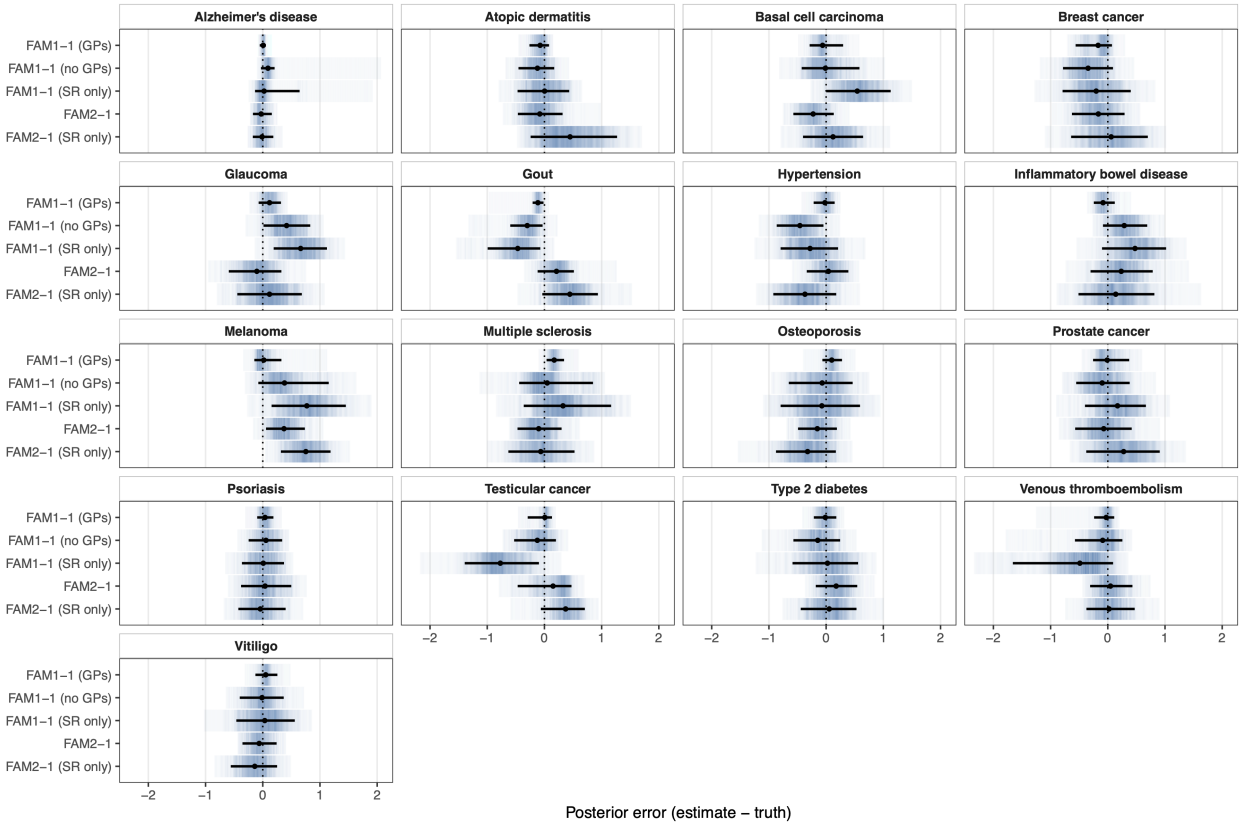

Figure 5: Polygenic score estimates for embryos with ultra-low-pass PGT-A data. Across 17 disease PGSs, we give the difference between the posterior PGS estimates and the PGS calculated from high-coverage data on the born child in units of PGS standard deviations. The black points give the posterior mean PGS, the whiskers denote the 95% equal-tailed credible interval, and the posterior densities are given by the intensity of shading. (SR only) indicates parental haplotypes were produced from short-reads only using statistical phasing. For FAM1, we give results for parental haplotypes produced using short and long reads, indicated with (no GPs), and for short and long reads with maternal grandparents (GPs), and for short reads only (SR only). For FAM2, we give results with short and long reads, and with short reads only (SR only).

right tail in the posterior distribution for Alzheimer's Disease for FAM4-2 (**Figure 6**) reflects the uncertainty in the inherited allele at that locus. This highlights the importance of sampling PGS posteriors for producing phenotype predictions that account properly for uncertainty in PGS values without making assumptions that posteriors are normally distributed. We discuss the implications of this in further detail in **Methods**.

For embryos with ULP sequence data, the MAE was 0.06407 SDs with parental short and long reads and both grandmothers (FAM1 only), 0.1566 SDs with short and long reads, and 0.2672 SDs with short reads only. In the base case, 88.2% of the estimated 95% CIs covered the true value of the PGS across the two embryos and traits.

As expected, the PGS estimates for embryos with genotyping array data are more accurate, achieving an MAE of 0.02977 SDs (0.06917 SDs when using only parental short reads). To contextualize this, we also calculated the PGSs from the embryo array data after applying reference-based imputation using a reference panel of 200k haplotypes (**Methods, Supplementary Table S4.6**). Array data plus reference based imputation is the most common approach used in academic research, and often involves using a much smaller reference panel. Error in PGS estimates as a result of reference-based imputation has been previously described[26,53], but remains underappreciated[54]. We found that the reference-imputed PGS MAE was 0.2637 and 0.1741 for the embryos genotyped using the Illumina HumanCytoSNP-12 (∼300k SNPs) and Axiom UKB Array (∼800k SNPs), respectively — substantially higher than the MAE from our method

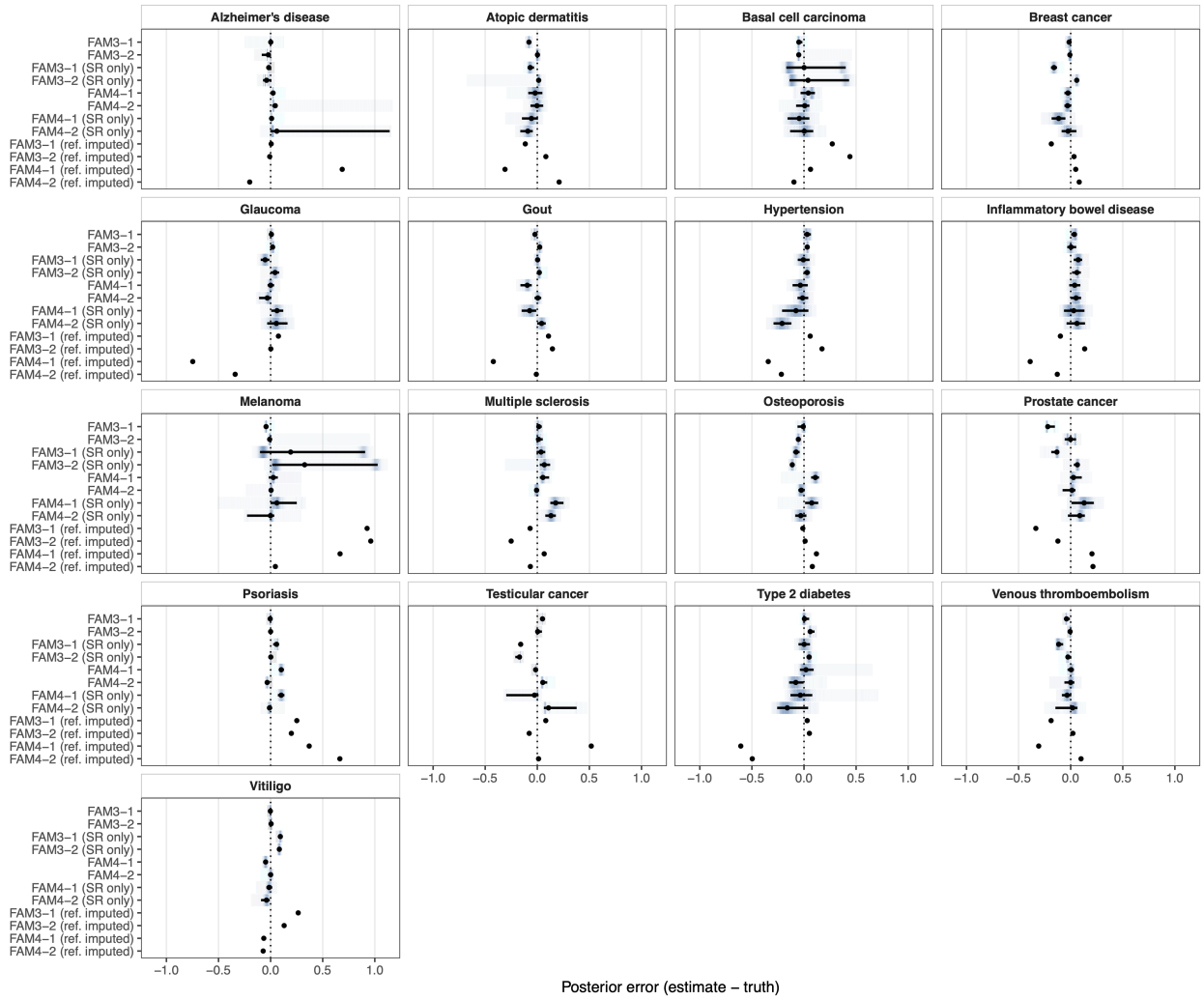

Figure 6: Polygenic score estimates for embryos with genotyping array data. Across 17 disease PGSs, we give posterior PGS estimates in units of standard deviations from the PGS calculated from high-coverage data on the born child. The black points denote the posterior mean PGS, the whiskers denote the 95% equal-tailed credible interval, and the posterior densities are given by the intensity of shading. (SR only) indicates parental haplotypes were produced from short-reads only using statistical phasing. The last four rows show the error of PGS estimates from the results of reference-based imputation.

applied to the embryo array data using parental short reads only (0.0692), and intermediate between the MAE from our method applied to ULP embryo data with parental short reads only (0.2672) and parental short and long reads (0.1566). This suggests that our PGS estimates achieve comparable accuracy to those derived from standard approaches applied to samples in academic human genetics research, with the added benefit of providing posterior uncertainties.

The largest estimated mean posterior variance across traits was for FAM1 with short reads only: assuming a within-family PGS variance of 0.5, this corresponds to an attenuation factor of 0.9124. For the embryo with the lowest mean posterior variance (FAM3-1), this factor is 0.9998, indicating effectively no loss in selection power. However, these estimates are probably somewhat over-optimistic due to imperfect calibration of posterior uncertainty in real data.

Discussion

Preimplantation genetic testing for aneuploidy (PGT-A) has become routine in contemporary IVF, with ultra-low-pass whole-genome (ULP) sequencing of trophectoderm biopsies deployed to identify embryos carrying chromosomal abnormalities. Although conceived as a single-purpose ploidy screen, PGT-A data has recently been repurposed for broader genomic discovery: for example, two recent studies[28,29] used ULP data from tens of thousands of embryos that underwent PGT-A to perform genome wide association studies on reproductive phenotypes. Yet, beyond these retrospective, population-level studies, PGT-A data have not been used for individual level embryo genotyping and any resulting implantation decisions during IVF. Here we show that the very same ULP data can be used to reconstruct the autosomal diploid genome of each embryo (excluding *de novo* variants), transforming a routine ploidy test into a comprehensive genotyping platform.

PGT-A data has been overlooked for applications other than its primary purpose because the assay's ultra-sparse nature ($\sim$0.002x-0.006x) is orders of magnitude lower than required for accurate genotyping from conventional methods. While low-pass sequencing followed by imputation has become increasingly common as a replacement for genotyping arrays in population studies[36,55], the target coverage of such assays (typically 0.1-1x) is two orders of magnitude higher than that generated in ULP PGT-A sequencing and typically does not produce accurate individual-level genotypes[53].

Our results show that when parental data is available, ULP PGT-A data can be used to obtain highly accurate embryo genotypes, with accuracy depending on the coverage of the embryo and the phasing quality of parental haplotypes. For trio-segregating sites, our approach achieves genotype dosage correlations of 0.9903 (when compared with high-quality, post-birth genotypes) when parents are phased using short and long reads in addition to grandparents, 0.9608 when grandparents are not available, and 0.9209 with only short reads (**Table 3**). Less commonly, PGT-A is performed using genotyping array data — which presents a less challenging problem for imputation — where our method achieved very high dosage correlation (>0.996) for all parental data types, higher than comparable existing methods[10]. We show that the attenuation in the efficacy of embryo selection due to posterior uncertainty is small when parental haplotypes are accurately estimated, with only a $\sim$5-10% reduction in expected gains, depending on whether both short and long reads are used for parental phasing.

Large ancestry-specific reference panels are becoming increasingly prevalent[56–58], yet are still not as readily available as predominantly European ancestry reference panels. For parents with ancestries well-represented in the reference panel, our method applied to ULP PGT-A data performs well enough for embryo polygenic scoring and selection when using statistically phased parental haplotypes derived from short reads only. For ancestries not well represented in the reference panel, the utility of parental long reads and grandparents increases. When even these data are not available, increased coverage from PGT-A sequencing data can compensate for lower quality parental phasing (**Figure 2**). Especially when the phasing quality of parental haplotypes is suboptimal, substantial posterior uncertainty can persist in embryo polygenic scores, highlighting the importance of propagating this uncertainty (**Figures 5 – 6**) into disease risk predictions (**Methods**).

Long reads additionally allow us to accurately impute rare variants that may not be present in a reference panel, which cannot be imputed using standard methods. We illustrate the power of this approach by detecting the presence of a rare pathogenic frameshift mutation in the father of one of the real PGT-A cases analyzed and accurately imputing it in the embryo even when no reads covered the site.

Our method bridges the gap between existing, widely-used laboratory workflows for PGT-A and the comprehensive genome-wide data required for PGT-P and detection of pathogenic rare variants, lowering the technological barrier to adoption of polygenic embryo screening and rare variant analyses in assisted reproduction.

Methods

The ImputePGTA model

ImputePGTA is a HMM used to impute offspring genotypes and PGSs from phased parental data and ULP sequence (or genotyping array) data on the offspring For each informative locus (where an informative locus is defined as those where at least one parent is heterozygous and at least one sequencing read maps for the offspring) $l = 0, 1, \ldots, L$, the hidden state of offspring $j = 1, \ldots, N$ is the inheritance vector $\mathbf{I}_j[l] = (\mathbf{I}_{pj}[l], \mathbf{I}_{mj}[l]) \in \{(0,0), (0,1), (1,0), (1,1)\}$, where $\mathbf{I}_{pj}[l]$ (resp. $\mathbf{I}_{mj}[l]$) indexes the paternal (resp. maternal) haplotype whose allele is transmitted. At each position l, genotype data (either sequence reads or genotype calls) on embryo $j = 1, \ldots, N$ is emitted: $\{\mathcal{G}_{lj}\}_{j=1}^{N}$. Based on Mendelian laws, the four possible values have equal prior probability, $1/4$. We make the Markovian assumption that $\mathbf{I}_j[l+1]$ is conditionally independent of $\mathbf{I}_j[l']$ given $\mathbf{I}_j[l]$ for $l' < l$.

The true parental haplotypes are denoted by $\mathbf{P}$ and $\mathbf{M}$, and the estimated haplotypes are denoted as $\hat{\mathbf{P}}, \hat{\mathbf{M}} \in [0,1]^{(L+1)\times 2}$. We assume the genotype calls are correct (or that the probabilities are accurate) but that the phasing may be incorrect. We account for switch errors by using a Baum–Welch algorithm to obtain maximum likelihood estimates of the paternal and maternal switch-error rates, $\lambda_p \mathrm{cM}^{-1}$ and $\lambda_m \mathrm{cM}^{-1}$. Given these parameters, we can obtain marginal posterior distributions over the inheritance vectors at each informative locus using the Forward-Backward algorithm[59].

Putting λ in context

In the literature, switch errors are typically quantified using the switch error *rate* (SER), which is the proportion of pairs of consecutive heterozygotes that are incorrectly phased. Switch errors can be further decomposed into type: a *single switch error* (SSE), in the words of Browning, is a switch error that is not "immediately preceded or followed by another switch error", whereas a *flip error* (FE) is when two consecutive switch errors occur[60]. A *phase error* is either a SSE or an FE. For ImputePGTA, it is primarily single switch errors that matter, as these result in long stretches of flipped alleles following a single switch; flip errors manifest merely as a few discordant sites where the flips occurred. Given the sparsity of the embryo data, these double switches are typically undetectable from the embryo data, so λ can be interpreted as measuring primarily the number of single switches that occur per cM.

Expected single switch-error rates in real data

Modern statistical phasing methods applied to the microarray genotypes in the White British cohort in the UK Biobank result in phase error rates equivalent to $\lambda \approx 0.03$ ([60]). However, this represents a best-case scenario, as this estimate derives from common variants on the UK Biobank (UKB) Axiom Array using a reference panel where the ancestry is well-represented. Phasing quality degrades with decreasing minor allele count in the reference panel and for poorly-represented genetic ancestries; as such, statistical phasing performance is highly dependent on the the reference panel used[52,61].

For individuals whose ancestries are not well-represented in readily available reference panels, long read and/or grandparental data can be collected in order to make up the difference[42].

Estimation of polygenic scores

In the context of PGT-P, the statistics of interest for an embryo are their polygenic scores, which are functions of the inheritance vectors. While the posterior mean PGS can be calculated from the marginal posterior decoding produced by the Forward-Backward algorithm, this does not give any information on posterior uncertainty in the PGS. We therefore sample inheritance vectors from the full joint posterior of $\mathbf{I}_j$ in $O(L)$ time, enabling efficient Monte Carlo estimates of any functional $f(\mathbf{I})$, such as a PGS.

Prediction of phenotypes and disease risks

Let Y_j be the phenotype of embryo j for a quantitative phenotype Y. Assuming a simple linear phenotype model based on a PGS, *i.e.*, $Y_j = r\mathrm{PGS}_j + \epsilon_j$, where ϵ_j is an independent error term, the predicted offspring phenotype is a simple linear function of the imputed mean offspring PGS:

$$\mathbb{E}\left[Y_j|\hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R}\right] = r\mathbb{E}\left[\mathrm{PGS}_j|\hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R}\right],$$

where $\mathbb{E}[\mathrm{PGS}_j|\hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R}]$ is the posterior mean PGS, which can be computed through posterior decoding without needing to sample from the joint posterior distribution over inheritance vectors.

A posterior distribution over the offspring phenotype can be computed by sampling inheritance vectors (above). The target is the conditional density of Y_j given the estimated parental haplotypes and offspring reads:

$$f\left(Y_j|\hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R}\right) = \sum_{\mathbf{I}_j} f\left(Y_j|\mathrm{PGS}_j\right) \mathbb{P}\left(\mathbf{I}_j|\hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R}\right),$$

where we have marginalized over the possible offspring inheritance vectors, PGS_j is a function of $\hat{\mathbf{P}}$, $\hat{\mathbf{M}}$, $\mathbf{I}_j$, and $f(Y_j|\mathrm{PGS}_j)$ is the density of the phenotype distribution conditional on the PGS, *e.g.*, $Y_j|\mathrm{PGS}_j \sim \mathcal{N}(r\mathrm{PGS}_j, 1 - r^2)$. As there are $2^{2(L+1)}$ possible inheritance vectors, it is not practical to evaluate this sum directly. We can instead use samples from the posterior distribution over the inheritance vectors, $\mathbb{P}(\mathbf{I}_j|\hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R})$, to produce a Monte-Carlo estimate of the posterior predictive distribution. Sample $\mathbf{I}_j^m$ from $\mathbb{P}(\mathbf{I}_j|\hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R})$ for $m = 1, \ldots, M$, then

$$f\left(Y_j|\hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R}\right) \approx \frac{1}{M} \sum_{m=1}^{M} f\left(Y_j|\mathrm{PGS}_j\left(\hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathbf{I}_j^m\right)\right)$$

where the PGS for the mth sample is computed using the mth sampled inheritance vector.

For disease prediction, the probability an offspring develops a disease is usually modelled as a non-linear function of the offspring PGS — for example, using a liability threshold model, potentially including family history and parental PGS — implying that the disease probability is not a linear function of the posterior mean PGS. Let D_j be the event that embryo j develops the disease. We assume that $\mathbb{P}(D_j|\hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R}, \mathbf{I}_j) = \mathbb{P}(D_j|\hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathrm{PGS}_j)$, *i.e.*, the disease probability model only depends on the inheritance vector and reads through the PGS. Then the posterior probability of disease is

$$\mathbb{P}\left(D_j|\hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R}\right) = \sum_{\mathbf{I}_j} \mathbb{P}(D_j|\hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathrm{PGS}_j)\mathbb{P}\left(\mathbf{I}_j|\hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R}\right),$$

which can be approximated using samples from the posterior distribution of the inheritance vector:

$$\mathbb{P}\left(D_j|\hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R}\right) \approx \frac{1}{M} \sum_{m=1}^{M} \mathbb{P}\left(D_j|\hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathrm{PGS}_j^m\right),$$

where PGS_j^m is the PGS computed from the mth inheritance vector sampled from the posterior.

Expected attenuation of selection efficacy when using imputed PGS

Here we derive an expression for the expected attenuation of selection efficacy as measured by the relative expected gain when selecting the lowest-risk embryo using posterior mean PGSs compared to when the true value of the PGS is known.

Assuming normality, Karavani et al.[46] show the expected gain from selecting the embryo with the max PGS value, $\max_j \mathrm{PGS}_j$ for $j = 1, \ldots, N$, is proportional to the within-family PGS standard deviation (although they do not make this explicit):

$$\mathbb{E}[\mathrm{gain}] \propto \sigma_w$$

where $\sigma_w^2 = \mathrm{Var}_j(\mathrm{PGS}_j)$ is the within-family variance of the PGS. This is because the expected max of an IID sample from a Normal scales in proportion to the SD.

In the context of ImputePGTA, the PGS of each embryo $j = 1, \ldots, N$ is estimated by the posterior mean: $\widehat{\mathrm{PGS}}_j = \mathbb{E}[\mathrm{PGS}_j | \hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R}]$, where $\hat{\mathbf{P}}$, $\hat{\mathbf{M}}$ are the estimated parental haplotypes, and $\mathcal{R}$ represents embryo reads. When using $\widehat{\mathrm{PGS}}_j$ to select an embryo, the gain is equal to the expected true PGS value of the embryo with the maximum imputed PGS value. Since the posterior mean PGS gives the expected value of true PGS for that embryo, the expected gain from selecting the embryo with the maximum imputed PGS value is simply $\max_j \hat{\mathrm{PGS}}_j$. Thus the expected gain is the expected max of the imputed PGS values, which, assuming normality, scales in proportion to the within-family SD of imputed PGS values: $\tilde{\sigma}_w = \mathrm{SD}_j(\hat{\mathrm{PGS}}_j)$.

From the law of total variance, we have that

$$\sigma_w^2 = \mathbb{E}\left[\mathrm{Var}(\mathrm{PGS} \mid \hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R})\right] + \mathrm{Var}\left(\mathbb{E}[\mathrm{PGS} \mid \hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R}]\right).$$

We can write the first term on the RHS as $\mathbb{E}[\nu]$, which denotes the average variance of the PGS posterior distribution for the embryos, which in practice we can obtain through sampling as described above, and the second term on the RHS is the within-family variance of $\hat{\mathrm{PGS}}_j = \mathbb{E}[\mathrm{PGS}_j | \hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R}]$. Then we can write $\tilde{\sigma}_w^2$ as:

$$\sigma_w^2 = \mathbb{E}[\nu] + \tilde{\sigma}_w^2 \implies \tilde{\sigma}_w^2 = \sigma_w^2 - \mathbb{E}[\nu].$$

Thus, when selecting on $\widehat{\mathrm{PGS}}_j$, the expected gain is

$$\mathbb{E}[\mathrm{gain}_{\mathrm{imputed}}] \propto \tilde{\sigma}_w = \sqrt{\sigma_w^2 - \mathbb{E}[v]},$$

The relative gain that is obtained by selecting on the posterior mean PGS compared to selecting on the true PGS, *i.e.* the *attenuation factor*, can then be written as the ratio of the gains and plugging in the results from above:

$$A_{\mathrm{imputed}} = \frac{\mathbb{E}[\mathrm{gain}_{\mathrm{imputed}}]}{\mathbb{E}[\mathrm{gain}]} = \frac{\tilde{\sigma}_w}{\sigma_w} = \frac{\sqrt{\sigma_w^2 - \mathbb{E}[\nu]}}{\sigma_w} = \sqrt{1 - \frac{\mathbb{E}[\nu]}{\sigma_w^2}}.$$

This ratio simplifies to $\sqrt{1 - 2\mathbb{E}[\nu]}$ for $\sigma_w^2 = 1/2$, the expected value of within-family PGS variance in a random-mating population with total PGS variance 1.

We now relate this to the within-family correlation between true PGS and imputed PGS:

$$\begin{aligned}
\mathrm{Cov}_j(\mathrm{PGS}_j, \widehat{\mathrm{PGS}}_j) &= \mathbb{E}[\widehat{\mathrm{PGS}}_j \mathrm{PGS}_j] - \mathbb{E}[\widehat{\mathrm{PGS}}_j]\mathbb{E}[\mathrm{PGS}_j] \\
&= \mathbb{E}[\mathbb{E}[\mathrm{PGS}_j \widehat{\mathrm{PGS}}_j \mid \hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R}]] - \mathbb{E}[\widehat{\mathrm{PGS}}_j]\mathbb{E}[\mathbb{E}[\mathrm{PGS}_j \mid \hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R}]] \\
&= \mathbb{E}[\widehat{\mathrm{PGS}}_j \mathbb{E}[\mathrm{PGS}_j \mid \hat{\mathbf{P}}, \hat{\mathbf{M}}, \mathcal{R}]] - \mathbb{E}[\widehat{\mathrm{PGS}}_j]^2 \\
&= \mathbb{E}[\widehat{\mathrm{PGS}}_j^2] - \mathbb{E}[\widehat{\mathrm{PGS}}_j]^2 \\
&= \mathrm{Var}(\widehat{\mathrm{PGS}}_j) = \tilde{\sigma}_w^2 \\
\implies \mathrm{Corr}(\mathrm{PGS}_j, \widehat{\mathrm{PGS}}_j) &= \frac{\tilde{\sigma}_w^2}{\tilde{\sigma}_w \sigma_w} = \frac{\tilde{\sigma}_w}{\sigma_w}.
\end{aligned}$$

Thus, the attenuation is equal to the within-family correlation between the true and estimated PGS.

In silico experiments

A description of the Platinum Pedigree data used for the simulations and in-silico downsampled offspring data is available in **Supplementary Note**. Briefly, we used a family comprising parents (NA12878 and NA12877, both of European ancestry) and five children (NA12879, NA12881, NA12882, NA12885, and NA12886) for which raw high-coverage short read sequence data and a highly validated phased truth-set comprising SNPs and small indels is publicly available.

Simulating offspring and sequencing reads

Simulated offspring genotypes were generated from the phased VCFs for the parents, NA12878 and NA12877, using a genetic recombination model. For each offspring, crossover events were simulated following a Poisson process with rates determined by sex-specific genetic maps[44], and offspring genotypes were determined given the transmission of haplotypes implied by the crossover events after first selecting an initial haplotype to transmit. This process was repeated to simulate 5 independent offspring.

Sequencing reads were then simulated from the offspring genotypes using a Poisson sampling process. For each variant site in each simulated offspring, the number of reads was drawn from a Poisson distribution with a rate equal to the target coverage level (ranging from 0.002x to 1x). When reads were present at a site, each read randomly sampled one of the two parental haplotypes with equal probability. Read errors were introduced by randomly changing the true base to one of the other three nucleotides with probability 0.001 (corresponding to Q30 base quality). All reads were assigned fixed Phred-scaled base quality scores of 30, representing high-quality reads.

Simulating switch errors in parental haplotypes

Phasing switch errors were introduced into the parental haplotypes using a Poisson process model. For each parent, haplotype data was processed chromosome-wise using sex-specific genetic map positions from the literature[44]. Switch error events were sampled at each variant position by drawing from a Poisson distribution with a rate equal to the genetic distance (in cM) between consecutive variants multiplied by λ, the specified error rate parameter. Here, λ can be interpreted as the expected number of switch errors per cM. Switch events occurred when the Poisson draw was odd. Once a switch error was introduced at a given position, all subsequent variants on that chromosome had their haplotype phases inverted until the next switch event occurred, thus modeling the propagation of phasing errors that persist along chromosomal segments. We introduced switch errors at the following rates (λ): 0, 0.01, 0.05, 0.1, 0.15, 0.2, 0.4, 0.6, 0.8, 1.0.

Generating downsampled short read data for real offspring

To characterize the empirical performance of our imputation method from ultra low pass sequencing data, we downsampled the publicly available high coverage BAMs of the offspring to the following coverages: 0.002x, 0.004x, 0.006x, 0.008x, 0.01x, 0.025x, 0.05x, 0.1x, 0.5x, 1x using `sambamba view -s`[62].

Imputation and evaluation of simulated reads and downsampled offspring data

For both the simulated and downsampled experiments, we ran ImputePGTA with either simulated or real reads as the data input for the offspring. As previously described, first, parent-specific switch error rates (λ_p, λ_m) were estimated from the parental haplotypes, the offspring reads, and a genetic map from the literature[44] using a Baum-Welch algorithm; these estimated values were then used in the inference step, where inheritance vectors were estimated and offspring genotypes are imputed.

Imputation accuracy was evaluated using the dosage correlation coefficient as well as genotype concordance. The dosage correlation coefficient is computed as the correlation between the dosage in the imputed VCF and the numeric genotype (*i.e.*, $0, 1, 2$) in the truth VCF. We also evaluated accuracy at "trio-segregating" (TS), defined as sites where at least one parent is heterozygous, the only sites that vary between siblings other than by *de novo* mutation.

Selection efficacy experiment

To characterize the attenuation of expected predicted gains when selecting on the basis of imputed PGS values versus selecting based on true PGSs, we simulated 500 offspring from the Platinum Pedigree parents used in the experiments described above and calculated PGSs for them. We simulated reads at a coverage of 0.004x for each offspring. We then grouped these simulated offspring into 100 batches of 5 embryos each. For each batch, we then simulated estimated parental haplotypes by inducing switch errors at $\lambda = 0.05$, 0.125, and 1.

The results of this procedure for each value of λ represents 100 realizations of embryo batches and estimated parental haplotypes under parameter combinations reflecting the real PGTA cases with ULP sequence data. We used ImputePGTA to impute the offspring genotypes from the simulated read data, and calculated the posterior mean PGS for the 17 traits for each offspring directly from the dosages resulting from posterior decoding. We also drew 100 posterior samples for each imputed offspring and calculated the PGS posterior variance.

We then simulated the process of embryo selection using two strategies: (a) selecting based on the lowest true PGS, and (b) selecting based on the lowest imputed PGS. For each batch of five embryos, we calculated the predicted gain as the difference between the true PGS of the selected embryo and the mean of the true PGSs within the five embryos in the batch. We also calculated the mean posterior variance per trait across all 500 embryos for each λ value as well as the true within-family PGS variance to obtain the predicted attenuation factor.

To translate these PGSs onto the risk scale, we used the disease prevalences and liability-scale R_l^2 values from Moore et al. 2025[9] (**Supplementary Tables S5.2-3**). We used the standard liability threshold model for disease where disease liability L is modeled as $L = G + E$ where G represents the genetic component distributed as $G \sim N(0, R_l^2)$, E represents the environmental component distributed as $E \sim N(0, 1 - R_l^2)$, and G and E are assumed to be uncorrelated. For a binary disease trait with population prevalence K, the liability threshold T is defined as $T = \Phi^{-1}(1 - K)$ where Φ^{-1} is the inverse standard normal cumulative distribution function. The probability of disease (*i.e.*, the absolute risk, or AR) for an individual with standardized PGS value g is then given by

$$P(\text{disease} \mid \text{PGS} = g) = \Phi\left(\frac{gR_l - T}{\sqrt{1 - R_l^2}}\right).$$

The absolute risk reduction for a given embryo in a batch is then the difference between its absolute risk AR and the mean absolute risk across embryos within the batch $\overline{\mathrm{AR}}$. The relative risk reduction (RRR) for a given embryo in a batch of embryos is then $\mathrm{RRR} = (\overline{\mathrm{AR}} - \mathrm{AR})/\overline{\mathrm{AR}}$. The absolute and relative risk reductions corresponding to the PGS-scale expected gains in **Figure 3** are shown in **Supplementary Figures S3– S4**.

Parental phasing in real data

Statistical phasing of variant calls is generally quite accurate; however, only variants that are present in a reference panel can be phased. An orthogonal approach is read-backed phasing, which uses direct read evidence (such as from long reads) to establish phase, allowing phase estimation at all heterozygotes with sufficient spanning reads[48]. However, even using modern long read technologies, this results in only multiple disjoint *phase sets* (*i.e.*, blocks of phased heterozygotes) that do not span full chromosomes; for example, for one of the real-world cases (FAM1), long read prephasing using high-coverage ONT long reads with `whatshap`[48] on chromosome 1 for the father resulted in phase resolution at >99.9% of all heterozygotes, but contained within 417 disjoint phase sets, meaning that while phase was locally resolvable for these variants, there was insufficient information to link these individual phased blocks together to achieve chromosome-level haplotypes.

Thus, in order to generate chromosome-level haplotypes for the parents that span both statistically phased common variants and long-read-phased rare variants, we used a combination of methods (**Figure 1C**). At a high level, the process is as follows:

1. Mendelian scaffolding (*i.e.*, inferring phase using Mendelian laws) using grandparental data (when available) at heterozygotes resolvable by Mendelian inheritance laws.

2. Statistical phasing using `SHAPEIT4`[63] with the scaffold (when grandparental data was available) option and the phase set option to phase common variants.

3. Read-backed phasing using ONT long reads with `whatshap` to obtain independent phase sets containing both common and rare variants with locally resolved phase.

4. A novel HMM-inspired dynamic programming algorithm we call phaseGrafter to combine the results from (2) and (3) into single unified haplotype estimates spanning both common and rare variants by "grafting" phase sets generated using long reads atop a scaffold of common variants.

For all parents, we used long read data for phasing purposes only, and relied on the short read variant calls for the genotypes. For each sample, we pre-phased and generated phase sets by using Oxford Nanopore long reads `whatshap` v2.6[48] to phase the short read VCFs, resulting in locally resolved phase sets of heterozygotes.

Statistical phasing

We used `SHAPEIT4` v4.2.2 to statistically phase the common variants present in each parent, incorporating Mendelian scaffold and long read phase sets as priors when available.

We used a reference panel of synthetic haplotypes generated using `RESHAPE`[64] that is based on the phased reference panel previously generated[52] for the UK Biobank 200k WGS release. We subset to variants with MAF $\geq 0.1\%$ prior to creating the synthetic haplotypes in order to control computational cost. The result of this is a *common variant scaffold* containing chromosome-level estimated haplotypes, but limited only to common variants overlapping the reference panel used.

Rare variant phasing and phaseGrafter

The output of `SHAPEIT4` is a set of statistically phased chromosome level haplotypes for each parent. However, statistical phasing can only resolve variants which are also present in the reference panel used — across the parents of the four families, $\sim 7\%$ of heterozygotes remain unresolved in the parents after statistical phasing due to their absence in the reference panel. On the other hand, prephasing has no such limitation, and heterozygotes can be phased where reads span at least two heterozygous variants. The limitation here is instead the disjoint nature of the phase sets, where two independently resolved sets of heterozygotes have internally resolved phase, but the phase relation between them is unclear.

We integrate these two sources of phase information using a novel HMM-inspired dynamic programming approach called phaseGrafter which "stitches" phase sets generated using `whatshap` onto the phased common variant scaffold output from statistical phasing (e.g. `SHAPEIT4`), resulting in unified haplotypes that retain the chromosome-scale consistency of the statistical scaffold at common variants while extending the phasing to rare heterozygotes carried by the individual.

The basic idea behind phaseGrafter is that most variants phased in one dataset will also be phased in in the other dataset. For each local haplotype in the read-derived phase sets, one can then compare the relative orientation of the phase estimates in the overlap and deduce the most consistent internal orientation of the heterozygotes within the phase set with respect to the scaffold, after which each heterozygote in the phase set which is *not* in the intersection can be grafted onto the scaffold accordingly.

Using this method, we are able to resolve the phase at $> 99.9\%$ of heterozygous sites in the parents in the base case (phasing using short and long read data) after applying `phaseGrafter`, thus enabling whole-genome reconstruction of both common and rare inherited variation within offspring. **Supplementary Table S4.3** contains the variant counts in the parents in addition to the number and percentage that were phased in each scenario.

Polygenic scores

We calculated PGSs for 17 disease traits described in a recent manuscript from Herasight[9]: Alzheimer's disease, atopic dermatitis, basal cell carcinoma, breast cancer, glaucoma, gout, hypertension, inflammatory bowel disease, melanoma, multiple sclerosis, osteoporosis, prostate cancer, psoriasis, testicular cancer, Type 2 diabetes, venous thromboembolism, and vitiligo. These PGS have nonzero weights at ~ 7.3M variants across the autosomes.

To calculate ancestry-adjusted PGS for a given individual, we first computed the PGS for each sample in the HGDP + 1KG dataset[65] using the `plink2 --score` function[66].

Using these individuals embedded in a common genetic PC space, we then fit per-trait linear models for the PGS conditional mean and residual variance on the first four PCs, then mapped each posterior sample from the target individual into this space and z-standardized the raw scores using the predicted mean and variance.

The results of this procedure are ancestry-adjusted standardized PGS for each posterior sample.

Real PGT-A Cases

Recruitment of participants and sample collection

We recruited four families to validate ImputePGTA. The parent participants recruited were informed about the study through their IVF clinic and enrolled under two approved IRB protocols (Solutions IRB protocol ID 0448 and Pearl IRB protocol ID 2025-0250). Informed consent in each case was obtained from both participating parents, and parental consent was obtained on behalf of the born child. For the family where grandparental data was obtained, consent was also obtained from the relevant born child's grandparents. **Supplementary Table S4.7** shows the details of the families. FAM3 and FAM4 both had two children for which we obtained both the PGT-A and born child high coverage read data. The parents underwent genomic sequencing using both Oxford Nanopore Technologies (ONT) long-read sequencing and Illumina short-read sequencing. The grandparents (if applicable) as well as the born child were sequenced using Illumina short-read sequencing.

To generate long-read data for the parents, whole blood was self-collected at home using the Tasso+ device with an EDTA microtainer tube. Genomic DNA from blood was extracted with the Qiagen Puregene Blood Kit following the manufacturer's instructions. Libraries were prepared with the Oxford Nanopore Technologies (ONT) Ligation Sequencing Kit and run on a PromethION P2.

To generate short-read data for the born children, grandparents, and parents, self-collection was performed to obtain either buccal epithelial cells, using the iSWAB-DNA-250 kit, or saliva, collected with GeneFiX or Oragene DX kits. For short-read sequencing, libraries were prepared with Illumina DNA Prep and sequenced on an Illumina NovaSeq X+ at Eurofins Clinical Enterprise (Louisville, KY).

The realized short read WGS coverages for aligned and deduplicated reads for the father, mother, and grandparents if applicable are tabulated in **Supplementary Table S4.7**; the average coverage for short reads achieved for the parents and grandparents was 49.9x and the average coverage of the born children was 36.6x. For the parental long-read sequencing, the average coverage was 34.0x for the long reads with an average N50 of ~ 26kb. Per-sample coverages can be found in **Supplementary Table S4.7**.

With assistance from the parents, we obtained the original PGT-A data used for screening during the IVF from the original data providers as described below.

Data methods for WGS samples

Alignment and variant calling

For the short reads, we aligned them to `GRCh38` (obtained from ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/seqs_for_alignment_pipelines.ucsc_ids/GCA_000001405.15_GRCh38_no_alt_analysis_set.fna.gz) using `bwa-mem2` v2.3. Long-read sequencing data for the parents were aligned to `GRCh38` using the `epi2me wf_human_variation` workflow v2.6.

The short read alignments were processed using a multi-step variant calling pipeline. For FAM1 and FAM2, we processed the mother, father, and born child using `DeepTrio` v1.9.0 for pedigree-informed variant calling. Joint genotyping of the trio gVCFs was then performed using GLnexus to produce a consolidated trio VCF. All variants were normalized using `bcftools norm` to split multi-allelic sites, filtered to retain only autosomal sites with non-missing genotype calls across all trio members. A further filtering step was done to retain only sites where both parents had confident genotype calls (GQ$\geq$30). Individual sample VCFs were extracted from the joint-called trio VCF. For the grandparents in FAM1, we generated single-sample VCFs using `DeepVariant` v1.9.0.

For FAM3 and FAM4, we generated single-sample VCFs for the parents and the two born children using `DeepVariant` v1.9.0. The same filters as for FAM1 and FAM2 were applied. The number of called heterozygous variants passing filters for each parent are tabulated in **Supplementary Table S4.3.**

Embryo PGT-A data

PGT-A data was obtained from the original data provider with assistance from the parents for the embryos corresponding with the born child(ren). The data processing steps for embryo ULP sequence and array data are described below.

Embryo ULP sequence data

For FAM1 and FAM2, the embryo data obtained was ULP sequence data in the form of a BAM file containing single-end reads aligned to `hg19`. For these samples, the PGT-A assay was performed using the Ion ReproSeq PGS Kit (Next Generation Sequencing) for 24 chromosome aneuploidy screening (Thermo Fisher Scientific, USA). The kit/assay was performed on the Ion Chef and Ion S5 System instruments (Thermo Fisher Scientific, Inc, MA, USA). Data analysis was performed with Ion Reporter software (IRv5.16) (Thermo Fisher Scientific, USA).

We reverted the BAMs to FASTQs using samtools and re-mapped the reads using `bwa` v0.7.18 before feeding the reads into ImputePGTA.

After alignment, we mark duplicates using `samtools markdup` and calculate the pileup using `samtools mpileup`, restricted to the set of sites at which at least one parent was variant and reads where the mapping quality (MQ) is ≥ 10. We only take into account single bases and thus only SNV variant sites are modeled as having observed data (reads at indels in the parents are ignored). For multi-allelic sites in the parents as well as sites which are covered by more than one read in the embryo, we model these as additional states in the HMM separated by a nonzero but negligible genetic distance (in principle, multiallelic variants should be explicitly modeled, but we find that this heuristic works well in practice). These reads after quality score recalibration (described below) are used as the input data for the embryos in ImputePGTA as described in the main text.

Embryo genotyping array data processing

For FAM3's children, the assay used for PGT-A was the Axiom UK Biobank Array[67] with ~ 800k SNPs. The data was obtained from the original provider, Genomic Prediction, in the form of VCFs with variant calls on hg19. We lifted these over to GRCh38 using Picard[68] and these VCFs were used for input to ImputePGTA. The VCFs obtained from Genomic Prediction did not contain variant-level quality scores.

For FAM4's children, the assay used for PGT-A was the Illumina HumanCytoSNP-12 BeadChip[69] with ~ 300k SNPs. The data was obtained from the original provider, Natera, in the form of IDAT files. We first converted these to GTC format using the Illumina Array Analysis Platform Genotyping Command Line Interface[70] followed by conversion to VCF on `GRCh38` using the `gtc2vcf` plugin for `bcftools`[71].

For each family, the genotype calls in the embryos at variants overlapping with the parental genotype calls are then used as the input to ImputePGTA after quality score recalibration as described below is applied.

We also imputed these arrays to the UKB reference panel used for parental phasing. Briefly, we pre-phased the original VCFs using `SHAPEIT5`[52], followed by imputation using `IMPUTE5`[72]. We filtered to sites called in the truth VCFs and calculated PGSs using the approach previously described.

Base/genotype quality score recalibration

We recalibrated call quality using an embryo-stratified, quantile-based empirical approach. For the sequencing data, Phred-scaled base qualities Q were converted to error probabilities as $p = 10^{-Q/10}$. For array data, we used the genotype qualities (the `FORMAT/IGC` field for IDAT VCF outputs); scores were treated as

accuracies and converted as $p = 1 - \mathrm{qual}$. For FAM3's embryo array data, the VCFs we received did not contain genotype qualities; as such, we assumed a conservative $\mathrm{qual} = 0.95$ for all sites.

Within each embryo, sites were partitioned into five quantiles of the raw quality score; a single stratum was used if the score had zero variance (*e.g.*, if quality scores are not available, one could choose to assume a fixed quality score at all sites). At Mendelian-informative sites where both parents were homozygous for the alternate allele, we defined an error as follows. For sequencing, an error was defined as a read base that was not equal to the alt allele at that site. For arrays, an error was defined as a genotype call on the embryo that was not homozygous for the alternate allele. For each embryo-quantile stratum we estimated the observed error rate using Laplace smoothing and compared it with the mean predicted error (based on the uncalibrated scores) in the same stratum.

We adjusted each call's error probability on the logit scale by adding a shrinkage-weighted offset equal to the difference between observed and predicted logit error rates:

$$\mathrm{logit}(p') = \mathrm{logit}(p) + w\,(\mathrm{logit}(\hat{e}) - \mathrm{logit}(\bar{p}))$$

where $w = n/(n + \tau)$. Here, p is the call's predicted error probability, $\hat{e}$ is the stratum's observed error rate, $\bar{p}$ is the stratum's mean predicted error, $\tau = 30$ is a hyperparameter that controls shrinkage toward no adjustment. and n is the number of variants in the stratum. When per-embryo strata were sparse or missing, we used quantile-level estimates aggregated across embryos in the same family. Adjusted probabilities were transformed back to the original scales: $Q' = -10 * \log(p')$ for Phred base qualities and $\mathrm{qual}' = 1 - p'$ for genotype quality.

Evaluation of genotype imputation accuracy

There are regions of the genome known to be technically difficult due to segmental duplications and low-mappability regions, high/low GC regions, tandem repeats, and difficult XY regions. All comparisons were therefore evaluated after applying a custom mask on the genome. Namely, we first mask technically difficult regions of the genome[73], defined in the BED file found here: https://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/release/genome-stratifications/v3.5/GRCh38@all/Union/GRCh38_notinalldifficultregions.bed.gz, before adding back any variants in the masked regions that were statistically imputed. In sum, this equates to masking out any rare variants within technically difficult regions. For the scenarios in which we only used short reads for phasing, no sites were masked (*i.e.*, all sites that could be statistically phased were used). Across the base case phasing results, an average of 2.56% sites across the four families were masked as a result of this procedure. We ignored sites with Mendelian errors in the born child when evaluating.

Genotype imputation accuracy was characterized by (1) the dosage correlation between the imputed dosage at each site and the numeric genotypes in the born child VCFs and (2) genotype concordance, the proportion of imputed hard genotype calls concordant with the genotypes in the born child VCFs. Results are presented primarily for sites that are segregating within the trio (*i.e.*, sites where at least one parent is heterozygous), as these are the sites where the inherited genotype is not known *a priori* based on Mendelian inheritance laws.

Pathogenic variant in FAM1

A pathogenic variant was detected in the father and child of FAM1. The rare pathogenic frameshift variant in VPS13C (NG_027782.1:g.145304dup, rs1315150327), resulting from a single thymine insertion at chr15:61920162 (`GRCh38`) was detected and called with high confidence in the trio. The child was called heterozygous with a sequencing depth of 51x at the variant and a GQ of 54; the father was called heterozygous with a depth of 47x and a GQ of 44; the mother was called homozygous reference (no mutation) with a depth of 51x and a GQ of 50. In the long read data generated for the father and mother, this site was covered at depths of 38x and 32x respectively. In the PGT-A data, this site received no sequencing reads.

References

[1] ASRM. US IVF usage increases in 2023, leads to over 95,000 babies born. *American Society for Reproductive Medicine* (2025). URL https://www.asrm.org/news-and-events/asrm-news/press-releasesbulletins/us-ivf-usage-increases-in-2023-leads-to-over-95000-babies-born/. Type: [Press release].

[2] Tecco, H., Shah, M. & Mbaye, M. *State of egg freezing: 2025 trends and insights* (Cofertility, 2024). URL https://www.cofertility.com/freeze-learn/state-of-egg-freezing-2025.

[3] Hamilton, B., Martin, J., Osterman, M. & Rossen, L. Births: Provisional data for 2018 (2019). https://www.cdc.gov/nchs/data/vsrr/vsrr-007-508.pdf.

[4] Vermeesch, J., Voet, T. & Devriendt, K. Prenatal and pre-implantation genetic diagnosis. *Nature Reviews Genetics* **17**, 643–656 (2016). URL https://doi.org/10.1038/nrg.2016.97.

[5] Morales, C. Current applications and controversies in pre-implantation genetic testing for aneuploidies (PGT-A) in in vitro fertilization. *Reproductive Sciences* **31**, 66–80 (2024). URL https://doi.org/10.1007/s43032-023-01301-0.

[6] Practice Committees of the American Society for Reproductive Medicine and the Society for Assisted Reproductive Technology. The use of preimplantation genetic testing for aneuploidy: a committee opinion. *Fertility and Sterility* **122**, 421–434 (2023). URL https://www.asrm.org/globalassets/_asrm/practice-guidance/practice-guidelines/pdf/use_of_preimplantation_genetic_testing_for_aneuploidy.pdf.

[7] Munné, S. & Griffin, D. K. Opinion: contemporary insights into the efficacy of preimplantation genetic testing for aneuploidy (pgt-a) by mining the society for assisted reproductive technology (sart) database (2024).

[8] Lencz, T. *et al.* Utility of polygenic embryo screening for disease depends on the selection strategy. *eLife* **10**, 64716 (2021). URL https://doi.org/10.7554/eLife.64716.

[9] Moore, S. *et al.* Development and validation of polygenic scores for within-family prediction of disease risks (2025). URL http://medrxiv.org/lookup/doi/10.1101/2025.08.06.25333145.

[10] Kumar, A. *et al.* Whole-genome risk prediction of common diseases in human pre-implantation embryos. *Nature Medicine* **28**, 513–516 (2022). URL https://doi.org/10.1038/s41591-022-01735-0.

[11] Widen, E., Lello, L., Raben, T., Tellier, L. & Hsu, S. Polygenic health index, general health and pleiotropy: Sibling analysis and disease-risk reduction. *Scientific Reports* **12**, 18173 (2022). URL https://doi.org/10.1038/s41598-022-22637-8.

[12] Li, S. *et al.* Concordance of whole-genome amplified embryonic DNA with the subsequently born child (2024).

[13] Ahangari, M. *et al.* Comparative evaluation of cross-ancestry polygenic risk scoring of type 1 diabetes in the all of us cohort. *medRxiv* (2025). URL https://www.medrxiv.org/content/early/2025/10/05/2025.10.02.25337098. https://www.medrxiv.org/content/early/2025/10/05/2025.10.02.25337098.full.pdf.

[14] Craig, A. *et al.* Stage-structured, distributional prediction of ivf outcomes with conditional updating. *medRxiv* (2025). URL https://www.medrxiv.org/content/early/2025/10/01/2025.09.27.25336680. https://www.medrxiv.org/content/early/2025/10/01/2025.09.27.25336680.full.pdf.

[15] De Souza, M. Testing Embryos for Polygenic Conditions and Traits. In *The Regulation of Embryo Testing in Australia*, 141–166 (Springer Nature Singapore, Singapore, 2025). URL https://link.springer.com/10.1007/978-981-96-5902-9_6.

[16] Anomaly, J. *Creating Future People: The Science and Ethics of Genetic Enhancement* (Taylor & Francis, 2024).

[17] Forzano, F. *et al.* The use of polygenic risk scores in pre-implantation genetic testing: An unproven, unethical practice. *European Journal of Human Genetics* **30**, 493–495 (2022). URL https://doi.org/10.1038/s41431-021-01000-x.

[18] Grebe, T. *et al.* Clinical utility of polygenic risk scores for embryo selection: A points to consider statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in Medicine* **26**, 101052 (2024). URL https://doi.org/10.1016/j.gim.2023.101052.

[19] Meyer, M., Tan, T., Benjamin, D., Laibson, D. & Turley, P. Public views on polygenic screening of embryos. *Science* **379**, 541–543 (2023). URL https://doi.org/10.1126/science.ade1083.

[20] Furrer, R. A. *et al.* Public attitudes, interests, and concerns regarding polygenic embryo screening. *JAMA Network Open* **7**, e2410832–e2410832 (2024).

[21] Haining, C. M., Savulescu, J., Keogh, L. & Schaefer, G. O. Polygenic risk scores and embryonic screening: considerations for regulation. *Journal of medical ethics* **51**, 719–728 (2025).

[22] Abdellaoui, A., Yengo, L., Verweij, K. & Visscher, P. Fifteen years of GWAS discovery: Realizing the promise. *American Journal of Human Genetics* **110**, 179–194 (2023). URL https://doi.org/10.1016/j.ajhg.2022.12.011.

[23] Xia, Y. *et al.* The first clinical validation of whole-genome screening on standard trophectoderm biopsies of preimplantation embryos. *F&S Reports* **5**, 63–71 (2024). URL https://linkinghub.elsevier.com/retrieve/pii/S2666334124000011.

[24] Wang, Q. *et al.* Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature* **597**, 527–532 (2021). URL https://www.nature.com/articles/s41586-021-03855-y. Publisher: Springer Science and Business Media LLC.

[25] Hernandez, R. D. *et al.* Ultrarare variants drive substantial cis heritability of human gene expression. *Nature Genetics* **51**, 1349–1355 (2019). URL https://www.nature.com/articles/s41588-019-0487-7. Publisher: Springer Science and Business Media LLC.

[26] Nguyen, D. T. *et al.* A comprehensive evaluation of polygenic score and genotype imputation performances of human SNP arrays in diverse populations. *Scientific Reports* **12**, 17556 (2022). URL https://www.nature.com/articles/s41598-022-22215-y.

[27] Heiser, H. *et al.* The embryo mosaicism profile of next-generation sequencing PGT-A in different clinical conditions and their associations. *Frontiers in Reproductive Health* **5**, 1132662 (2023). URL https://doi.org/10.3389/frph.2023.1132662.

[28] Sun, S. *et al.* Identifying risk variants for embryo aneuploidy using ultra-low coverage whole-genome sequencing from preimplantatilimion genetic testing. *The American Journal of Human Genetics* **110**, 2092–2102 (2023).

[29] Li, S. *et al.* Ultra-low-coverage genome-wide association study—insights into gestational age using 17,844 embryo samples with preimplantation genetic testing. *Genome medicine* **15**, 10 (2023).

[30] Davies, R. W., Flint, J., Myers, S. & Mott, R. Rapid genotype imputation from sequence without reference panels. *Nature Genetics* **48**, 965–969 (2016). URL https://www.nature.com/articles/ng.3594. Publisher: Springer Science and Business Media LLC.

[31] Rubinacci, S., Ribeiro, D. M., Hofmeister, R. J. & Delaneau, O. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nature Genetics* **53**, 120–126 (2021). URL https://www.nature.com/articles/s41588-020-00756-0. Publisher: Springer Science and Business Media LLC.

[32] Viotti, M. Preimplantation Genetic Testing for Chromosomal Abnormalities: Aneuploidy, Mosaicism, and Structural Rearrangements. *Genes* **11**, 602 (2020). URL https://www.mdpi.com/2073-4425/11/6/602. Publisher: MDPI AG.

[33] Wasik, K. *et al.* Comparing low-pass sequencing and genotyping for trait mapping in pharmacogenetics. *BMC Genomics* **22**, 197 (2021). URL https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-021-07508-2.

[34] Davies, R. W. *et al.* Rapid genotype imputation from sequence with reference panels. *Nature Genetics* **53**, 1104–1111 (2021).

[35] Li, J., Mazur, C., Berisa, T. & Pickrell, J. Low-pass sequencing increases the power of GWAS and decreases measurement error of polygenic risk scores compared with genotyping arrays. *Genome Research* **31**, 529–537 (2021). URL https://doi.org/10.1101/gr.266486.120.

[36] Martin, A. *et al.* Low-coverage sequencing cost-effectively detects known and novel variation in underrepresented populations. *The American Journal of Human Genetics* **108**, 656–668 (2021).

[37] Natesan, S. *et al.* Genome-wide karyomapping accurately identifies the inheritance of single-gene defects in human pre-implantation embryos in vitro. *Genetics in Medicine* **16**, 838–845 (2014). URL https://doi.org/10.1038/gim.2014.45.

[38] Yan, L. *et al.* Live births after simultaneous avoidance of monogenic diseases and chromosome abnormality by next-generation sequencing with linkage analyses. *Proceedings of the National Academy of Sciences* **112**, 15964–15969 (2015). URL https://doi.org/10.1073/pnas.1523297113.

[39] Backenroth, D. *et al.* Haploseek: a 24-hour all-in-one method for preimplantation genetic diagnosis (pgd) of monogenic disease and aneuploidy. *Genetics in Medicine* **21**, 1390–1399 (2019).

[40] Backenroth, D. *et al.* SHaploseek is a sequencing-only, high-resolution method for comprehensive pre-implantation genetic testing. *Scientific Reports* **13**, 18036 (2023). URL https://doi.org/10.1038/s41598-023-45292-z.

[41] Janssen, A. E. J. *et al.* Clinical-grade whole genome sequencing-based haplarithmisis enables all forms of preimplantation genetic testing. *Nature Communications* **15** (2024). URL https://www.nature.com/articles/s41467-024-51508-1. Publisher: Springer Science and Business Media LLC.

[42] Choi, Y., Chan, A., Kirkness, E., Telenti, A. & Schork, N. Comparison of phasing strategies for whole human genomes. *PLoS genetics* **14**, 1007308 (2018).

[43] Kronenberg, Z. *et al.* The Platinum Pedigree: a long-read benchmark for genetic variants. *Nature Methods* **22**, 1669–1676 (2025).

[44] Bhérer, C., Campbell, C. & Auton, A. Refined genetic maps reveal sexual dimorphism in human meiotic recombination at multiple scales. *Nature communications* **8**, 14994 (2017).

[45] Gorjanc, G., Bijma, P. & Hickey, J. M. Reliability of pedigree-based and genomic evaluations in selected populations. *Genetics Selection Evolution* **47**, 65 (2015). URL http://www.gsejournal.org/content/47/1/65.

[46] Karavani, E. *et al.* Screening Human Embryos for Polygenic Traits Has Limited Utility. *Cell* **179**, 1424–1435.e8 (2019). URL https://linkinghub.elsevier.com/retrieve/pii/S0092867419312103.

[47] Wedel, I. Falconer, D. S.: Introduction to Quantitative Genetics. Oliver and Boyd, Edinburgh and London 1960; 365 S., 35 s. *Biometrische Zeitschrift* **4**, 140–141 (1962). URL https://onlinelibrary.wiley.com/doi/10.1002/bimj.19620040211.

[48] Martin, M. *et al.* WhatsHap: fast and accurate read-based phasing. *BioRxiv* **085050** (2016).

[49] O'Connell, J. *et al.* A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. *PLoS Genetics* **10**, e1004234 (2014). URL https://dx.plos.org/10.1371/journal.pgen.1004234.

[50] Lesage, S. *et al.* Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy. *The American Journal of Human Genetics* **98**, 500–513 (2016). URL https://linkinghub.elsevier.com/retrieve/pii/S0002929716000483. Publisher: Elsevier BV.

[51] Chen, S. *et al.* A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625**, 92–100 (2024).

[52] Hofmeister, R., Ribeiro, D., Rubinacci, S. & Delaneau, O. Accurate rare variant phasing of whole-genome and whole-exome sequencing data in the UK Biobank. *Nature genetics* **55**, 1243–1249 (2023).

[53] Petter, E. *et al.* Genotype error due to low-coverage sequencing induces uncertainty in polygenic scoring. *The American Journal of Human Genetics* **110**, 1319–1329 (2023). URL https://linkinghub.elsevier.com/retrieve/pii/S0002929723002409.

[54] Chen, S.-F. *et al.* Genotype imputation and variability in polygenic risk score estimation. *Genome Medicine* **12**, 100 (2020). URL https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-020-00801-x.

[55] Rubinacci, S., Hofmeister, R., Mota, B. & Delaneau, O. Imputation of low-coverage sequencing data from 150 119 UK Biobank genomes. *Nature Genetics* **55**, 1088–1090 (2023). URL https://doi.org/10.1038/s41588-023-01438-3.

[56] Al Kuwari, H. *et al.* The Qatar Biobank: background and methods. *BMC Public Health* **15**, 1208 (2015). URL http://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-015-2522-7.

[57] Kawai, Y. *et al.* Exploring the genetic diversity of the Japanese population: Insights from a large-scale whole genome sequencing analysis. *PLOS Genetics* **19**, e1010625 (2023). URL https://dx.plos.org/10.1371/journal.pgen.1010625.

[58] Yu, C. *et al.* A high-resolution haplotype-resolved Reference panel constructed from the China Kadoorie Biobank Study. *Nucleic Acids Research* **51**, 11770–11782 (2023). URL https://academic.oup.com/nar/article/51/21/11770/7327062.

[59] Rabiner, L. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* **77**, 257–286 (1989). URL http://ieeexplore.ieee.org/document/18626/. Publisher: Institute of Electrical and Electronics Engineers (IEEE).

[60] Browning, B. & Browning, S. Genotype error biases trio-based estimates of haplotype phase accuracy. *The American Journal of Human Genetics* **109**, 1016–1025 (2022).

[61] Williams, C. M. *et al.* Phasing millions of samples achieves near perfect accuracy, enabling parent-of-origin analyses. *Human Genetics and Genomics Advances* **6** (2025).

[62] Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015). URL https://academic.oup.com/bioinformatics/article/31/12/2032/214758. Publisher: Oxford University Press (OUP).

[63] Delaneau, O., Zagury, J., Robinson, M., Marchini, J. & Dermitzakis, E. Accurate, scalable and integrative haplotype estimation. *Nature communications* **10**, 5436 (2019).

[64] Cavinato, T., Rubinacci, S., Malaspinas, A.-S. & Delaneau, O. A resampling-based approach to share reference panels. *Nature Computational Science* **4**, 360–366 (2024). URL https://www.nature.com/articles/s43588-024-00630-7.

[65] Koenig, Z. *et al.* A harmonized public resource of deeply sequenced diverse human genomes. *Genome Research* **34**, 796–809 (2024). URL http://genome.cshlp.org/lookup/doi/10.1101/gr.278378.123. Publisher: Cold Spring Harbor Laboratory.

[66] Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, s13742–015–0047–8 (2015). URL https://academic.oup.com/gigascience/article/doi/10.1186/s13742-015-0047-8/2707533.

[67] Thermo Fisher Scientific. UK Biobank Axiom™ Array (2025). https://www.thermofisher.com/order/catalog/product/902502.

[68] Broad Institute. Picard Tools (2025). Available from: https://broadinstitute.github.io/picard/index.html.

[69] Illumina. HumanCytoSNP-12 BeadChip Support (2025). Available from: https://support.illumina.com/array/array_kits/humancyto-snp-12_v2-1_dna_analysis_kit.html.

[70] Illumina. Illumina Array Analysis Platform Genotyping Command Line Interface (2025). Available from: https://emea.support.illumina.com/downloads/iaap-genotyping-cli.html.

[71] Genovese, G. freeseek/gtc2vcf (2025). URL https://github.com/freeseek/gtc2vcf. Original-date: 2018-07-06T22:52:57Z.

[72] Rubinacci, S., Delaneau, O. & Marchini, J. Genotype imputation using the Positional Burrows Wheeler Transform. *PLOS Genetics* **16**, e1009049 (2020). URL https://dx.plos.org/10.1371/journal.pgen.1009049.

[73] The Global Alliance for Genomics and Health Benchmarking Team *et al.* Best practices for benchmarking germline small-variant calls in human genomes. *Nature Biotechnology* **37**, 555–560 (2019). URL https://www.nature.com/articles/s41587-019-0054-x.

Acknowledgements

We thank Shai Carmi, Joseph Pickrell, Lex Flagel, Robert Maier, Jonathan Anomaly, and Santiago Munne for feedback on an earlier version of the manuscript.

Figure 1 was created using BioRender. This research has been conducted using the UK Biobank Resource under Application Number 103244.

Author contributions

J.H.L. performed the computational experiments and analyses. A.S.Y. developed the original mathematical and statistical methodology for ImputePGTA; M.C. assisted in the original conceptualization of the method. T.W. implemented the original model and software for ImputePGTA with assistance from A.S.Y.. J.H.L. contributed improvements to ImputePGTA's implementation and algorithm. A.S.Y. and J.H.L developed extensions to the original mathematical treatments. J.H.L. developed and implemented the phase unification method (phaseGrafter). Sample processing, sequencing, and primary and secondary analyses were validated and performed by J.S. and J.S.. J.H.L., T.W., A.S.Y. designed the study and wrote the manuscript. All authors reviewed the final manuscript. I.D. and J.S. were responsible for setting up the IRB study and provided operational support.

Competing interests

J.H.L., T.W., I.D., J.S., J.S., S.M., D.S., and M.C. are employed by Herasight Inc., a private company developing tests for preimplantation genetic screening. A.S.Y. is an advisor and shareholder of Herasight Inc..

Supplementary Note: Platinum Pedigree and 1000 Genomes Data Description

Platinum Pedigree dataset

The Platinum Pedigree resource is a publicly available gold standard pedigree dataset comprising whole genome sequence data from five technologies across a four-generation pedigree[43]. As of the time of writing this comprises the most fully-validated publicly available dataset for variant calls across multiple generations. We focused on generations 2 and 3 (G2, G3) as part of this study (the "parents" and "offspring"). G2 comprises NA12878 and NA12877, the mother and father respectively of the individuals in G3 which we analyzed: NA12879, NA12881, NA12882, NA12885, and NA12886.

The two main datasets we used in this study were (1) the "Pedigree consistent merged small variant calls (truthset)" VCF which comprise a unified callset with pedigree-consistent phased SNPs and small indels for G2 and G3, and (2) the high-depth short read whole genome sequences in BAM format available for G3. (1) was used as the source for the parental haplotypes input to ImputePGTA and as the starting point for the simulations described below. (1) was also used to derive the truth calls for G3. (2) was used as a starting point for the downsampling experiments described below.

The VCF truthset was obtained from the following s3 path:
s3://platinum-pedigree-data/variants/small_variant_truthset/GRCh38/CEPH1463.GRCh38.family-truthset.ov.vcf.gz

The high-coverage short read alignments for the offspring were acquired from the following s3 path:
s3://platinum-pedigree-data/data/illumina/mapped/GRCh38/

We preprocessed the VCF truthset by splitting multiallelics using `bcftools norm -m -any` and filtering to autosomes, resulting in 5949154 variants after deduplication.

HGDP + 1KG Reference Panel

We used the HGDP + 1KG dataset[2], a cosmopolitan dataset representing a diverse set of human genomes, as a reference panel for PGS ancestry adjustment. Briefly, we obtained the raw genotype calls on GRCh38 from the gnomAD website[51], annotated them using dbSNP build 157, and filtered them to the variants in our PGS. Ancestry adjustment was then performed as described in the **Methods**.

Supplementary Figures

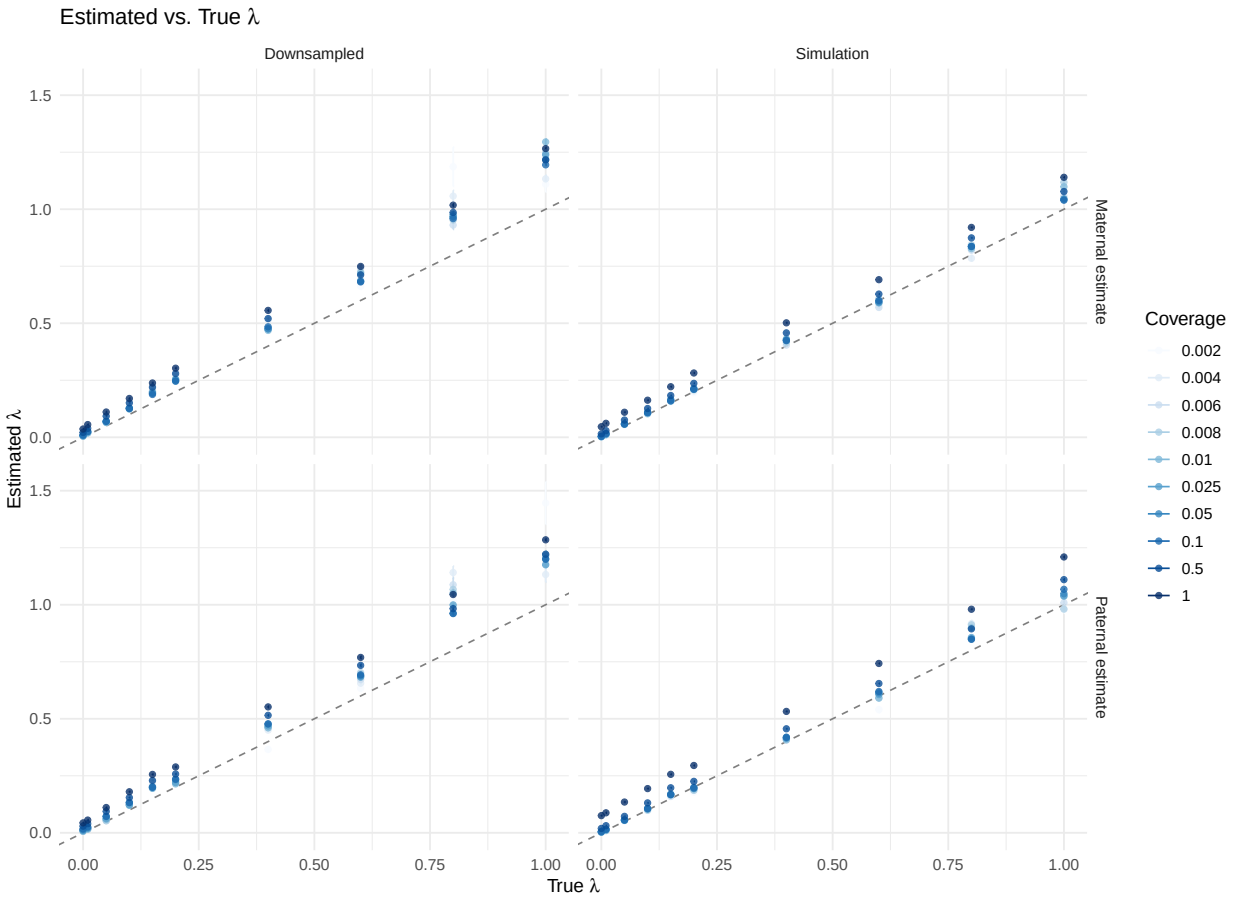

Figure S1: Calibration of lambda estimates. Estimated vs true switch error rates as described by λ (switch errors/cM) for the simulated offspring and downsampled offspring data. The true λ reflects the rate at which we introduced switch errors into the parental haplotypes. Estimates for the paternal and maternal switch error rates are shown on the y-axis and the true λ values are shown on the x-axis. The estimates are generally well-calibrated but display an upward bias at higher coverages and higher levels of λ.

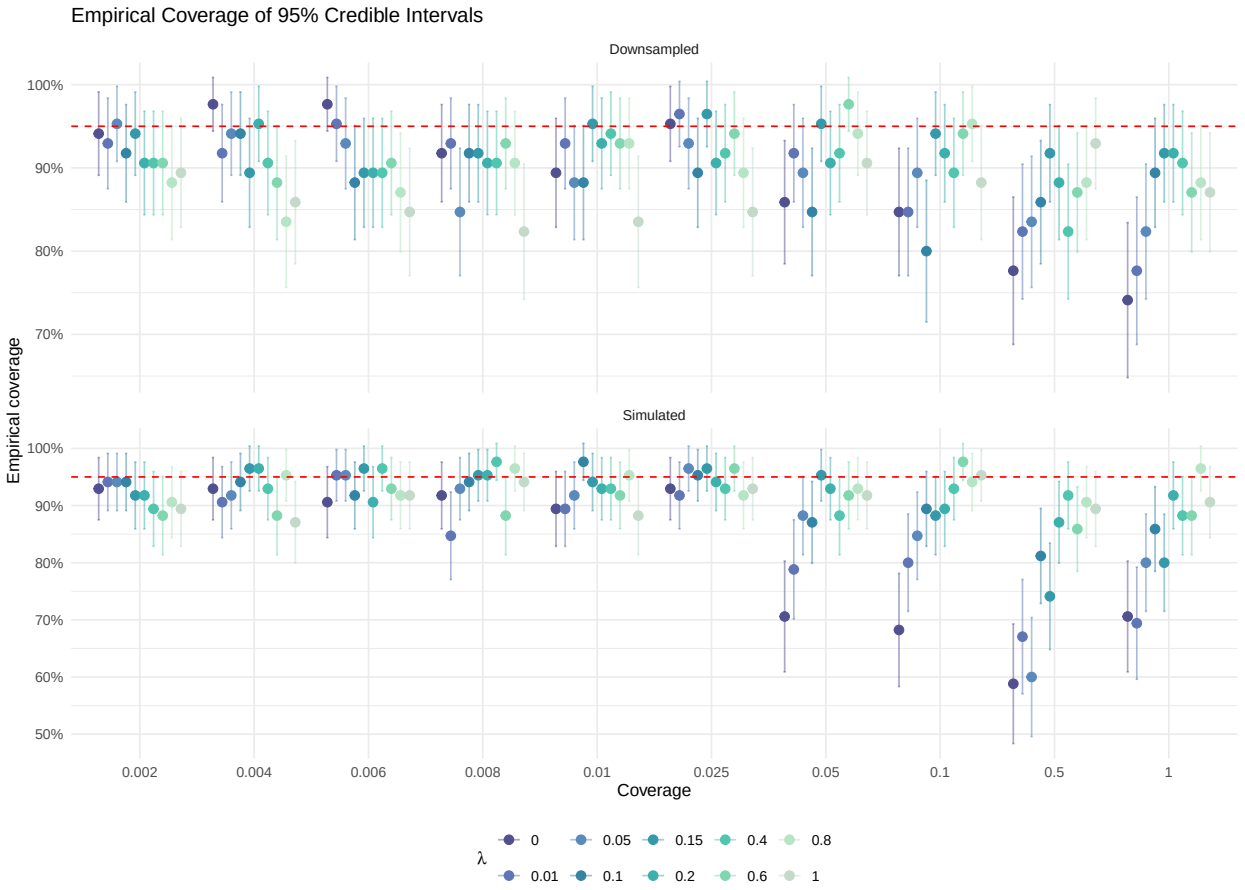

Figure S2: Empirical coverage of PGS posterior 95% credible intervals. The empirical coverage of the 95% equal tailed credible intervals for the in silico experiments. For each coverage x lambda combination, the empirical coverage of the credible intervals across all 5 samples and 17 traits is shown with the standard errors. We observe overall good coverage especially in the more difficult cases — *i.e.*, at lower coverages and higher switch error rates. The empirical coverage is somewhat lower at higher coverages when λ is low, though the MAE of the posterior mean at these parameter regimes is extremely low.

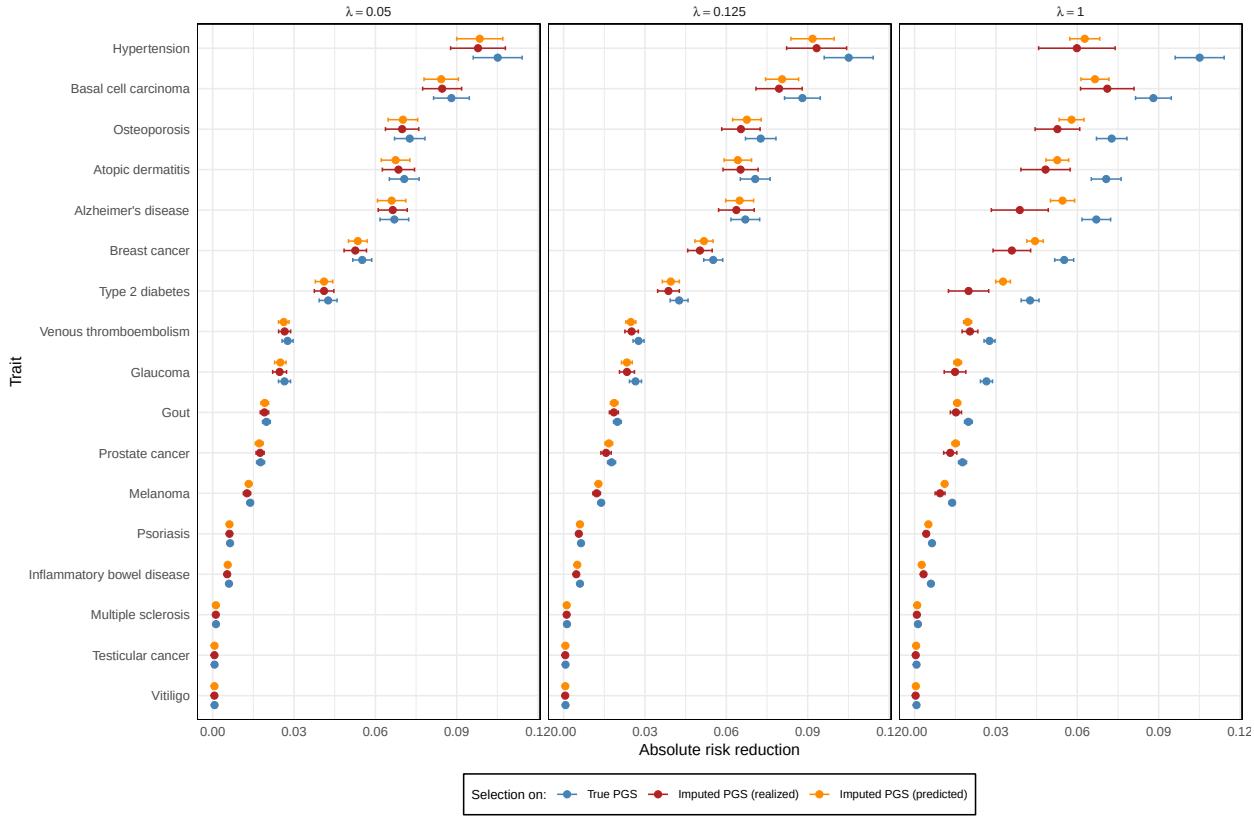

Figure S3: Absolute risk reduction due to selection on imputed PGS. The x-axis shows the mean absolute risk reduction (95% CIs) from 100 instances of 5-embryo families with the same parents assuming a liability threshold disease model. The panels show results for three values of λ, the switch-error rate in parental haplotypes: the value based on empirical results for parents phased with short reads and long reads, ($\lambda = 0.05$), short reads only ($\lambda = 0.125$), and the highest value analyzed in the *in silico* experiments ($\lambda = 1$). For each panel, three results are shown: the first two are the absolute risk reductions from two selection strategies: (a) selection based on true PGS values (blue), and (b) selection based on the imputed PGS values (red). The third is the theoretical absolute risk reduction predicted by the posterior variances (orange). The difference between the red and blue data points reflects the actual decrease in absolute risk reduction (based on true PGS values) resulting from imputation uncertainty.

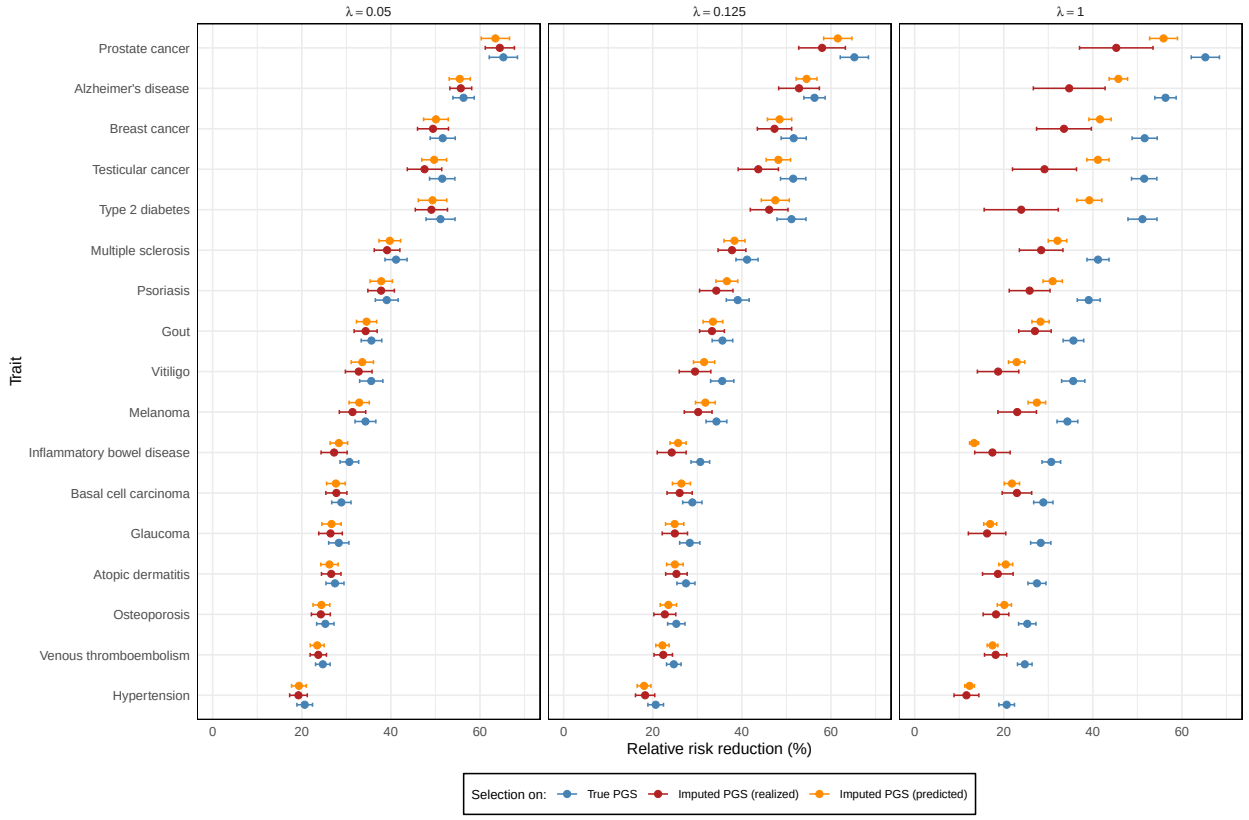

Figure S4: Relative risk reduction due to selection on imputed PGS. The x-axis shows the mean relative risk reduction (95% CIs) with respect to the risk corresponding to the within-family embryo mean PGS. These are calculated from 100 instances of 5-embryo families with the same parents assuming a liability threshold disease model. The panels show results for three values of λ, the switch-error rate in parental haplotypes: the value based on empirical results for parents phased with short reads and long reads, ($\lambda = 0.05$), short reads only ($\lambda = 0.125$), and the highest value analyzed in the *in silico* experiments ($\lambda = 1$). For each panel, three results are shown: the first two are the relative risk reductions from two selection strategies: (a) selection based on true PGS values (blue), and (b) selection based on the imputed PGS values (red). The third is the theoretical relative risk reduction predicted by the posterior variances (orange). The difference between the red and blue data points reflects the actual decrease in relative risk reduction (based on true PGS values) resulting from imputation uncertainty.

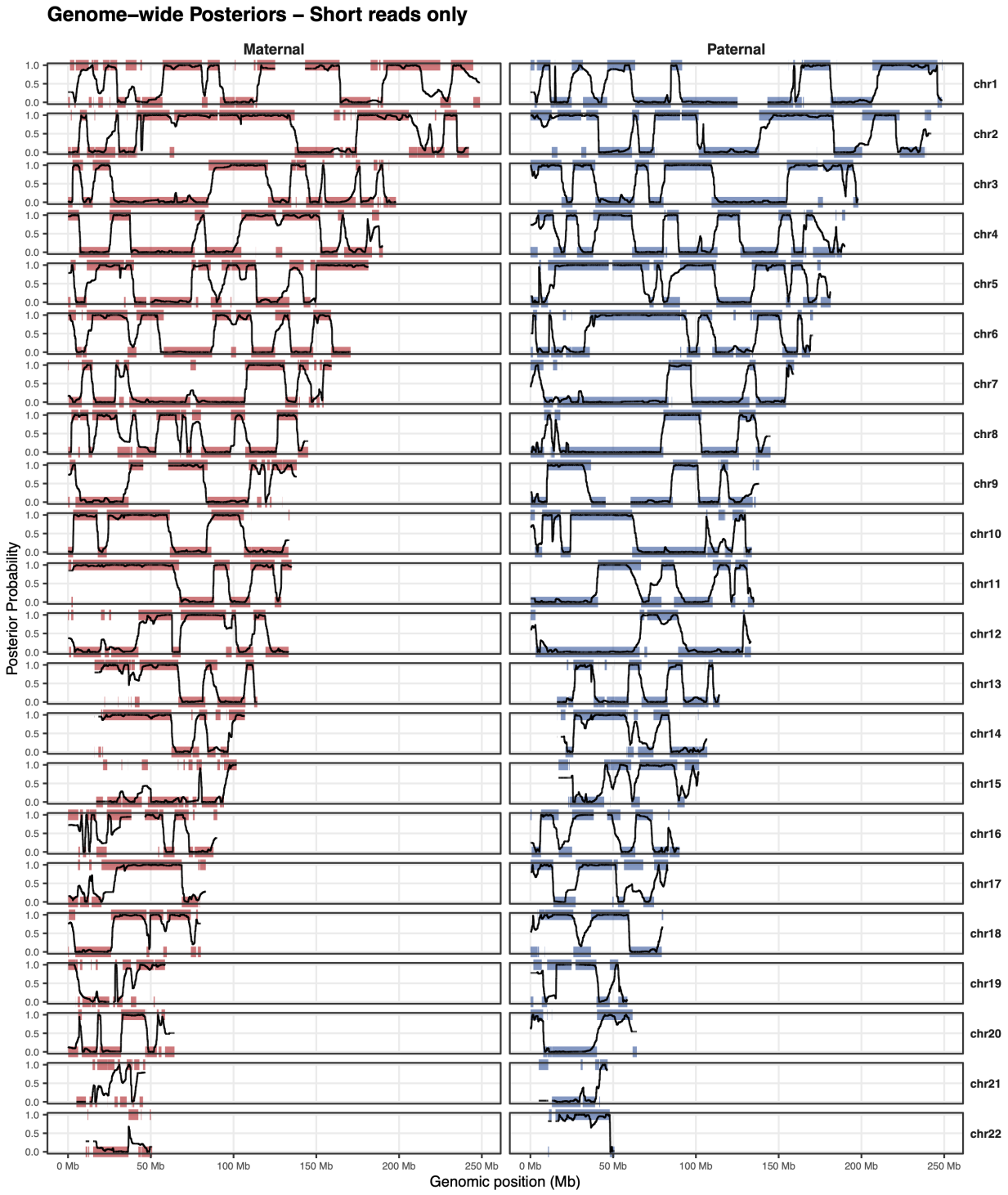

Figure S5: Genome-wide posterior plots for FAM1 parents using short reads only. Each pane depicts in thick colored lines the smoothed posterior probabilities obtained from running ImputePGTA on the born child's genotypes obtained from high-coverage WGS. Probabilities at 0 and 1 indicate inheritance of one or the other of the parental haplotypes. In black are shown the raw posterior probabilities resulting from running ImputePGTA on the ULP PGT-A data. Gaps on the x-axis reflect regions of at least 1Mb where no variants were genotyped in the parents.

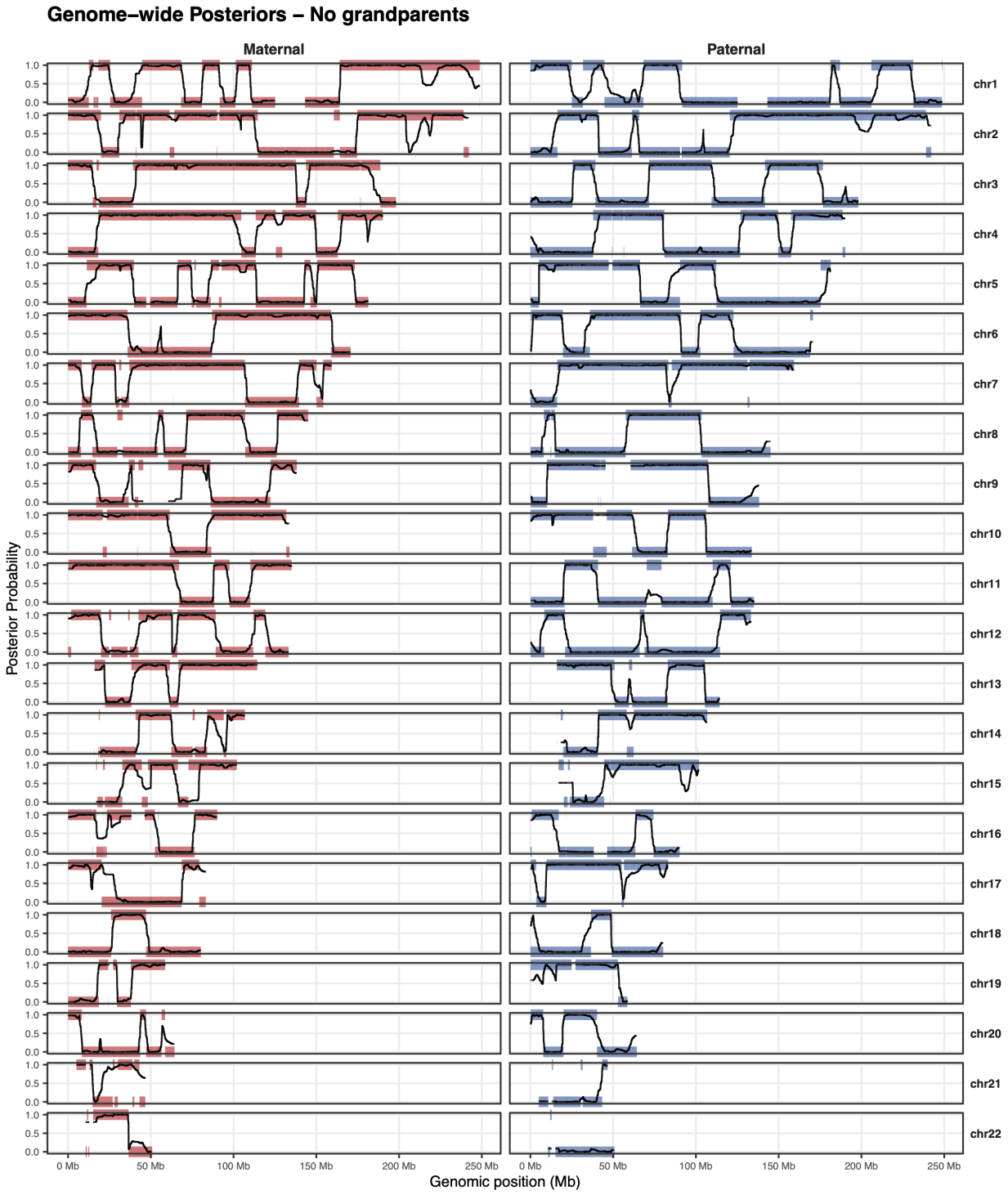

Figure S6: Genome-wide posterior plots for FAM1 parents using short and long reads. Each pane depicts in thick colored lines the smoothed posterior probabilities obtained from running ImputePGTA on the born child's genotypes obtained from high-coverage WGS. Probabilities at 0 and 1 indicate inheritance of one or the other of the parental haplotypes. In black are shown the raw posterior probabilities resulting from running ImputePGTA on the ULP PGT-A data. Gaps on the x-axis reflect regions of at least 1Mb where no variants were genotyped in the parents.

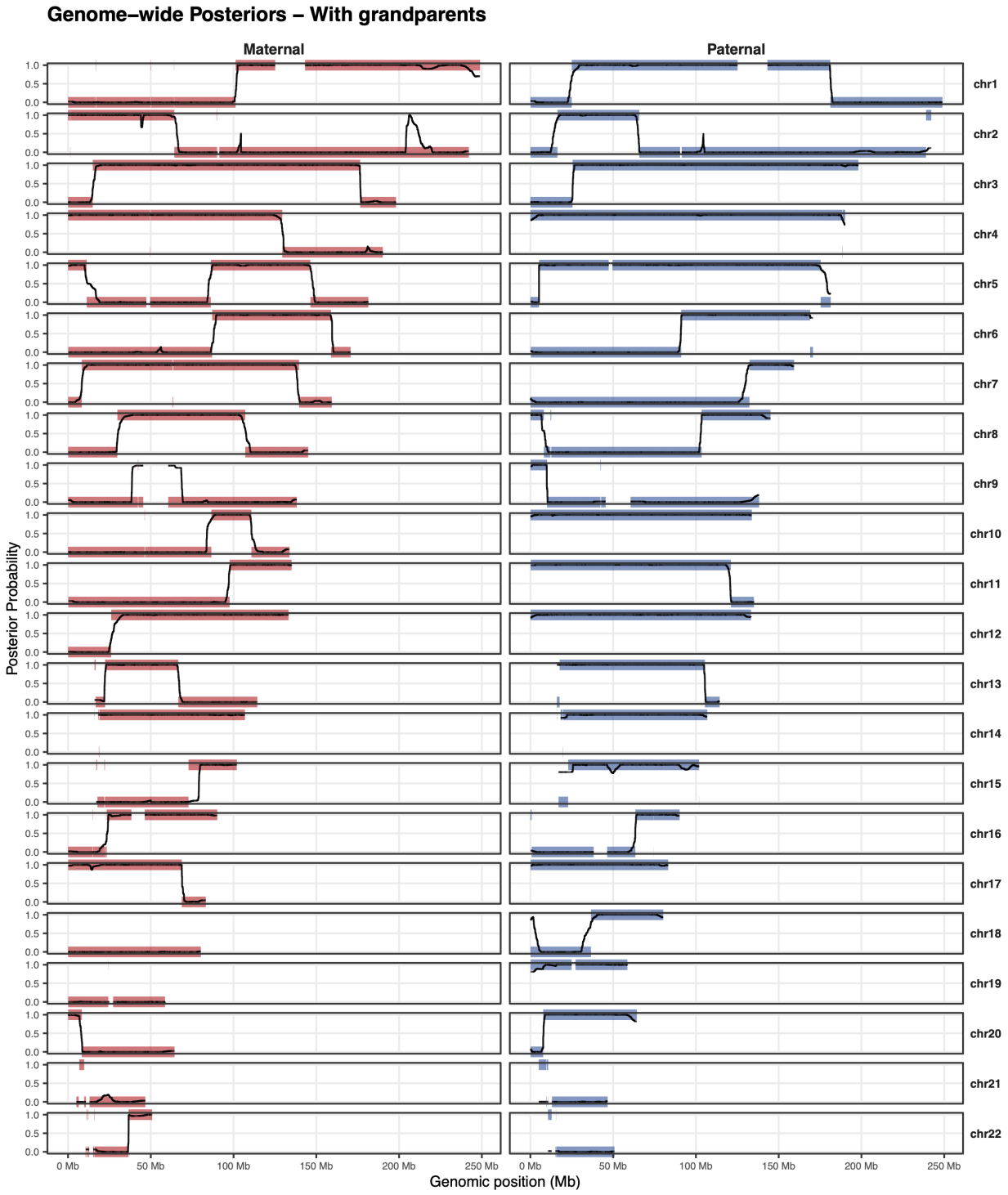

Figure S7: Genome-wide posterior plots for FAM1 parents using short and reads and grandparental data. Each pane depicts in thick colored lines the smoothed posterior probabilities obtained from running ImputePGTA on the born child's genotypes obtained from high-coverage WGS. Probabilities at 0 and 1 indicate inheritance of one or the other of the parental haplotypes. In black are shown the raw posterior probabilities resulting from running ImputePGTA on the ULP PGT-A data. Gaps on the x-axis reflect regions of at least 1Mb where no variants were genotyped in the parents.