Comparison and assessment of family- and population-based genotype imputation methods in large pedigrees

Ehsan Ullah, Raghvendra Mall, Mostafa M. Abbas, et al.

Genome Res. published online December 4, 2018
Access the most recent version at doi:10.1101/gr.236315.118

Comparison and Assessment of Family- and Population-based Genotype Imputation Methods in Large Pedigrees

Ehsan Ullah[1†], Raghvendra Mall[1†], Mostafa M. Abbas[1‡], Khalid Kunji[1‡], Alejandro Q. Nato Jr[2,3], Halima Bensmail[1], Ellen M. Wijsman[2,4], Mohamad Saad[1*]

[1]Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar

[2]Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, USA

[3]Department of Biomedical Sciences, Joan C. Edwards School of Medicine, Marshall University, Huntington, WV, USA

[4]Department of Biostatistics, University of Washington, Seattle, WA, USA

[†,‡] Contributed equally

*Corresponding author: Mohamad Saad, Qatar Computing Research Institute, Office 1117A, HBKU – Research Complex, Hamad Bin Khalifa University, P.O. Box: 5825, Doha - Qatar

Email : msaad@hbku.edu.qa; Phone : +974 4454 7746

Running title: Genotype Imputation in Large Pedigrees

Keywords: Genotype imputation, Pedigree data, Association analysis, MaCH, minimac, SHAPEIT, Eagle, Beagle, IMPUTE, GIGI, Merlin, Imputation accuracy.

ABSTRACT

Genotype imputation is widely used in genome-wide association studies to boost variant density, allowing increased power in association testing. Many studies currently include pedigree data due to increasing interest in rare variants coupled with the availability of appropriate analysis tools. The performance of population-based (subjects are unrelated) imputation methods is well established. However, the performance of family- and population-based imputation methods on family data has been subject to much less scrutiny. Here, we extensively compare several family- and population-based imputation methods on family data of large pedigrees with both European and African ancestry. Our comparison includes many widely used family- and population-based tools and another method, Ped_Pop, which combines family- and population-based imputation results. We also compare four subject selection strategies for full sequencing to serve as the reference panel for imputation: GIGI-Pick, ExomePicks, PRIMUS, and Random selection. Moreover, we compare two imputation accuracy metrics: the Imputation Quality Score and Pearson correlation R^2 for predicting power of association analysis using imputation results. Our results show that: 1) GIGI outperforms Merlin, 2) family-based imputation outperforms population-based imputation for rare variants but not for common ones, 3) combining family- and population-based imputation outperforms all imputation approaches for all minor allele frequencies, 4) GIGI-Pick gives the best selection strategy based on the R^2 criterion, and 5) R^2 is the best measure of imputation accuracy. Our study is the first to extensively evaluate the imputation performance of many available family- and population-based tools on the same family data, and provides guidelines for future studies.

INTRODUCTION

Genome-wide association studies (GWAS) have led to the discovery of hundreds of loci associated with complex diseases (Manolio et al. 2009; Visscher et al. 2017; Marigorta et al. 2018). Large sample sizes are required to achieve the necessary statistical power to identify such loci (Wang et al. 2005). Research consortia have attained large sample sizes by combining data from several studies using joint or meta-analysis (Evangelou et al. 2007; International Parkinson's Disease Genomics Consortium (IPDGC) and Wellcome Trust Case Control Consortium 2 (WTCCC2) 2011; Nalls et al. 2011; Siddiq et al. 2012; Nalls et al. 2014; Sniekers et al. 2017). These studies involved imputation of missing genotypes to allow association analysis of the same SNPs in multiple studies. Imputation facilitates performing joint or meta-analysis and also permits increasing the genomic coverage by searching for association on a much denser map. For all these reasons, performing imputation in GWAS data has become a common step (Marchini and Howie 2010). However, despite the enormous samples used, substantial heritability is not explained by the identified associations, leading to the conclusion that there is substantial rare variation that is also important and may explain part of the missing heritability (Maher 2008; Manolio et al. 2009; Visscher et al. 2017). Association with rare variants is difficult to find in analysis of unrelated subjects but can be identified in family-based designs, raising interest, once again, in family-based studies (Wijsman 2012).

To efficiently impute rare variants, imputation approaches that work well in general pedigrees are needed. To date, only a few methods have been proposed for family-based imputation designs, including Merlin (Burdick et al. 2006), GIGI

(Cheung et al. 2013) (coupled with gl_auto (Thompson 2011)), PRIMAL (Livne et al. 2015), and cnF2freq (Nettelblad 2012). These approaches use, e.g., sequencing data on a small set of subjects from the studied pedigrees and infer the missing genotypes on the remaining subjects (Saad and Wijsman 2013). Unlike Merlin and GIGI, PRIMAL and cnF2freq are not set up for general use. Merlin and GIGI rely on Identity by Descent (IBD) computation, which is mostly identical for these tools and is based on the Lander-Green algorithm (Lander and Green 1987). The two main differences between the programs are their different approach to the treatment of alleles in founders when such alleles are undefined by the data within the pedigree, together with their different capabilities for large pedigrees. GIGI can handle large pedigrees efficiently, while Merlin cannot, thus requiring pedigree splitting or trimming. In previous studies, the performance of GIGI was compared to several population-based imputation methods (Saad et al. 2016), while the performance of Merlin was separately evaluated on trimmed pedigrees (Lent et al. 2016). The two programs were not compared directly on large pedigrees, although other studies have shown that both GIGI and Merlin perform well for rare variant imputation but not as well for common variants (Chen et al. 2012; Saad and Wijsman 2014). To date, there has not been an evaluation of all the approaches on the same data, used in a way that produces comparable results.

Population-based imputation coupled with phasing methods also exist. Phasing approaches include Eagle (Loh et al. 2016), MaCH (Li et al. 2010), IMPUTE (Howie et al. 2012), Beagle (Browning and Browning 2007), and SHAPEIT (Delaneau et al. 2012). Imputation approaches include minimac (Fuchsberger et al. 2015), IMPUTE (Howie et al. 2009), and Beagle (Browning and Browning 2016), and are more developed than family-based methods. Some of these methods have been

compared in previous studies in both real and simulated data of unrelated subjects (Marchini and Howie 2010), and have also been extensively used in GWAS applications on real data of complex diseases (Nalls et al. 2011; Al-Tassan et al. 2015). The population-based imputation approaches can be used for imputation in family-based GWAS, but they ignore the IBD information and rely only on Linkage Disequilibrium (LD) information, which leads to a loss of information. They may lead to good imputation of common variants, but not for rare variants because of the minimal LD between rare variants (Saad and Wijsman 2014). Moreover, in genomic regions where the LD is minimal between common variants, or the number of typed SNPs is low, IBD-based imputation methods may outperform population-based imputation methods for both rare and common variants. To benefit from both LD and IBD information, one can use the Ped_Pop (https://bioinformatics.qcri.org/ped_pop) approach (Saad and Wijsman 2014), which combines family-based and population-based imputation methods using the best features of each to impute rare and common variants with higher accuracy. Although the original implementation of Ped_Pop combined GIGI and Beagle imputation results, the approach is general, and other combinations of methods could be used just as well. A comprehensive assessment of both family- and population-based imputation on pedigree data has not yet been done.

Imputation accuracy can be evaluated by several metrics. These include the Concordance Rate (CR), the Imputation Quality Score (IQS) (Lin et al. 2010), and Pearson's squared correlation (R^2). For common variants, these metrics provide similar accuracies, but for rare variants, this is not the case. For instance, the CR yields overestimated accuracy because common alleles are easily imputed (Lin et al. 2010). There is a need to know which accuracy metrics work well. Previous studies

compared R^2 and IQS (Ramnarine et al. 2015) but ignored the different meaning of these metrics, in that the R^2 value is the squared correlation and the IQS is an agreement ratio. Moreover, the range of both metrics is not the same, with R^2 varying from 0 to 1 while the upper bound of the IQS is one but the minimum could be negative. This precludes direct comparison of R^2 metric with IQS.

In imputation analysis, the selection of the reference dataset has a great impact on the imputation accuracy. Unlike population-based imputation, which allows the use of external reference datasets (e.g., 1000 Genomes Project (The 1000 Genomes Project Consortium 2015), Haplotype Reference Consortium (HRC) (McCarthy et al. 2016), and UK10K (The UK10K Consortium 2015)), family-based imputation requires the reference dataset subjects to belong to the same pedigrees (Cheung et al. 2013; Saad and Wijsman 2013). Several subject selection approaches exist for pedigree data: GIGI-Pick (Cheung et al. 2014), ExomePicks (see http://genome.sph.umich.edu/wiki/ExomePicks), and PRIMUS (Staples et al. 2013). These approaches aim to select the pedigree members to be sequenced, forming the reference dataset. GIGI-Pick capitalizes on the concept of Inheritance Vectors (IV) that represent the descent of chromosomes in a pedigree at specified positions. ExomePicks selects units of related subjects from the oldest to youngest generations, thus encouraging determination of haplotypes across loci. PRIMUS aims to identify a set of maximally unrelated subjects. The impact of these three approaches on the imputation accuracy of the different phasing and imputation approaches has not been thoroughly compared for both rare and common variants.

Here we show the results of an extensive comparison of imputation methods in family-based data. Our dataset consists of a collection of real pedigrees with small to large sizes. We simulated genetic data on these pedigrees to mimic the minor

allele frequency spectrum and LD of the 1000 Genomes Project to evaluate results for both European and African ancestries. We compared the main family- and population-based combinations of phasing and imputation algorithms: gl_auto, GIGI, Merlin, Eagle, SHAPEIT (with and without the duoHMM feature), MaCH, minimac, IMPUTE, Beagle, and Ped_Pop. To run Merlin, we split all pedigrees into smaller sub-pedigrees that can fit in the memory and then combined the sub-pedigree results. We also compared the effect of four subject selection strategies, GIGI-Pick, ExomePicks, PRIMUS, and Random Selection on the imputation accuracy. Finally, we compared the imputation accuracy measures R^2 and IQS for various Minor Allele Frequency (MAF) intervals with respect to the power of association analysis using a linear mixed model. We ignored the Concordance Rate because of the well-known limitation mentioned earlier (Ramnarine et al. 2015). Our paper represents the first comprehensive guideline to the choice of imputation methods in family-based human genetic data, and delivers answers regarding the choice of the best subject selection for downstream association analysis, and which phasing and imputation methods to use, depending on the context and scenarios of a study.

RESULTS

Mean Squared Correlation (R^2)

The R^2 values were estimated between the imputed and true observed dosages. Here the dosage is the estimated (or observed) fraction of minor alleles in the genotype. This computation was performed for all SNPs except the observed GWAS SNPs (i.e. 500 in both EUR and AFR), which were not imputed. The results of mean R^2 for random selection are summarized in Figure 1A and 1B. This figure shows that for rare or infrequent variants (MAF in (0,0.05)), family-based imputation

methods outperformed population-based methods. GIGI had the same performance in both European (EUR) and African (AFR). The same trend was also observed for Merlin. Within family-based approaches, GIGI (using full pedigrees) outperforms Merlin (using sub-pedigrees) across all MAF intervals for both EUR and AFR. When applied on the same sub-pedigrees, GIGI outperformed Merlin for the rare variants but not for the common ones (Supplemental Fig. S1). Zooming in on the (0,0.05) MAF interval, Supplemental Figure S2 shows how the different methods behave for rare variants and how the clear improvement of population-based approaches starts to be apparent between [0.03,0.04) and [0.04,0.05).

Imputation of the more common variants was better with population-based than pedigree-based methods (MAF in [0.05,0.5]) (Fig. 1A and 1B). The better performance for population-based imputation is more evident in the EUR compared to the AFR sample, for which GIGI and Merlin were not substantially outperformed for the common variants. Within population-based approaches, duoHMM for phasing followed by minimac for imputation (duoHMM+minimac) and SHAPEIT+minimac were the best combinations for the EUR across all MAF intervals (Fig. 1A). In the AFR, IMPUTE+IMPUTE performed as well as those two previous combinations (Fig. 1B). To check if the differences between population-based approaches were simply due to sampling variation, we re-generated 100 new genetic datasets using the random selection strategy and we performed imputation combinations MaCH+minimac, SHAPEIT+minimac, duoHMM+minimac, IMPUTE+IMPUTE, SHAPEIT+IMPUTE, and duoHMM+IMPUTE for EUR and AFR. The same differences and trends were observed again, which suggests that these difference are systematic (Supplemental Fig. S3). Note that Beagle had the lowest imputation accuracy. Moreover, for all population-based methods, our results showed that

imputation accuracy was greater for EUR than AFR (Fig. 1A vs Fig. 1B). Notably, the hybrid approach Ped_Pop, which combined both family- and population-based strengths, had the greatest performance for EUR and AFR in case of both rare and common variants (Fig. 1). Supplemental Figure S4 shows an example of GIGI, duoHMM+minimac, and Ped_Pop performances for imputing two arbitrary chosen SNPs (one with MAF=0.006 and one with MAF=0.46) in a pedigree of 24 subjects of whom 5 subjects were fully observed. For the SNP with low allele frequency (MAF=0.006), GIGI perfectly imputed all 10 genotypes with at least one copy of the minor allele, while duoHMM+minimac could not impute any of them. For the SNP with high allele frequency (MAF=0.46), duoHMM+minimac accurately imputed all 14 genotypes with at least one copy of the minor allele, while GIGI could impute 11. In both cases, Ped_Pop accurately imputed genotypes with at least one copy of the minor allele.

Eagle vs SHAPEIT for phasing

Figure 2A and 2B show that the use of SHAPEIT or duoHMM for phasing was more successful at yielding high-quality imputed data than the use of Eagle. When phasing was done with Eagle, imputation accuracy dropped, especially with minimac. This pattern was apparent for both the EUR (Fig. 2A) and AFR (Fig. 2B) data and for all MAF intervals.

IQS vs R^2 and Statistical Power as a baseline

IQS and R^2 were not always in agreement in summarizing imputation accuracy (Fig. 1A and 1B vs Fig. 1C and 1D, respectively). Similar conclusions reached using either IQS or R^2 were that: (1) Ped_Pop performed best overall, (2) family-based approaches had higher values for rare variants, whereas population-

based approaches had higher values for common variants, and finally (3) both IQS and R^2 yielded a better performance in the EUR data compared to AFR data. IQS differed from R^2, with: (1) the approaches involving IMPUTE for imputation were better than the other approaches for both EUR and AFR (Fig. 1C vs Fig. 1D), (2) the approaches involving minimac for imputation appeared to be less accurate, (3) MaCH+minimac, and not Beagle, appeared to be the worst approach, and finally, (4) GIGI was slightly outperformed by Merlin for common variants. Altogether, the IQS values were smaller than the R^2 values for all approaches as can be seen for instance for GIGI and duoHMM+minimac in Supplemental Figure S5. This does not necessarily mean that R^2 values overestimate imputation accuracy as claimed by (Ramnarine et al. 2015).

To determine the metric that appears to be better for imputation accuracy, we computed the statistical power of association analysis and used this as a baseline to identify the imputation approach that leads to the highest statistical power. We focused on the cases where a disagreement between IQS and R^2 was observed. For these cases we compared the corresponding powers. Type 1 error rates were well controlled for all observed genotype and imputed dosages for the Random and GIGI-Pick selection strategies for α=0.05, 0.01, and 0.001 with the exception of evidence for slight conservatism for the population-based imputation programs in the bin with the lowest MAF (Supplemental Table S1-S6). Power results for the random selection strategy for α=0.05 are shown in Table 1, and the remaining tables are shown in the Supplementary Material (Supplemental Table S7-S11).

The power results showed that SHAPEIT+minimac and duoHMM+minimac were better than SHAPEIT+IMPUTE and duoHMM+IMPUTE. They also showed that the power of Beagle was slightly lower than MaCH+minimac. Finally, the power

results showed that GIGI had at least similar power compared to Merlin. All these results are in agreement with the R^2 values. In addition, these conclusions were most obvious for common variants. For rare variants, the trend was similar despite the small values of both IQS and R^2.

Not surprisingly, in the association analysis, the direct power estimates for Ped_Pop were the largest among all the imputation approaches (Table 1). Power to detect association was smaller in AFR than EUR for all population-based imputation. On the other hand, power achieved through the use of GIGI for imputation was similar in EUR and AFR. Finally, power for imputation involving the GIGI-Pick selection was higher than for the random selection, for all α levels examined (Table 1 vs Supplemental Table S9, Supplemental Table S7 vs Supplemental Table S10, and Supplemental Table S8 vs Supplemental Table S11). Again, all these results are congruent with the imputation R^2 accuracy values.

Subject selection strategies

The four subject selection strategies we compared performed differently depending on the phasing+imputation combination used. In Figure 3A and 3B, we show the mean R^2 for three imputation approaches in EUR and AFR: GIGI (best of family-based imputation), duoHMM+minimac (best of population-based imputation), and Ped_Pop (combination of GIGI and duoHMM+minimac). As expected, PRIMUS did not perform well for any imputation method; It was even worse than the Random selection strategy. The reason behind this result is that in family data, subjects with vertical and horizontal relationships in pedigrees (e.g., parent-offspring, siblings) improve the phasing process and therefore the imputation. By repeating a random

selection on all simulated datasets, such relationships were present more often than with PRIMUS, which forces the selection of a set of maximally unrelated subjects.

The selection of subjects using GIGI-Pick led to better imputation accuracy than ExomePicks for both GIGI and Ped_Pop using the R^2 criterion. Note that when imputing rare variants, ExomePicks selection led to a better imputation accuracy than GIGI-Pick for SHAPEIT+minimac, SHAPEIT+IMPUTE, MaCH+minimac, duoHMM+minimac, duoHMM+IMPUTE, and Beagle (Supplemental Fig. S6 part A). However, imputation accuracies of these approaches remained much smaller than GIGI's accuracy. For common variants, the same trend was observed, but with smaller differences between GIGI-Pick and ExomePicks (Supplemental Fig. S6 part B). All the conclusions above were the same for AFR.

DISCUSSION

Our study is the first to address several challenges faced in imputation in family data and evaluate the performance of many available family- and population-based tools in GWAS analysis on the same family data of both European and African ancestry, and provides guidelines for future studies. We showed that family-based imputation outperforms population-based imputation for rare variants. For common variants, population-based approaches are expected to be better except when the amount of LD between SNPs is low. This explains why population-based imputation yielded more accurate results on data from European than African samples (mean LD computed within non-overlapping windows of 100 SNPs in EUR was $R^2 = 0.032$ vs 0.02 in AFR). It is worth noting, however, that family-based imputation was not affected by the ancestry differences because this approach relies on IBD rather than LD.

Of the population-based tools, the combination of SHAPEIT (v2) with the duoHMM feature for phasing and minimac (v3) for imputation outperformed all other combinations. For family-based imputation, we found that GIGI outperformed Merlin for rare variants but not for the common ones. This is because Merlin uses LD information by incorporating the fastPHASE algorithm (Scheet and Stephens 2006) when IBD information cannot determine the phase. But because population-based approaches outperformed both GIGI and Merlin for common variants, GIGI would be preferable to Merlin from both an accuracy and computational point of view. Merlin presented great computational challenges even for small sets of SNPs (<11,000 SNPs), and also required splitting pedigrees into smaller sub-pedigrees. Therefore, running Merlin on a GWAS or Whole Genome Sequence level would be impractical. The solution that worked best across the full range of allele frequencies was that implemented in Ped_Pop, which combines the strengths of both family- and population-based imputation. By considering both EUR and AFR populations, we combined the results of GIGI and duoHMM+minimac (see Supplementary Methods for further details). This approach led to the greatest imputation accuracy and largest association power for both rare and common variants.

The accuracy measure used to evaluate the imputation performance is of great importance. The accuracy results computed with R^2 were concordant with the power of association analysis, contrary to the results computed with IQS, which provided an inconsistent predictor of statistical power. Overall, our results suggest that IQS underestimates imputation accuracy, but R^2 better defines the imputation accuracy and should continue to be used to evaluate imputation accuracy for both rare and common variants in future studies. In addition, R^2 has a direct relationship with power in association studies. For example, in the case of imputing one SNP on

1000 unrelated subjects, an imputation accuracy of R^2=0.8 means that to achieve the same power when using perfect genotypes, the sample size must be $\frac{1}{0.8} = 1.25$ times higher.

The choice of subjects to be sequenced from pedigrees has a large impact on the imputation performance and therefore on the association results. A careful and optimal choice of selected subjects at the study design level would likely increase the imputation accuracy and therefore the power of association tests. In our study, we evaluated four strategies to select subjects for sequencing and analysis with an association test on the imputation accuracy: Random selection strategy, GIGI-Pick, ExomePicks, and PRIMUS. Our results showed that if one is interested in rare variant imputation, the selection of subjects should be done using GIGI-Pick and imputation should be done using GIGI. If one is interested in both rare and common variant imputation, the selection of subjects and imputation should be done using GIGI-Pick and Ped_Pop, respectively.

Here we did not evaluate pedigrees chosen through selective phenotypic ascertainment, but we believe that our general conclusions should still hold under such ascertainment. Pedigree ascertainment through selective phenotypes can only affect power to detect true associations. Power is a function of the number of subjects with high-quality imputed genotypes, particularly in individuals who are not closely related. Of the available tools for subject selection, GIGI-pick, with its genome-wide option used here, already does the best job of balancing the conflicting needs of re-sequencing inherited copies of genomic regions for phasing, while also distributing the sequencing across individuals without shared inheritance. More importantly, GIGI-pick is also currently the only subject selection tool that can also

take into account realized IBD in a region of interest for subject selection. Such a region may be determined with pedigrees selected through members' phenotypic status followed by genotyping with a low-cost SNP array and linkage analysis. As was shown previously, this option can have a marked positive effect on the overall number of high-quality, imputed, relatively-independent genotypes in the region (Cheung et al. 2014), and thus the power of an association test.

In our study, we did not evaluate the performance of phasing approaches but only compared the imputation accuracy with respect to the different phasing methods used. Two of the best population-based phasing algorithms are Eagle and SHAPEIT. In our simulated data, we observed that SHAPEIT outperformed Eagle, which was also observed in (Herzig et al. 2018), contrary to what was observed by Eagle's authors (Loh et al. 2016). For family-based phasing, we used gl_auto to phase the set of sparse markers. Like GIGI, gl_auto does not use LD for phasing and imputation. Future incorporation of such useful information into GIGI will most certainly improve its performance, and may, ultimately, outperform the Ped_Pop approach. Until this happens, the results of our study suggest that the approach of Ped_Pop (https://bioinformatics.qcri.org/ped_pop) provides a pragmatic approach that combines both pedigree- and population-based strengths.

METHODS

Simulated Data

Genetic Data

We simulated sequence data on a collection of 20 extended pedigrees consisting of 1200 total subjects with sizes ranging from 10 to 174 subjects and with median and

mean sizes of 47 and 60 subjects, respectively. The sibship sizes ranged from 1 to 11 siblings, with median and mean sizes of 1 and 1.86 sibling, respectively. The number of generations ranged from 3 to 9, with median and mean sizes of 8 and 6.65 generations, respectively. The pedigree, generation, and sibship sizes were modelled on those of real pedigrees (EM Wijsman, pers. comm.). The generated pedigrees and the simulated sequenced data are available online (see Data Access section). We used the same simulation strategy used in a previous study (Saad and Wijsman 2013) to obtain 100 semi-realistic sequence datasets that mimic the LD structure and MAF spectrum of the 1000 Genomes Project for European and African ancestries (we will call them EUR and AFR throughout). Briefly, for each ancestry we simulated 20,000 haplotypes for a region of approximately 6 Megabase (Mb) pairs on Chromosome 22 (Genome built GRCh37, 26443384 - 32049917) using HAPGEN (Su et al. 2011). From the pool of 20,000 haplotypes, we started by randomly selecting haplotypes, without replacement, for the unrelated founders. Then, we dropped the haplotypes from the founders down through the 3-5 generations in pedigrees using a recombination rate of 1% per centiMorgan (cM) per meiosis under the assumption that 1 cM is 1000 kilobase (kb) pairs. We used the same pedigrees for both EUR and AFR. The number of SNPs in the EUR and AFR 1000 Genomes data were 8,954 and 11,891, respectively. For both EUR and AFR, 500 SNPs (~5% of the total number of SNPs) were randomly selected to form the GWAS list of SNPs. The whole process was repeated 100 times to finally obtain 100 simulated datasets for the two ancestries. The distributions of MAFs for both EUR and AFR are shown in Supplemental Figure S7. Note that reference assembly for sequence read alignment (e.g., GRCh37 or GRCh38) has no significant impact on the linkage

disequilibrium, or IBD. Therefore, our conclusions will not be affected by the use of GRCh37 reference assembly is used.

Quantitative traits

We simulated quantitative traits by sampling from two models: $(H_0)\ Y = \epsilon$, $(H_a)\ Y = \beta_j X_j + \epsilon$. In both models, ϵ follows a multivariate normal distribution $N(0, \Sigma)$ where $\Sigma = h^2\Phi + (1 - h^2)I$, Φ is the matrix of twice the kinship coefficient between pairs of subjects, I is the identity matrix, and h^2 is the heritability, set to 0.5 by setting the total variance to 2 and the genetic variance due to polygenic effects to 1. In model H_a, X_j is the variable of known genotypes of the *j*th SNP coded as 0, 1, or 2 copies of the minor alleles, β_j is the effect size of the corresponding SNP and calculated as $\beta_j = \sqrt{\frac{v_a}{2 \times MAF_j \times (1 - MAF_j)}}$ where MAF_j is the minor allele frequency and v_a, set to 0.01, is the additive variance of SNP *j*.

For each SNP and each genetic dataset, we simulated 10 quantitative traits for the H_0 model (hypothesis of no association) to compute the type 1 error rate and 10 quantitative traits for the H_a model (hypothesis of association) to compute the statistical power rate. Since there are 100 genetic datasets and 10 quantitative traits for each dataset, we calculated the rate for each SNP as a proportion of these 1,000 datasets. We computed both rates for each SNP as the proportion of replicates with a p-value smaller than a given α and then averaged all SNP rates within the following MAF bins: (0,0.01), [0.01,0.05), [0.05,0.1), [0.1,0.2), [0.2,0.3), [0.3,0.4), and [0.4,0.5]. This process yielded the following number of tests within the respective MAF bins: 2,497,900, 1,708,680, 955,460, 911,850, 766,810, 652,040, 461,260 for EUR and 2,621,320, 3,333,930, 1,646,340, 1,384,900, 840,620, 582,290, and

481,600 for AFR. We used three values of α: 0.05, 0.01, and 0.001. Note that more stringent thresholds could be applied, which will likely yield lower power. For the sake of comparing the different imputation approaches and not evaluating the statistical power per se, the α thresholds we used would be enough to reach our main conclusions.

Imputation and Association Analyses

In all approaches evaluated, imputation relies on inferring missing genotypes in study subjects using a reference dataset of fully sequenced subjects. The study subjects are genotyped on a sparse map of SNPs. All imputation methods are based on two steps: phasing and imputation. For family-based imputation, the reference dataset only needs to contain subjects from the pedigrees under study. For population-based imputation methods on pedigree data, the reference dataset was formed by combining all sequenced subjects across pedigrees. In all the imputation analyses, we selected subjects based on the pedigree structure from which we simulated sequence data in order to obtain the reference dataset. In all tools we compared, the default parameters suggested by their respective developers were used.

Selection of Reference Dataset

An optimal selection choice, in the context of trait mapping, depends on several factors: the pedigree structure, the availability of phenotype, the severity of disease, and the availability of good quality DNA. In our simulation study, we only used the pedigree structure to select subjects. We compared four selection strategies: GIGI-Pick (Cheung et al. 2014), ExomePicks ([http://genome.sph.umich.edu/wiki/ExomePicks]), PRIMUS (Staples et al. 2013;

Staples et al. 2014), and Random selection. In all imputation analyses, 20% of subjects were selected for sequencing from each of the 20 analyzed pedigrees. This proportion was used across all pedigrees, resulting in 240 subjects. For all four selection strategies, the set of 240 subjects obtained was used for EUR and AFR imputation because the pedigree structure was the same. A detailed description about the subject selection procedures is provided in the Supplementary material.

In the following section, we list the proposed phasing and imputation algorithms that were assessed in our study, and we briefly summarize them in Table 2. A more thorough description can be found in the Supplementary material.

Family-based design imputation

We used pedigree-based imputation computer programs GIGI (Cheung et al. 2013) and Merlin (Abecasis et al. 2002). GIGI requires a pre-phasing step where the IBD is computed by the program gl_auto, implemented in MORGAN (Thompson 2011). On the other hand, Merlin performs this step internally. Running Merlin on large pedigrees requires too much memory, which can be predicted by the number of bits in the pedigree (Kruglyak et al. 1996), $bits = 2n - f$, where n is the number of non-founders and f is the number of founders. The pedigrees we are using ranged from 5 to 165 bits. In our data, imputation of pedigrees with 19 bits required 110 Gb of memory to impute 8,954 SNPs. To get results from Merlin, we split large pedigrees into small computable sub-pedigrees defined with a maximum of 19 bits.

Automated methods for subdividing the pedigree structures exist, such as PedCut (Liu et al. 2008) and PedStr (Kirichenko et al. 2009). However, we found them

unsatisfactory, resulting in excessively small sub-pedigrees without the flexibility to ensure that at least some sequenced subjects are in each sub-pedigree. We instead opted to manually construct the sub-pedigrees to include a greater number of subjects in each sub-pedigree and to be close to the upper limit of 19 bits, while including both vertical and horizontal relationships (Grand-parents, Parents, Offspring, Siblings, etc.). We often included the same individuals in several of the sub-pedigrees to be closer to 19 bits. In these cases, we retained the imputation results for these individuals from the largest sub-pedigree when combining the results.

Population-based design imputation

For phasing, we used the following programs: SHAPEIT (v2) (Delaneau et al. 2012), DuoHMM (O'Connell et al. 2014), IMPUTE (Howie et al. 2009), MaCH (Li et al. 2010), Beagle (Browning and Browning 2007), and Eagle (Loh et al. 2016). For imputation, we used: IMPUTE, minimac (v3) (Fuchsberger et al. 2015), and Beagle (v4.1) (Browning and Browning 2016). The versions of the tools we used are the latest only from an algorithmic point of view. For example, we used IMPUTE (v2), which has a newer version (v4, see https://jmarchini.org/impute-4/). However, IMPUTE phasing and imputation algorithms did not change in the new version. The new phasing and imputation versions that are being proposed are mainly aiming at handling larger datasets.

Combination of Family- and Population-based imputation

To benefit from both IBD and LD information, we combined family- and population-based imputation results using Ped_Pop (https://bioinformatics.qcri.org/ped_pop, (Saad and Wijsman 2014). Ped_Pop can combine imputation results from any

family- and population-based methods. In this study, the family- and population-based approaches with overall best performance were combined (see Supplemental Methods for further details).

Imputation accuracy measures

Several measures to compute imputation accuracy have been proposed. Examples are: Pearson's correlation R^2, the Imputation Quality Score (IQS (Lin et al. 2010) based on the Kappa statistic), and the Concordance Rate (CR). The performance of these metrics depends on the MAF of imputed SNPs. For example, CR over-estimates the imputation accuracy for rare variants. Moreover, it also has been claimed that R^2 overestimates imputation accuracy for rare variants (Lin et al. 2010). However, the claim was based on comparing R^2 and IQS values, without reference to a baseline to decide which metric is better, while knowing that these values cannot be compared directly. Here we compared R^2 and IQS and their behavior with respect to the statistical power of the association test to determine the best metric. The imputation that leads to the highest association power is the best imputation approach in this context. The use of type 1 error rates does not allow comparison of R^2 and IQS because they are expected to be similar and close to the α threshold we set (e.g. $\alpha = 0.05$) in all scenarios. Note that we chose not to use the CR in our comparison because of its known limitations.

Association Testing

We constructed the following linear mixed model to test for association between SNPs and quantitative traits: $Y = \beta_j X_j + \epsilon$ where X_j is the variable representing the genotype of the j^{th} SNP coded as the number of copies of the minor

alleles, β_j is the corresponding regression coefficient, and $\epsilon \sim N(\beta_j X_j, \sigma_g^2 \Phi + \sigma_e^2 I)$ where Φ is the matrix of twice the kinship coefficient between pairs of subjects, σ_e^2 is the residual variance and σ_g^2 is the polygenic variance. Association tests were performed twice (under the null and alternative hypotheses) on all SNPs except the first and the last 500 SNPs, where imputation results for population-based imputation are poor due to lack of buffer downstream and upstream. The association test was performed on the data of true genotypes and also on the data of all imputation combinations without applying any poor-quality imputation filtering. The number of tests we performed were: 2,497,900, 1,708,680, 955,460, 911,850, 766,810, 652,040, 461,260 for EUR and 2,621,320, 3,333,930, 1,646,340, 1,384,900, 840,620, 582,290, and 481,600 for AFR for the respective intervals [0,0.01), [0.01,0.05), [0.05,0.1), [0.1,0.2), [0.2,0.3), [0.3,0.4), and [0.4,0.5]**.** These analyses were conducted for the observed genotype dosages and the imputed ones for the Random and GIGI-Pick selection strategies using the 'lmekin' function in the 'kinship2' R package (https://cran.r-project.org/web/packages/kinship2/index.html).

Finally, in our study, we evaluated several combinations of phasing and imputation methods in family- and population-based designs in both EUR and AFR. These combinations (Phasing+Imputation) were: MaCH+minimac, SHAPEIT+minimac, duoHMM+minimac, Eagle+minimac, IMPUTE+IMPUTE, SHAPEIT+IMPUTE, duoHMM+IMPUTE, Eagle+IMPUTE, Beagle+Beagle, GIGI (using gl_auto), Merlin, and Ped_Pop (combining GIGI and duoHMM+minimac). All approaches were performed for the four selection strategies: GIGI-Pick, ExomePicks, PRIMUS, and Random selection. Note that Merlin was run for the sub-pedigrees only with the GIGI-Pick and Random selection strategies. To have a fair

comparison with GIGI, GIGI was run on the same sub-pedigrees and also on the full large pedigrees.

Data Access

Simulated genetic sequence data and pedigree structures used in our study are available at DOI: 10.5281/zenodo.1485557 and on our site at https://bioinformatics.qcri.org/IRD.

ACKNOWLEDGEMENTS

A great part of the computation used in this study was performed on the Raad2 cluster at Texas A&M University in Qatar.

Partially supported by grant support from the US National Institutes of Health, R01 HD088431, R37 GM046255, R24 OD021324, and P50 AG005136.

DISCLOSURE DECLARATION

All authors declare no conflict of interest

Figures and Tables

Figure 1: Mean correlation R^2 and IQS between true and imputed genotypes for all approaches: (A) R^2 for EUR and (B) R^2 for AFR (C) IQS for EUR, and (D) IQS for AFR, using the random selection strategy. The first/second of a pair of programs in the key indicates phasing/imputation functions. Computation of the mean of R^2 and IQS is based on all 100 genetic datasets with a sample size of 960 subjects, each having 7,954 SNPs for EUR and 10,891 SNPs for AFR.

Figure 2: Mean correlation R^2 between true and imputed genotypes for SHAPEIT+minimac, duoHMM+minimac, SHAPEIT+IMPUTE, duoHMM+IMPUTE, Eagle+minimac, and Eagle+IMPUTE for (A) EUR and (B) AFR using the random selection strategy. The first/second of a pair of programs indicates phasing/imputation functions. Computation of the mean of R^2 is based on all 100

genetic datasets with a sample size of 960 subjects, each having 7,954 SNPs for EUR and 10,891 SNPs for AFR.

Figure 3: Mean correlation R^2 between true and imputed genotypes for the four selection strategies (GIGI-Pick, ExomePicks, PRIMUS, and Random selection) for Ped_Pop, GIGI, and duoHMM+minimac. (A) EUR and (B) AFR. Computation of the mean of R^2 is based on all 100 genetic datasets with a sample size of 960 subjects, each having 7,954 SNPs for EUR and 10,891 SNPs for AFR.

Table 1: Table 1: Power of association tests performed in European and African data for the different combination of phasing+imputation approaches using the random selection strategy for α=0.05. The sample size was 960.

Table 2: Phasing and Imputation approach summary.

REFERENCES

Abecasis GR, Cherny SS, Cookson WO, Cardon LR. 2002. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics* **30**: 97-101.

Al-Tassan NA, Whiffin N, Hosking FJ, Palles C, Farrington SM, Dobbins SE, Harris R, Gorman M, Tenesa A, Meyer BF et al. 2015. A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer. *Sci Rep* **5**: 10442.

Browning BL, Browning SR. 2016. Genotype Imputation with Millions of Reference Samples. *Am J Hum Genet* **98**: 116-126.

Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics* **81**: 1084-1097.

Burdick JT, Chen WM, Abecasis GR, Cheung VG. 2006. In silico method for inferring genotypes in pedigrees. *Nature Genetics* **38**: 1002-1004.

Chen MH, Huang J, Chen WM, Larson MG, Fox CS, Vasan RS, Seshadri S, O'Donnell CJ, Yang Q. 2012. Using family-based imputation in genome-wide association studies with large complex pedigrees: the Framingham Heart Study. *PLoS One* **7**: e51589.

Cheung CYK, Blue EM, Wijsman EM. 2014. A statistical framework to guide sequencing choices in pedigress. *American Journal of Human Genetics* **94**: 257-267.

Cheung CYK, Thompson EA, Wijsman EM. 2013. GIGI: An approach to effective imputation of dense genotypes on large pedigrees. *American Journal of Human Genetics* **92**: 504-516.

Delaneau O, Marchini J, Zagury JF. 2012. A linear complexity phasing method for thousands of genomes. *Nature Methods* **9**: 179-181.

Evangelou E, Maraganore DM, Ioannidis JP. 2007. Meta-analysis in genome-wide association datasets: strategies and application in Parkinson disease. *PLoS One* **2**: e196.

Fuchsberger C, Abecasis GR, Hinds DA. 2015. minimac2: faster genotype imputation. *Bioinformatics* **31**: 782-784.

Herzig AF, Nutile T, Babron MC, Ciullo M, Bellenguez C, Leutenegger AL. 2018. Strategies for phasing and imputation in a population isolate. *Genet Epidemiol* **42**: 201-213.

Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. 2012. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet* **44**: 955-959.

Howie BN, Donnelly P, Marchini J. 2009. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**: e1000529.

International Parkinson's Disease Genomics Consortium (IPDGC) and Wellcome Trust Case Control Consortium 2 (WTCCC2) 2011. A two-stage meta-analysis identifies several new loci for Parkinson's disease. *PLoS Genet* **7**: e1002142.

Kirichenko AV, Belonogova NM, Aulchenko YS, Axenovich TI. 2009. PedStr software for cutting large pedigrees for haplotyping, IBD computation and multipoint linkage analysis. *Ann Hum Genet* **73**: 527-531.

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. 1996. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* **58**: 1347-1363.

Lander ES, Green P. 1987. Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci U S A* **84**: 2363-2367.

Lent S, Deng X, Cupples LA, Lunetta KL, Liu CT, Zhou Y. 2016. Imputing rare variants in families using a two-stage approach. *BMC Proc* **10**: 209-214.

Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. 2010. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* **34**: 816-834.

Lin P, Hartz SM, Zhang Z, Saccone SF, Wang J, Tischfield JA, Edenberg HJ, Kramer JR, A MG, Bierut LJ et al. 2010. A new statistic to evaluate imputation reliability. *PLoS One* **5**: e9697.

Liu F, Kirichenko A, Axenovich TI, van Duijn CM, Aulchenko YS. 2008. An approach for cutting large and complex pedigrees for linkage analysis. *European Journal of Human Genetics* **16**: 854-860.

Livne OE, Han L, Alkorta-Aranburu G, Wentworth-Sheilds W, Abney M, Ober C, Nicolae DL. 2015. PRIMAL: Fast and accurate pedigree-based imputation from sequence data in a founder population. *PLoS Comput Biol* **11**: e1004139.

Loh PR, Danecek P, Palamara PF, Fuchsberger C, Y AR, H KF, Schoenherr S, Forer L, McCarthy S, Abecasis GR et al. 2016. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* **48**: 1443-1448.

Maher B. 2008. Personal genomes: The case of the missing heritability. *Nature* **456**: 18-21.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A et al. 2009. Finding the missing heritability of complex diseases. *Nature* **461**: 747-753.

Marchini J, Howie B. 2010. Genotype imputation for genome-wide association studies. *Nat Rev Genet* **11**: 499-511.

Marigorta UM, Rodriguez JA, Gibson G, Navarro A. 2018. Replicability and Prediction: Lessons and Challenges from GWAS. *Trends Genet* **34**: 504-517.

McCarthy S Das S Kretzschmar W Delaneau O Wood AR Teumer A Kang HM Fuchsberger C Danecek P Sharp K et al. 2016. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**: 1279-1283.

Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, DeStefano AL, Kara E, Bras J, Sharma M et al. 2014. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet* **46**: 989-993.

Nalls MA, Plagnol V, Hernandez DG, Sharma M, Sheerin UM, Saad M, Simon-Sanchez J, Schulte C, Lesage S, Sveinbjornsdottir S et al. 2011. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* **377**: 641-649.

Nettelblad C. 2012. Inferring haplotypes and parental genotypes in larger full sib-ships and other pedigrees with missing or erroneous genotype data. *BMC Genet* **13**: 85.

O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, Traglia M, Huang J, Huffman JE, Rudan I et al. 2014. A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet* **10**: e1004234.

Ramnarine S, Zhang J, Chen LS, Culverhouse R, Duan W, Hancock DB, Hartz SM, Johnson EO, Olfson E, Schwantes-An TH et al. 2015. When Does Choice of Accuracy Measure Alter Imputation Accuracy Assessments? *PLoS One* **10**: e0137601.

Saad M, Nato AQ, Jr., Grimson FL, Lewis SM, Brown LA, Blue EM, Thornton TA, Thompson EA, Wijsman EM. 2016. Identity-by-descent estimation with population- and pedigree-based imputation in admixed family data. *BMC Proc* **10**: 295-301.

Saad M, Wijsman EM. 2013. Power of Family-Based Association Designs to Detect Rare Variants in Large Pedigrees Using Imputed Genotypes. *Genet Epidemiol* doi:10.1002/gepi.21776.

Saad M, Wijsman EM. 2014. Combining family- and population-based imputation data for association analysis of rare and common variants in large pedigrees. *Genet Epidemiol* **38**: 579-590.

Scheet P, Stephens M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet* **78**: 629-644.

Siddiq A Couch FJ Chen GK Lindstrom S Eccles D Millikan RC Michailidou K Stram DO Beckmann L Rhie SK et al. 2012. A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Hum Mol Genet* **21**: 5373-5384.

Sniekers S, Stringer S, Watanabe K, Jansen PR, Coleman JRI, Krapohl E, Taskesen E, Hammerschlag AR, Okbay A, Zabaneh D et al. 2017. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat Genet* **49**: 1107-1112.

Staples J, Nickerson DA, Below JE. 2013. Utilizing graph theory to select the largest set of unrelated individuals for genetic analysis. *Genet Epidemiol* **37**: 136-141.

Staples J, Qiao D, Cho MH, Silverman EK, University of Washington Center for Mendelian G, Nickerson DA, Below JE. 2014. PRIMUS: rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. *Am J Hum Genet* **95**: 553-564.

Su Z, Marchini J, Donnelly P. 2011. HAPGEN2: simulation of multiple disease SNPs. *Bioinformatics* **27**: 2304-2305.

The 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature* **526**: 68-74.

The UK10K Consortium, Walter K, Min JL, Huang J, Crooks L, Memari Y, McCarthy S, Perry JR, Xu C, Futema M et al. 2015. The UK10K project identifies rare variants in health and disease. *Nature* **526**: 82-90.

Thompson EA. 2011. The structure of genetic linkage data: from LIPED to 1M SNPs. *Human Heredity* **71**: 86-96.

Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J. 2017. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet* **101**: 5-22.

Wang WYS, Barratt BJ, Clayton DG, Todd JA. 2005. Genome-wide association studies: Theoretical and practical concerns. *Nature Reviews Genetics* **6**: 109-118.

Wijsman EM. 2012. The role of large pedigrees in an era of high-throughput sequencing. *Human Genetics* **131**: 1555-1563.

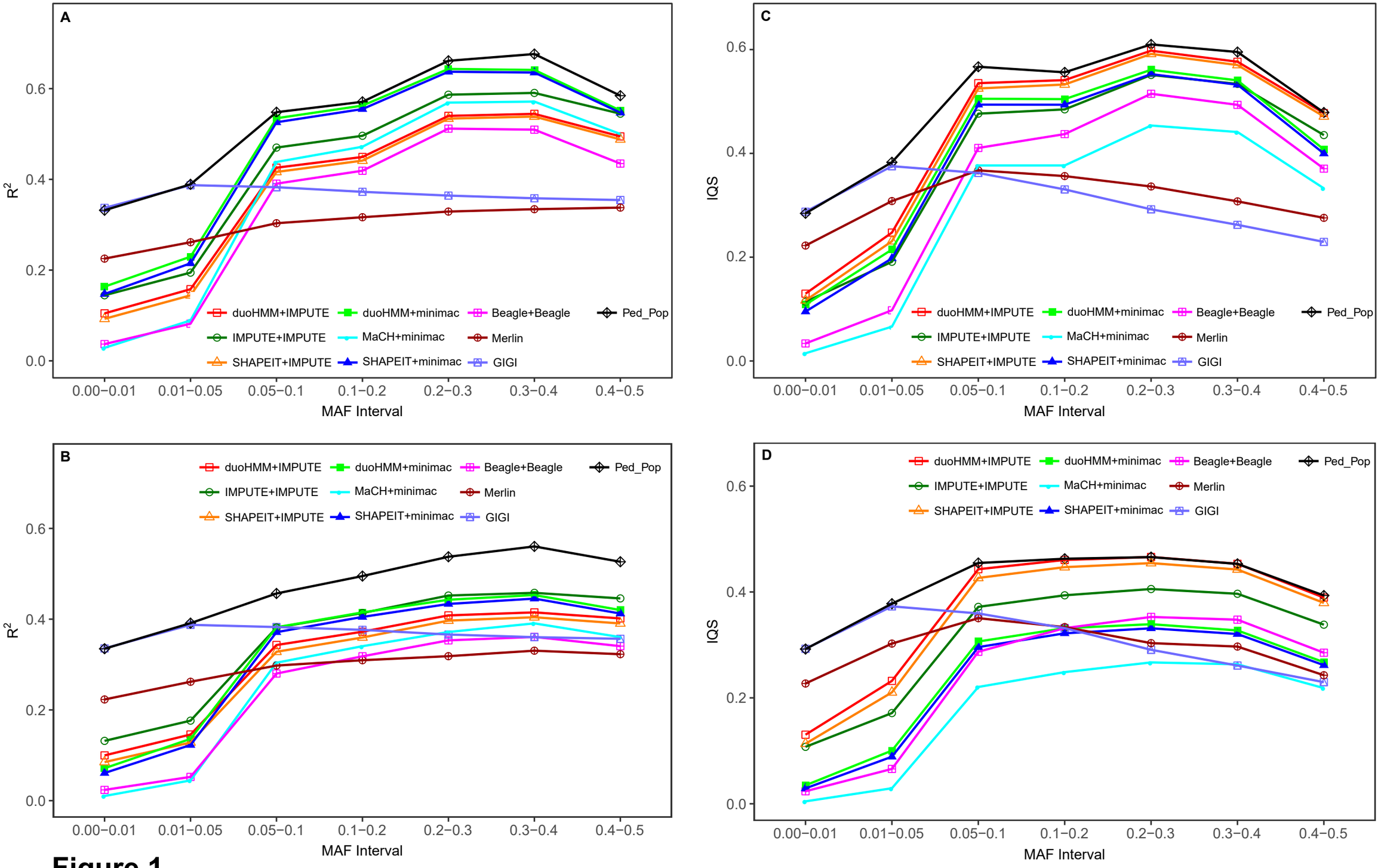
A
B
C
D
R^2
IQS
0.0
0.2
0.4
0.6
0.00−0.01
0.01−0.05
0.05−0.1
0.1−0.2
0.2−0.3
0.3−0.4
0.4−0.5
MAF Interval
duoHMM+IMPUTE
duoHMM+minimac
Beagle+Beagle
Ped_Pop
IMPUTE+IMPUTE
MaCH+minimac
Merlin
SHAPEIT+IMPUTE
SHAPEIT+minimac
GIGI

Figure 1

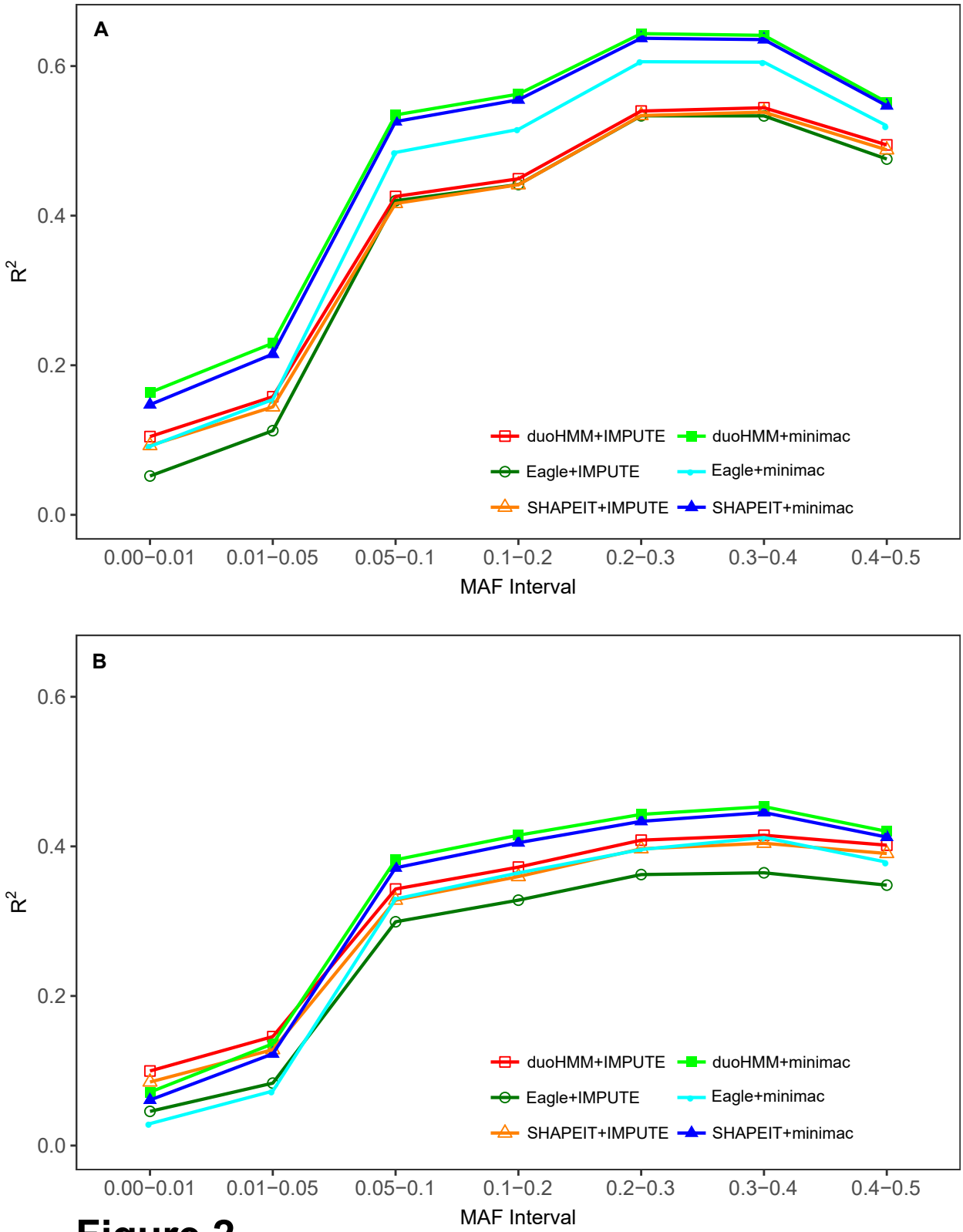
A
B
R^2
0.0
0.2
0.4
0.6
0.00−0.01
0.01−0.05
0.05−0.1
0.1−0.2
0.2−0.3
0.3−0.4
0.4−0.5
MAF Interval
duoHMM+IMPUTE
duoHMM+minimac
Eagle+IMPUTE
Eagle+minimac
SHAPEIT+IMPUTE
SHAPEIT+minimac

Figure 2

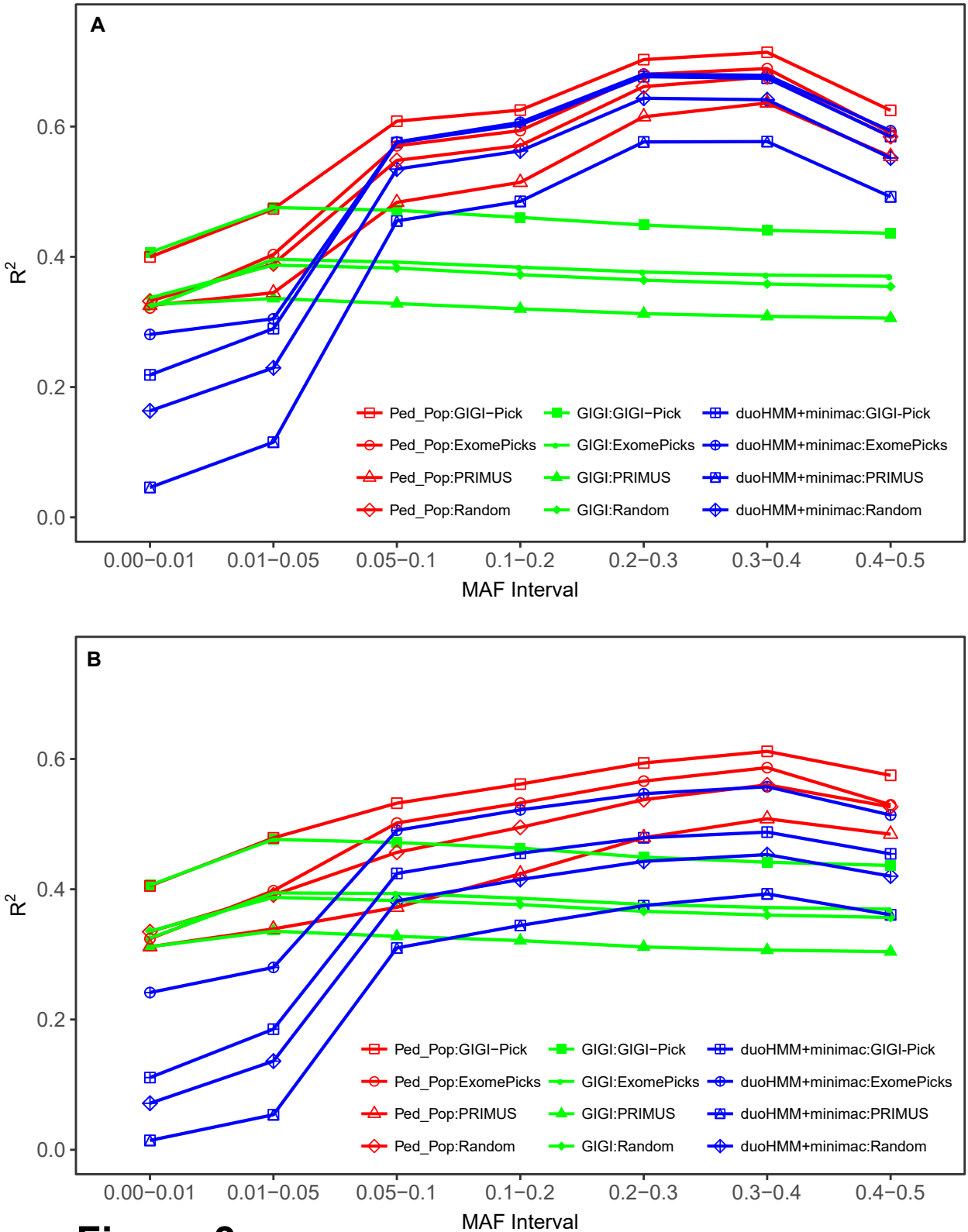
A
B
R^2
0.0
0.2
0.4
0.6
0.00−0.01
0.01−0.05
0.05−0.1
0.1−0.2
0.2−0.3
0.3−0.4
0.4−0.5
MAF Interval
Ped_Pop:GIGI−Pick
Ped_Pop:ExomePicks
Ped_Pop:PRIMUS
Ped_Pop:Random
GIGI:GIGI−Pick
GIGI:ExomePicks
GIGI:PRIMUS
GIGI:Random
duoHMM+minimac:GIGI-Pick
duoHMM+minimac:ExomePicks
duoHMM+minimac:PRIMUS
duoHMM+minimac:Random

Figure 3

Table 1: Power of association tests performed in European and African data for the different combination of phasing+imputation approaches using the random selection strategy for α=0.05.

Method	MAF bin [0,0.01)		[0.01,0.05)		[0.05,0.1)		[0.1,0.2)		[0.2,0.3)		[0.3,0.4)		[0.4,0.5]	
	EUR	**AFR**	**EUR**	**AFR**	**EUR**	**AFR**	**EUR**	**AFR**	**EUR**	**AFR**	**EUR**	**AFR**	**EUR**	**AFR**
Observed genotypes	0.546	0.616	0.807	0.813	0.806	0.812	0.805	0.812	0.800	0.811	0.806	0.810	0.808	0.812
Ped_Pop	0.296	0.314	**0.403**	**0.403**	**0.513**	**0.427**	**0.529**	**0.433**	**0.591**	**0.448**	**0.606**	**0.471**	**0.541**	**0.470**
GIGI	**0.301**	**0.315**	**0.403**	0.400	0.393	0.380	0.383	0.359	0.372	0.345	0.370	0.356	0.369	0.368
Merlin	0.216	0.225	0.291	0.297	0.343	0.347	0.355	0.354	0.358	0.354	0.354	0.362	0.349	0.340
MaCH+minimac	0.052	0.045	0.122	0.086	0.408	0.301	0.441	0.334	0.517	0.363	0.525	0.379	0.476	0.360
SHAPEIT+minimac	0.118	0.079	0.240	0.159	0.490	0.367	0.515	0.396	0.573	0.422	0.576	0.430	0.516	0.406
duoHMM+minimac	0.127	0.086	0.253	0.173	0.498	0.377	0.521	0.405	0.578	0.430	0.581	0.437	0.519	0.413
IMPUTE+IMPUTE	0.112	0.119	0.219	0.204	0.476	0.403	0.504	0.430	0.569	0.460	0.573	0.466	0.537	0.447
SHAPEIT+IMPUTE	0.087	0.093	0.179	0.166	0.435	0.360	0.460	0.385	0.528	0.416	0.532	0.425	0.491	0.402
duoHMM+IMPUTE	0.094	0.103	0.192	0.182	0.443	0.373	0.467	0.396	0.533	0.427	0.538	0.434	0.497	0.412
Beagle+Beagle	0.055	0.052	0.119	0.093	0.389	0.298	0.422	0.332	0.499	0.366	0.504	0.381	0.443	0.354

Numbers in bold are the greatest in each column ignoring the first row (i.e. Observed genotypes); Underlined numbers are the greatest within each group of approaches; Total number of replicates=100 (see Methods: Quantitative traits); For Ped_Pop, we combined the imputation results of GIGI and duoHMM+minimac (see Supplementary Methods for further details); The number of tests (in millions) we performed were: 2.50, 1.71, 0.96, 0.91, 0.77, 0.65, 0.46 for EUR and 2.62, 3.33, 1.65, 1.38, 0.84, 0.58, and 0.48 for AFR for the respective intervals [0,0.01), [0.01,0.05), [0.05,0.1), [0.1,0.2), [0.2,0.3), [0.3,0.4), and [0.4,0.5].

Table 2: Phasing and Imputation approach summary

	Phasing	Imputation
Family-Based Approaches	• gl_auto (MORGAN) – Samples inheritance vectors (IVs) from the pedigree and genotype data from the set of most informative sparse markers – Uses a combination of exact and Markov Chain Monte Carlo (MCMC) based estimations – Allows multi-point IBD estimation on the complete pedigree – Uses the Lander-Green algorithm as well as the Elston-Stewart algorithm as part of computations	• GIGI – Relies on correlation resulting from inheritance in pedigrees through the inheritance of shared segments of a chromosome as represented by IVs – Uses a sparse set of “framework markers” typed on most subjects plus a set of “dense markers” typed on a few subjects • Merlin – Relies on the Lander-Green algorithm for traversing the pedigree
Population-Based Approaches	• SHAPEIT (v2) – A Hidden Markov model-based (HMM) approach where haplotypes of each sample are updated iteratively and inference is done using Gibbs sampling – Creates a graph capturing the haplotype structure of all haplotypes using a greedy approach – Uses a different space to represent haplotypes that are consistent with a subject’s genotypes – Incorporates surrogate family phasing approach • duoHMM – An HMM based approach with a constant number of hidden states, four in this case, and sixteen possible transitions between states per meiosis – Corrects the results of SHAPEIT that are inconsistent with pedigree information such as switch and genotyping errors along with detection of recombination events – Algorithmic complexity is O(nL), where n is the number of non-founders and L is the number of SNPs • IMPUTE (v2) – An MCMC based approach that probabilistically samples phased haplotypes for each subject, conditional on the current haplotype guesses for the rest of the subjects – Incorporates surrogate family phasing approach – Uses k templates from the set of available haplotypes of a subject to reduce the runtime • MaCH – A Markov Chain (MC) based approach that iteratively samples a new pair of haplotypes for each subject using an HMM, which describes the haplotype pair as an imperfect mosaic of the other haplotypes – Typically uses 20-100 iterations to construct a consensus haplotype by merging the haplotypes sampled from each round – Algorithmic complexity is O(ns^2), where n is the number of rounds and s is the number of states • Beagle (v4.1) – An empirical LD model based approach that adapts to the local structure of the data by modeling haplotype frequencies on a local scale – Initially clusters the haplotypes at each marker such that the haplotypes in the same cluster tend to have similar probabilities for alleles at downstream markers – A diploid HMM is used with ordered pairs of edges at each level of the model to find the most likely haplotype pairs for each subject given the subject's known genotypes • Eagle (v2) – A BEAGLE based approach – Uses a new data structure based on the positional Burrows-Wheeler transform and a rapid search algorithm that explores only the most relevant paths through the HMM	• IMPUTE (v2) – Searches for reference haplotypes that share high sequence identity with the haplotypes of the subject being imputed – Considers the genetic distance of a locus of interest to its neighbors for imputation – Assumes a uniform mutation rate across the genome • minimac (v3) – Searches for reference haplotypes that share high sequence identity with the haplotypes of the subject being imputed – Requires the haplotypes for all subjects, which can be obtained by using one of the phasing methods described above • Beagle (v4.1) – An HMM based approach, which uses the set of all ordered pairs such that the first element is an aggregate genotyped marker and the second element is a reference haplotype – Considers a fixed genetic distance, 0.005 cM by default, to combine the sets of consecutive markers within that distance into an aggregate set of genotyped markers in the study dataset
Combined Approach	• Ped_Pop – Combines family- and population-based imputation results to benefit from both IBD and LD information – Compares the variance of the three genotype posterior probabilities between one family-based and one population-based method, and select the probabilities with the highest variance	