Genotype imputation accuracy and the quality metrics of the minor ancestry in multi-ancestry reference panels

Mingyang Shi,[1] Chizu Tanikawa,[2] Hans Markus Munter,[3] Masato Akiyama,[4] Satoshi Koyama,[5] Kohei Tomizuka,[6] Koichi Matsuda,[2] Gregory Mark Lathrop,[3] Chikashi Terao,[6] Masaru Koido,[1] and Yoichiro Kamatani[1]

[1] Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Japan.

[2] Laboratory of Clinical Genome Sequencing, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Japan.

[3] Department of Human Genetics, School of Biomedical Sciences, Faculty of Medicine and Health Sciences, McGill University, Canada.

[4] Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Japan.

[5] Cardiovascular Disease Initiative, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, US.

[6] Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Japan.

Correspondence: M.K. (mkoido@edu.k.u-tokyo.ac.jp) and Y.K. (kamatani.yoichiro@edu.k.u-tokyo.ac.jp)

24 Tel: +81 (0)3-5449-5286

25 Fax: None

Key Words

Genotype imputation, Imputation quality metric, East Asian, Under-represented population, TOPMed reference panel, Template switching rate

Abstract

Large-scale imputation reference panels are now available and have contributed to efficient genome-wide association studies through genotype imputation. However, it is still under debate whether large-size multi-ancestry or small-size population-specific reference panels are the optimal choices for under-represented populations. We imputed genotypes of East Asian (EAS; 180k Japanese) subjects using the Trans-Omics for Precision Medicine (TOPMed) reference panel and found that the standard imputation quality metric (Rsq) substantially overestimated the dosage r^2 (squared correlation between imputed dosage and true genotype). Variance component analysis of Rsq revealed that the increased imputed-genotype certainty (dosages closer to 0, 1, or 2) caused upward bias, indicating some systemic bias in the imputation. Through systematic simulations using different template switching rates (θ value) in the hidden Markov model, we uncovered that the lower θ value increased the imputed-genotype certainty and Rsq; however, dosage r^2 was insensitive to the θ value, thereby causing a deviation. In simulated reference panels with different sizes and ancestral diversities, the θ value estimates from Minimac decreased with the size of a single ancestry and increased with the ancestral diversity. Thus, Rsq could overestimate or underestimate dosage r^2 for a subpopulation in the multi-ancestry panel and the deviation represents different imputed-dosage distributions. Finally, despite the impact of θ value, distant

ancestries in the reference panel contributed only a few additional variants passing a predefined Rsq threshold. We conclude that the θ value has a substantial impact on the imputed dosage and the imputation quality metric value.

Introduction

Genotype imputation is a cost-efficient technique to expand the number of markers in genetic studies. Li and Stephen's hidden Markov model (HMM) and the pre-phasing-imputation pipeline are employed by the popular tools Minimac, IMPUTE, and BEAGLE, with slightly different implementations but similar accuracy [1–5].

The imputation accuracy (*e.g.*, the squared correlation between imputed dosage and true genotype; dosage r^2) [6], the scores generated by software without the true genotype (Rsq and INFO) [3,7,8], and the number of variants passing a predetermined Rsq threshold (high-Rsq variants) are widely used quality metrics to evaluate imputation performance [9]. The performance primarily depends on the reference panel size and genetic similarity between the reference panel and target sample [10]. In studies using the International HapMap Project (HapMap) and the 1000 Genomes Project (1KGP) [11,12], pooling multiple ancestries together to maximize the panel size typically improved the imputation performance compared to using only the matched ancestry [8,13,14]. Deelen et al. combined the population-specific whole genome sequencing (WGS) dataset of the Genome of the Netherlands (GoNL) with the 1KGP and showed higher dosage r^2 than when using either GoNL or 1KGP alone [15]. Huang et al. combined the UK10K with the 1KGP to generate more variants with high INFO [9].

Moreover, larger multi-ancestry reference panels, such as the Haplotype Reference Consortium (HRC) and the Trans-Omics for Precision Medicine (TOPMed) [16,17], greatly improved the imputation accuracy and increased the number of high-Rsq variants in European (EUR), African (AFR), and Admixed American (AMR) [10,18].

However, constructing large reference panels by combining diverse ancestries is not always beneficial. Small-size population-specific panels, such as Norwegian and Estonian panels [19,20], achieve similar performance to the HRC panel in EUR. Moreover, even though the HRC panel includes all 1KGP samples, the imputation accuracy in non-EUR populations can be inferior to that of the 1KGP panel alone [10,21]. Bai et al. found that adding 27 samples from a different ancestry to the Han Chinese (1KGP-CHB, size = 103) resulted in better imputation accuracy than using the 1KGP-CHB or 1KGP panel [22]. In studies of Asian, combining population-specific WGS datasets with the 1KGP had improved the imputation accuracy in some instances [23–25], but not in others [26–28]. So far, the reason for these discrepancies is unclear.

Two sets of parameters are used in the HMM-based imputation algorithms. Transition probability is governed by the template switching rate (θ) between adjacent markers, which models the recombination and relatedness, and emission probability is determined by the error rate (ε) for each marker, which models genotyping error, gene conversion, and recurrent mutation [3,7]. These parameters are crucial to the HMM for calculating matching probabilities between the template and target during the imputation process, but dosage r^2 is robust to these parameters [3,5]. Many studies have

95 suggested different Rsq or INFO thresholds to achieve a similar dosage r^2 [29,30],
96 making it difficult to ascertain the relationship between the imputation process, Rsq, and
97 dosage r^2.
98
99 In this study, we imputed the East Asian (EAS) using the TOPMed reference panel
100 (EAS comprised 1.22% of the total samples) and found that Rsq substantially
101 overestimated dosage r^2. We introduced novel variance component analysis for Rsq
102 and analytically investigated why Rsq was overestimated and characterized the
103 relationship between the template switching rate used in the HMM, quality metrics, and
104 imputed dosage. Further, we evaluated the θ value, Rsq, dosage r^2, the deviation
105 between them, and the number of high-Rsq variants from matched or distant ancestry in
106 cases where the target ancestry was the major or minor component of the multi-
107 ancestry panels.
108

109 **Material and methods**

110 **Subjects, genotyping, and quality control (QC)**

111 Cohort specifications of Biobank Japan (BBJ-180k), genotyping, and QC are provided in
112 **Supplementary Method 1**. For imputation using the TOPMed panel, we lifted the
113 coordinates of the genotyping array from hg19 to hg38 using LiftOver [31]. Palindromic
114 variants (reference/alternative (ref/alt) alleles were G/C or A/T) and variants not in the
115 TOPMed freeze5b (https://bravo.sph.umich.edu/freeze5/hg38/) were excluded. The

116 ref/alt alleles were swapped and/or reverse-complemented according to the hg38
117 reference sequence. A total of 515,587 autosomal variants remained.
118

119 **Pre-phasing, reference panel construction, and imputation**

120 EAGLE v2.4.1 was used for pre-phasing without an external reference [32]. The 1KGP
121 reference panel was downloaded from https://genome.sph.umich.edu/wiki/Minimac3.
122 Methods used to construct the population-specific panels (named the BBJ1k and
123 JEWEL3k) are provided in **Supplementary Method 2**. We used an in-house server to
124 perform imputation using the 1KGP, BBJ1k, and JEWEL3k panels. Minimac3 v2.0.3
125 was used to estimate the HMM parameters and prepare m3vcf files.[3] Minimac4 v1.0.2
126 was used for imputation [33]. Imputation using the TOPMed and HRC panel were
127 performed using the TOPMed and Michigan imputation servers [3,17], respectively.
128

129 **Evaluation of the imputation performance using different reference panels**

130 Rsq from the Minimac4's info file was extracted for the subsequent analyses. Variants
131 with Rsq < 0.3 were removed [34]. Imputation performance was empirically evaluated
132 using 993 samples with WGS (named WGS_{993}). QC and processing are elaborated in
133 **Supplementary Method 3**. Coverage of a reference panel was the fraction of variants
134 in WGS_{993} that could be imputed with Rsq ≥ 0.3. Dosage r^2 was the squared Pearson
135 correlation coefficient between the imputed dosage and true genotype (encoded as 0, 1,

or 2). Minor allele frequency (MAF) and minor allele count (MAC) of WGS_{993} were used to stratify variants into bins.

Quantification of the deviation between Rsq and dosage r^2

We modeled the relationship between the imputed allelic dosage ($\boldsymbol{y} = (y_1, \dots, y_n)$, $y_i \in [0,1]$) and the true allele ($\boldsymbol{x} = (x_1, \dots, x_n)$, $x_i \in \{0,1\}$) using the simple linear regression formula:

$$\boldsymbol{y} = \beta_0 + \boldsymbol{x}\beta_{imp} + \boldsymbol{\varepsilon}$$

(Equation 1)

where β_{imp} is a scalar of the regression coefficient with $\hat{\beta}_{imp} = Cov\,(\boldsymbol{x}, \boldsymbol{y})/Var\,(\boldsymbol{x})$, β_0 is the intercept, and $\boldsymbol{\varepsilon} = (\epsilon_1, \dots, \epsilon_n)$ is an error term. The Minimac's EmpRsq metric, which is the squared Pearson correlation coefficient between ***x*** and ***y*** (https://genome.sph.umich.edu/wiki/Minimac3_Info_File), equals the ratio between the regression sum of squares (SS_{reg}) and the total sum of squares (SS_{tot}) in this simple linear regression. Then we have:

$$EmpRsq = \frac{Cov(\boldsymbol{x}, \boldsymbol{y})^2}{Var(\boldsymbol{x})Var(\boldsymbol{y})} = \frac{SS_{reg}}{SS_{tot}}$$

(Equation 2)

where Cov(***x***, ***y***), Var(***x***), and Var(***y***) are the covariance between ***x*** and ***y***, the variance of ***x***, and the variance of ***y***, respectively.

156 Rsq is the ratio between Var(***y***) and $p(1-p)$, where *p* is the alternative allele frequency

157 (AAF) in the imputed dataset [3,7]. Then,

158 $$Rsq = \frac{Var(\boldsymbol{y})}{p(1-p)} = \frac{SS_{tot}/n}{p(1-p)} = \frac{SS_{reg}/n}{p(1-p)} + \frac{SS_{res}/n}{p(1-p)}$$

159 (Equation 3)

160 where SS_{res} is the residual sum of squares that follows $SS_{tot} = SS_{reg} + SS_{res}$, and *n* is

161 the number of imputed haplotypes. Hence, Rsq comprises two parts: regression-related

162 and residual-related. We define $\frac{SS_{res}/n}{p(1-p)}$ as MAF-Adjusted-Residual-Error (MARE). Then:

163 $$Rsq = \frac{SS_{reg}/n}{p(1-p)} + MARE = \frac{EmpRsq \times SS_{tot}/n}{p(1-p)} + MARE = \frac{EmpRsq \times Var(\boldsymbol{y})}{p(1-p)} + MARE$$

164 (Equation 4)

165 By assuming the equal AAF in ***x*** and ***y***, *i.e.*, $Var(\boldsymbol{x}) = p(1-p)$, Rsq could be further

166 treated as:

167 $$Rsq = \frac{Var(\boldsymbol{y})Cov(\boldsymbol{x},\boldsymbol{y})^2}{Var(\boldsymbol{x})Var(\boldsymbol{x})Var(\boldsymbol{y})} + MARE = \frac{Cov(\boldsymbol{x},\boldsymbol{y})^2}{Var(\boldsymbol{x})^2} + MARE = \hat{\beta}_{imp}^2 + MARE$$

168 (Equation 5)

169 Finally, *MARE* and $\hat{\beta}_{imp}$ could be obtained from Rsq and EmpRsq:

170 $$MARE = Rsq - \frac{Var(\boldsymbol{y})EmpRsq}{p(1-p)} = Rsq - Rsq \times EmpRsq = (1 - EmpRsq)Rsq$$

171 (Equation 6)

172 $$\hat{\beta}_{imp} = \sqrt{Rsq - (1 - EmpRsq)Rsq} = \sqrt{EmpRsq \times Rsq}$$

173 (Equation 7)

174 In Equation 7, $\hat{\beta}_{imp}$ is negative if ***x*** and ***y*** are negatively correlated. We do not consider

175 that situation. Hence, each combination of Rsq and EmpRsq indicates specific values of

176 *MARE* and $\hat{\beta}_{imp}$. Hereafter, we refer to $\hat{\beta}_{imp}$ as the β_{imp} metric and *MARE* as the MARE
177 metric.
178
179 Because the WGS dataset is unphased diploid data, dosage r^2 calculated from that will
180 be slightly different from EmpRsq. **Supplementary Note 1** provides the detailed
181 methods to calculate MARE and β_{imp} from haploid or diploid data, and the concordance
182 between values obtained using Equations 6–7 and calculated from the imputed dosage
183 in both haploid and diploid cases.
184

185 **Imputation using the simulated 1KGP reference panels with different θ values**

186 The 1KGP vcf files were downloaded from the Minimac3 website (see above).
187 Parameters were estimated using Minimac3. The θ value was extracted (denoted as
188 "Recom" in the m3vcf file) [3], scaled by 21 folds manually (0.01–100), and replaced in
189 the original file. A brief explanation of the θ value is provided in **Supplementary Note 2**.
190 The array data of WGS_{993}, a subset of BBJ-180k, was used as the target sample. The
191 imputation procedure was the same as described above.
192

193 **Evaluation of the imputation performance using the simulated reference panels**

194 We obtained the imputed allelic dosage (LooDosage) from Minimac4's empiricalDose
195 file (by turning the "--meta" option on) [35]. It was derived from the leave-one-out (LOO)
196 method by hiding the markers on the array during the imputation. Rsq, EmpRsq, MARE,

197 and β_{imp} were calculated from the LooDosage and array data. MAF was obtained from
198 the array data.
199

200 **Simulation of reference panels and estimating the θ value**

201 We sampled subsets from the 1KGP and JEWEL3k, shuffled the sample order 10 times,
202 and created new vcf files using bcftools v1.14 (https://samtools.github.io/bcftools/) [36].
203 We estimated the parameters using Minimac3. **Supplementary Method 4** provides the
204 detailed methods. The total θ value along chr19 (by summing the θ values between
205 adjacent markers) was used to evaluate the reference panel size and ancestral
206 diversity's impact. A discussion of the θ value qualification is provided in
207 **Supplementary Note 3**.

208

209 **EUR-EAS reference panel simulation and imputation**

210 We randomly sampled 403 individuals (named 1KGP-EUR_{403}) from the 1KGP-EUR and
211 combined them with the 1KGP-EAS and 6 subsets (size = 500, 1,000, 1,500, 2,000,
212 2,500, and 3,256) of the 3,256 JPT WGS samples (named as JPT_{3256}) in JEWEL3k.
213 The remaining 100 individuals in the 1KGP-EUR were used as the target sample. We
214 extracted the 10,375 polymorphic variants (chr19) on the Illumina Global Screening
215 Array v3.0 (https://support.illumina.com/content/dam/illumina-
216 support/documents/downloads/productfiles/global-screening-array-24/v3-0/infinium-

217 global-screening-array-24-v3-0-a1-b151-rsids.zip) to simulate the genotyping array.
218 Parameter estimation and imputation were the same as above.
219

220 **JPT-1KGP reference panel simulation and imputation**

221 We sampled 7 subsets (size = 100, 500, 1,000, 1,500, 2,000, 2,500, and 3,256) from the
222 JPT_{3256} and then combined the JPT_{3256} with 6 subsets (1KGP-JPT and 1–5 ancestries)
223 of the 1KGP. WGS_{993} was used as the target. The other processing methods were the
224 same as above.
225

226 **Number of confident alleles and high-Rsq variants**

227 Haploid dosage (HDS) is the imputed alternative allele's allelic dosage at the haploid
228 level, which could be obtained by the "--format HDS" option in Minimac4. Briefly, it is in
229 a per-variant and per-individual manner and a higher value represents an imputed
230 alternative allele is more certain. We quantified the θ value's impact on the imputed
231 dosages by the number of confident alleles (HDS > 0.9). Rsq was used to judge how
232 many high-Rsq (Rsq > 0.7) variants could be passed to the downstream analyses.
233

234 Furthermore, we determined how many confident alleles and high-Rsq variants could be
235 obtained only from the additional distant ancestries in the multi-ancestry reference
236 panel. We defined ancestry-specific variants as follows: EUR-only variants only existed
237 in the 1KGP-EUR_{403}, and it has no non-EUR variants. JPT_{3256}-only and 1KGP-EAS-only

238 variants only existed in the JPT_{3256} and 1KGP-EAS, respectively. Non-EAS variants
239 were not found in the JPT_{3256} or 1KGP-EAS.
240

241 **Replication of the TOPMed imputation pipeline**

242 We followed the TOPMed imputation pipeline (Accessed on Oct. 15, 2022,
243 https://topmedimpute.readthedocs.io/) and used the 1KGP and JEWEL3k reference
244 panels. Specifically, the HapMap2 genetic map was used as a reference for the θ value
245 instead of estimating it using Minimac3. WGS_{993} was used as the target. As the θ value
246 was not outputted by default, we modified the Minimac4's source code to obtain the
247 transformed θ value (**Supplementary Method 5**).
248

249 **Results**

250 **BBJ imputation using the four reference panels**

251 The BBJ-180k was imputed using the TOPMed, 1KGP, BBJ1k, and JEWEL3k reference
252 panels (**Figure 1A**). Characteristics of each panel and the target sample are listed in
253 **Table 1**. We categorized the imputed variants by MAF and Rsq in each imputed
254 dataset. With more EAS samples in the panel, more variants with low MAF passed each
255 Rsq threshold (**Figure 2A** and **Supplementary Table 1**). In addition to the absolute
256 number, unique variants were imputed from each panel (**Figure 2B**). These results
257 reproduced the benefits of using large and different reference panels [19].

258

259 We then used WGS_{993} to empirically evaluate the imputation performance. MAF and
260 MAC of WGS_{993} were used to categorize the variants into six bins: common (MAF ≥
261 5%), low-frequency (5% > MAF ≥ 1% and 1% > MAF ≥ 0.5%), rare (0.5% > MAF >
262 0.1%), doubleton (MAC = 2), and singleton (MAC = 1). For common single nucleotide
263 polymorphisms (SNVs), the TOPMed imputation result showed the highest coverage
264 (0.950), whereas BBJ1k (0.890) and JEWEL3k (0.907) were lower than 1KGP (0.948).
265 This loss of variants was possibly due to the additional QC steps in combining the
266 1KGP with Japanese WGS (**Supplementary Method 2**). As MAF decreased, the
267 coverage also decreased; however, more EAS samples in the reference panel mitigated
268 this decrease, as expected (**Figure 2C**).

269

270 The 155,297 shared SNVs (on chr19) imputed by the four panels were used to compare
271 imputation accuracy. The mean dosage r^2 using the TOPMed panel was between those
272 of 1KGP and BBJ1k (**Figure 2D**). However, Rsq was substantially upward biased only
273 in the TOPMed imputation result (**Figure 2E**). Using all SNVs and short insertions and
274 deletions (indels) with Rsq ≥ 0.3, although with higher coverage, the TOPMed
275 imputation result showed an inferior dosage r^2 than 1KGP (**Supplementary Figure 1**).

276

277 **Quantifying the deviation between Rsq and dosage r^2**

278 The deviation between Rsq and dosage r^2 was persistent in all MAF bins only when
279 using the TOPMed panel (**Supplementary Figure 2**), which indicated a potential
280 systematic bias from the imputation pipeline or the reference panel. To justify our
281 observation, we analytically derived the relationship between Rsq and dosage r^2
282 (**Methods**). Two novel metrics, MARE and β_{imp}, were introduced to quantify the
283 deviation. MARE, a MAF-adjusted form of residual error, takes a value between 0 and
284 1, and increases with SS_{res}. β_{imp} describes the distinguishability between the mean
285 imputed dosage of each true genotype group. Rsq is the ratio between observed and
286 expected variance and dosage r^2 shows the correlation. Under the assumption of "well-
287 calibration" (the posterior allele probability from imputation equals the expected true
288 allele dose), Rsq equals dosage r^2 [10]. Our analytical derivation of the relationship
289 between Rsq and dosage r^2 did not assume the "well-calibration" (**Supplementary**
290 **Figure 3–7** and **Supplementary Note 1**). Any two of Rsq, dosage r^2, MARE, and β_{imp}
291 could entirely quantify the relationship between the imputed dosage and true genotype
292 and determine the other two metrics (Equations 6–7), whereas Rsq or dosage r^2 alone
293 could not.

294

295 To exhibit the relationships, we plotted the theoretical values of MARE and β_{imp} on the
296 coordinates of Rsq and dosage r^2 (**Figure 3A** and **3F**). It clearly showed that the
297 overestimated Rsq was accompanied by a higher MARE (**Figure 3A**). We used
298 rs142572000 as an example (**Figure 3B–E**). In the TOPMed imputation result, imputed

genotypes were more certain (defined as the imputed dosage closer to 0, 1, or 2) (**Figure 3B**) [37], compared to the other three panels (**Figure 3C–E**). The high certainty increased Var(***y***) and Rsq. In **Supplementary Note 4**, the positive relationship between imputed-genotype certainty and Rsq is demonstrated. As discussed further below, high certainty or Rsq did not mean the imputation is more accurate. As shown in **Figure 3B**, many heterozygotes were incorrectly imputed with a dosage of approximately 0 in the TOPMed imputation, causing higher MARE and Rsq, and an even low dosage r^2.

Variants with dosage r^2 < Rsq showed a lower β_{imp} (**Figure 3F**). We used rs671 as another example (**Figure 3G–J**). Except for the TOPMed imputation result, imputed dosages were shrunk to the AAF in EAS (0.247) (**Figure 3H–J**), causing uncertainty in the imputed genotype, and decreased β_{imp}, MARE, and Rsq. However, imputed dosages were still highly correlated with the true genotypes (**Figure 3F**), causing dosage r^2 > Rsq. Taking the two examples together, imputation results of a similar Rsq or dosage r^2 may have a great difference in the imputed dosage, thus, more comprehensive evaluations are necessary and the deviation between Rsq and dosage r^2 should not be ignored.

We categorized MARE into Rsq bins and β_{imp} into dosage r^2 bins to compare them between reference panels and to the expected values when assuming Rsq equals dosage r^2. The mean MARE of the TOPMed result was above the expected for Rsq bins

320 0.35–0.9, and the mean β_{imp} of the 1KGP result was below those of the other panels
321 (**Supplementary Figure 8**). These results suggested that the TOPMed result was more
322 certain and might contain more wrongly imputed genotypes (**Figure 3B** and **3G**), while
323 the 1KGP result might possess higher shrinkage in the imputed dosage, as shown
324 above (**Figure 3C** and **3H**).

325

326 **Template switching rate impacts the deviation between Rsq and dosage r^2**

327 The high certainty suggested an overconfident matching between the reference panel
328 and target sample. We investigated how the template switching rate (θ) used in the
329 HMM affects the imputed-allele certainty using the 1KGP reference panel and 21
330 scalings of the θ value (0.01–100-fold of that estimated by Minimac3; **Methods**; **Figure**
331 **1B**). Low θ values caused Rsq > EmpRsq (EmpRsq is the alternate of dosage r^2;
332 **Methods**) (**Supplementary Figure 9A–C**), while high θ values caused Rsq < EmpRsq
333 (**Supplementary Figure 9E–G**). Using rs10410162 as an example, Rsq and MARE
334 increased as the θ value decreased (**Figure 4A**), with increasing certainty of the
335 imputed alleles (**Supplementary Figure 10A–C**) and the deviation toward Rsq >
336 EmpRsq (**Figure 4B**). As the θ value increased, the imputed allelic dosages were
337 shrunk to the AAF (0.351) (**Supplementary Figure 10E–G)**, resulting in a more drastic
338 decrease in Rsq than EmpRsq (**Figure 4A**), thereby causing EmpRsq > Rsq. Notably,
339 EmpRsq and β_{imp} were roughly maintained unless the θ value was scaled up by 2-fold
340 or higher (**Figure 4A** and **4C**), suggesting EmpRsq was insensitive to altering the θ
341 value, particularly the downscaling. In contrast, Rsq was sensitive to the θ value and

342 consistently increased with downscaling of the θ value, leading to the deviation between
343 Rsq and EmpRsq.

344

345 We also showed that low θ values increased the imputed-allele certainty using a
346 randomly selected target haplotype **(Figure 4D** and **Supplementary Figure 11).** High
347 certainty might cause some variants to be wrongly imputed as the opposite allele
348 (**Supplementary Figure 12**). An explanation is provided in **Supplementary Note 5**.
349 Such imputed-dosage properties were also revealed by the mean MARE and β_{imp}
350 across all variants on chr19 (**Supplementary Figure 13**). Taken together, the simulated
351 imputation using an extremely low θ value mimics the high certainty and Rsq
352 overestimation in the TOPMed imputation result in EAS (**Figure 3**).

353

354 **Template switching rate impacts the imputation performance**

355 Using the modified 1KGP panels, although the mean EmpRsq was insensitive to the
356 scaling of the θ value (maximum difference of 0.034 for the θ value scaling between
357 0.01 and 2-fold; **Figure 5**), low θ value increased Rsq (**Figure 5**). We evaluated the
358 number of confident alleles (HDS > 0.9) and high-Rsq variants (Rsq > 0.7) (**Methods**).
359 Downscaling the θ value increased the number of confident alleles and high-Rsq
360 variants (**Supplementary Table 2**). When the θ value was 0.5-fold, the number of
361 confident alleles and high-Rsq variants increased by 5.46% and 8.89%, respectively,
362 and if the θ value was 2-fold, these numbers decreased by 11.8% and 13.8%,

363 respectively. These results indicated that the θ value shaped the imputed allelic dosage
364 and changed Rsq and the number of high-Rsq variants while leaving EmpRsq almost
365 the same.

366

367 **Reference panel and θ estimates**

368 We evaluated how Minimac3's parameter estimation changed with the composition of
369 the reference panel (**Supplementary Method 4**, **Supplementary Figure 14**, and
370 **Supplementary Note 6**). The θ estimates decreased with the sample size of a single
371 ancestry and increased with ancestral diversity when size was fixed (**Figure 6A** and
372 **6B**). On pooling samples from different ancestries together at a small panel size, the θ
373 estimates still decreased as the panel size increased (**Figure 6C**); however, at a larger
374 panel size, it increased with simultaneously increasing the panel size and ancestral
375 diversity (**Figure 6D**), suggesting that the θ estimates were in a trade-off between the
376 panel size and ancestral diversity (**Figure 6E**). **Supplementary Note 6** explains these
377 effects.

378

379 **Fitness of the θ value, deviation, and imputation performance**

380 To elucidate the multi-ancestry reference panel's impact on the imputation result, we
381 simulated two scenarios using the multi-ancestry reference panels: the target sample
382 was from the (1) minor and (2) major ancestry (**Figure 1C**).

383

384 *Scenario 1: The target sample was from the minor ancestry*

385 We simulated 8 EUR-EAS reference panels and used 100 EUR samples as the target
386 (**Methods**). As the panel size increased and θ value decreased (**Table 2**), Rsq and
387 MARE were upwardly biased, as expected (**Supplementary Figure 15** and **S16**). The
388 mean EmpRsq decreased consistently with the addition of EAS samples to the
389 reference panel (**Table 2**). However, only a marginal difference was observed
390 (maximum difference of 0.020).

391

392 There were 114,606 EUR-only and 696,017 non-EUR variants (**Methods**). As the panel
393 size increased and θ value decreased, the number of confident alleles increased from
394 84,797 to 101,116 and from 0 to 458 for EUR-only and non-EUR variants, respectively
395 (**Supplementary Table 3**). The number of high-Rsq variants increased from 163,468 to
396 200,161, 18,646 to 27,941, and 0 to 521 for all, EUR-only, and non-EUR variants,
397 respectively (**Supplementary Table 3**). These results revealed that a lower θ value
398 increases the number of confident alleles and high-Rsq variants, without improving the
399 EmpRsq (**Supplementary Table 3**). Meanwhile, even if the reference panel comprised
400 90.3% EAS samples, non-EUR variants only comprised 0.26% of the total variants in
401 the imputation result (when setting a cutoff of Rsq > 0.7). The majority of variants
402 gaining from the larger panel when imputing the under-represented ancestry was
403 because of the lower θ value.

404

405 *Scenario 2: The target sample was from the major ancestry*

406 We simulated 13 JPT-1KGP reference panels and used WGS_{993} as the target
407 (**Methods**). When only JPT samples were in the panel, the mean EmpRsq and Rsq
408 increased with the panel size, as expected (**Supplementary Figure 17**). When
409 combined with the 1KGP subsets, the mean EmpRsq was highest when using
410 JPT_{3256}+1KGP-EAS for variants with MAF ≥ 1% and JPT_{3256}+1KGP-JPT for variants
411 with 1% > MAF ≥ 0.5% (**Table 3**). Adding other ancestries decreased the mean
412 EmpRsq marginally (maximum difference < 0.01 for all MAF categories). The mean Rsq
413 was the highest when using JPT_{3256} and decreased with the addition of more
414 ancestries, with maximum differences of 0.007, 0.024, and 0.048 for variants with MAF
415 ≥ 5%, 5% > MAF ≥ 1%, and 1% > MAF ≥ 0.5%, respectively (**Table 3**). None of the
416 imputation results showed a noticeable deviation in the MARE and β_{imp}. However, the
417 MARE and β_{imp} of the JPT_{3256}, JPT_{3256}+1KGP-JPT, and JPT_{3256}+1KGP-EAS results
418 were closer to the expected values (**Supplementary Figure 18**), while adding other
419 ancestries made MARE and β_{imp} correspond to the condition of using a higher θ value.

420

421 There were 74,490 JPT_{3256}-only, 7,627 1KGP-EAS-only, and 495,489 non-EAS variants
422 (**Methods**). From JPT_{3256} to JPT_{3256}+1KGP, the number of confident alleles decreased
423 from 24,599 to 23,666, increased from 0 to 131, and increased from 0 to 254 for these
424 three groups of variants (**Supplementary Table 4**). The number of high-Rsq variants

425 decreased from 274,343 to 264,221 (a decrease of 3.69%) when using JPT_{3256} or
426 JPT_{3256}+1KGP (**Supplementary Table 4**). Only 265 non-EAS variants reached an
427 Rsq > 0.7 when using JPT_{3256}+1KGP, which was 0.10% of the total variants. Hence,
428 combining with the 1KGP caused fewer confident alleles and high-Rsq variants to
429 remain in the imputation results, as expected from using higher θ values. Both
430 scenarios suggest that the distant ancestry in the reference panel affects the θ
431 estimates and the number of high-Rsq variants. However, the expanded haplotypes and
432 variant sets from distant ancestries provided only a few additional variants.

433

434 **EAS imputation using the public reference panels**

435 We used the TOPMed imputation pipeline (**Methods**) and observed an upward
436 deviation of Rsq, particularly when using the 1KGP panel (**Supplementary Figure 19**).
437 The θ value transformed from the genetic map was 0.15- and 0.27-fold of that estimated
438 by Minimac3 when using the 1KGP and JEWEL3k panels, respectively
439 (**Supplementary Table 5**).

440

441 As the θ estimates decreased with the size of the major ancestry (**Figure 6C**), it would
442 be underestimated for the minor ancestry in the multi-ancestry panel regardless of using
443 the genetic map or estimating it by the Mimimac3. One example was the HRC panel,
444 which did not use the genetic map but induced Rsq overestimation in the BBJ-180k
445 (**Supplementary Figure 20**). Thus, caution may be required when imputing the under-

446 represented population using large ancestry-imbalanced panels under the current
447 framework.

448

449 **Discussion**

450 The deviation between Rsq and dosage r^2 in the imputation result is raised from the
451 imputed dosage, and has been widely observed using different reference panels and
452 software [9,15,38]. One reason was that the θ value used in the HMM does not fit the
453 panel size, ancestral components in the reference panel, and the target population.
454 When using the multi-ancestry reference panel, distant ancestries affect the θ estimates
455 (in Minimac3/4), then the imputed dosage and Rsq. Our simulations indicated that the
456 lower θ value used by the larger panel could increase the imputed-genotype certainty
457 and the number of high-Rsq variants but not dosage r^2. It may cause confusing
458 benchmarking results and increase the chance of false positives in association tests.
459 Ferwerda et al. reported height and body mass index association signals in an ethnically
460 diverse cohort only when imputing against the TOPMed panel [39]. Bai et al. reported
461 the highest Rsq but lowest dosage r^2 when imputing the Han Chinese using the HRC
462 panel, compared to the 1KGP and a population-specific panel [22], similar to what we
463 had observed in Japanese. In addition to simply comparing dosage r^2 or Rsq, we
464 provided the changes in those metrics, the θ estimates, imputed-dosage certainty, and
465 the number of high-Rsq variants passing to downstream analyses. We provided the
466 script (https://github.com/shimaomao26/impumetric) to allow users to check their results

467 using the leave-one-out imputation of Minimac4 (external WGS not required) or
468 additional WGS data.

469

470 The TOPMed imputation pipeline used the θ value transformed from the HapMap2
471 genetic map by assuming 0.01 centimorgan (cM) corresponds 1% switching rate. The
472 switching rate would thereby be fixed given the variant's base pair position and cM. Two
473 reasons may explain why it works well. First, the genetic map did not significantly
474 change the imputation accuracy, as verified in a Finnish study [40]. We had also shown
475 that dosage r^2 was insensitive to the θ value. Second, the θ value was only
476 underestimated for the EAS (size = 1,184), but the sizes of EUR, AFR, and AMR (size =
477 17,085–47,159) might fit this value. Previous studies reported that Rsq in the TOPMed
478 imputation result was sometimes misleadingly high in AFR [38] and even EUR [41]. A
479 recent paper reported that the TOPMed panel was more robust to the low-density
480 genotyping array than the HRC and 1KGP panels [18]. These results also suggested
481 that a fixed low θ value might be used by the TOPMed pipeline, as we have inferred.
482 Further studies are warranted to comprehensively elucidate the impacts on genetic
483 studies.

484

485 Our results suggest that avoiding ancestral diversity is best when more than 3,000 WGS
486 samples are available to construct a JPT reference panel. Increased diversity would
487 then only have a marginal impact on dosage r^2 but would affect the individual's imputed
488 dosage and cause fewer variants to pass a predetermined Rsq filter. Zhang et al. and

489 Cong et al. similarly observed that adding the 1KGP to about 3,000 Chinese WGS
490 samples would neither benefit nor harm dosage r^2 [26,27]. We focused on explaining
491 the reason in this work. Further studies are warranted to take advantage of combining
492 population-specific and public WGS datasets while avoiding our identified problems.
493
494 Our study has several limitations. First, our simulations of the reference panel were
495 study-specific. As discussed, the θ estimates decreased with the panel size and
496 increased with the ancestral diversity. However, a larger panel typically increases the
497 diversity simultaneously. Therefore, the θ value and imputation result need to be
498 discussed case by case. This limitation also implies that the general experience may not
499 work well for a new reference panel and target dataset. Second, in the simulations of
500 multi-ancestry reference panels (scenarios 1 and 2), WGS datasets were merged using
501 IMPUTE2 [9]. We did not investigate the impact of IMPUTE2 but treated the merged
502 dataset similarly to the WGS dataset. It might underestimate the number of high-Rsq
503 variants from the distant ancestry (**Supplementary Note 7)**. Our results using the 1KGP
504 (not modified by IMPUTE2) also revealed that only a limited number of confident alleles
505 and high-Rsq variants in the imputation result were from the distant ancestry
506 (**Supplementary Note 7** and **Supplementary Table 2**). Therefore, the conclusion that
507 distant ancestry in the reference panel affects the θ estimates rather than providing
508 high-Rsq variants would be valid. However, further studies should be conducted to
509 determine the impact of the panel-merging method and the net gain of variants from
510 distant ancestries.
511

512 In summary, we explained that the HMM parameter could be a potential reason for
513 inaccurate Rsq and inferior dosage r^2 when using large multi-ancestry reference panels.
514 This is also the first study of the relationship between the template switching rate,
515 imputed-genotype certainty, Rsq, and dosage r^2. We envision that our methods and
516 conclusions could provide insights into benchmarking studies, construction of reference
517 panels, and development of imputation algorithms and pipelines in the future.

Funds

This research was supported by the Ministry of Education, Culture, Sports, Sciences and Technology (MEXT) of Japanese government and the Japan Agency for Medical Research and Development (AMED) under grant numbers JP18km0605001 (the BioBank Japan project) and JP19km0405215 (to C.T., K.M., and Y.K.).

Acknowledgments

We thank the staff of the BBJ for collecting and managing samples and clinical information. We acknowledged the Human Genome Center, the Institute of Medical Science, the University of Tokyo (http://sc.hgc.jp/shirokane.html) for providing the super-computing resources.

Web resources

Impumetric, https://github.com/shimaomao26/impumetric

TOPMed imputation server, https://imputation.biodatacatalyst.nhlbi.nih.gov/#!

Michigan Imputation server, https://imputationserver.sph.umich.edu/index.html

Minimac3, https://genome.sph.umich.edu/wiki/Minimac3

Minimac4, https://genome.sph.umich.edu/wiki/Minimac4

EAGLE, https://alkesgroup.broadinstitute.org/Eagle/

Bcftools, https://samtools.github.io/bcftools/

LiftOver, https://genome.ucsc.edu/cgi-bin/hgLiftOver

Data and code availability

We provided the tool to replicate the metrics used in this work (https://github.com/shimaomao26/impumetric). Other scripts were also deposited there. Genotype data for BBJ and the TOPMed imputation results were deposited at NBDC Human Database (research ID: hum0014 and hum0311, respectively).

Appendix

Imputation quality metrics

Rsq: The standard quality metric in Minimac. The ratio between observed variance and expected variance of imputed allelic dosage.

INFO: The standard quality metric in IMPUTE. The ratio between observed information and complete information of imputed genotype distribution.

Dosage r^2: The squared Pearson correlation between imputed dosage and true genotype.

EmpRsq: The squared Pearson correlation between imputed allelic dosage and true allele dose. Provided by Minimac and is only available for the genotyped markers on array.

MARE: The minor allele frequency adjusted residual error in linear regression between imputed dosage and true genotype.

β_{imp}: The regression coefficient between imputed dosage and true genotype.

561 **References**

562 1. Li N, Stephens M. Modeling linkage disequilibrium and identifying recombination
563 hotspots using single-nucleotide polymorphism data. Genetics 2003; 165:2213–2233
564 2. Howie B, Fuchsberger C, Stephens M, et al. Fast and accurate genotype imputation
565 in genome-wide association studies through pre-phasing. Nat. Genet. 2012; 44:955–
566 959
567 3. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and
568 methods. Nat. Genet. 2016; 48:1284–1287
569 4. Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation
570 Method for the Next Generation of Genome-Wide Association Studies. PLOS Genet.
571 2009; 5:e1000529
572 5. Browning BL, Browning SR. Genotype Imputation with Millions of Reference
573 Samples. Am. J. Hum. Genet. 2016; 98:116–126
574 6. Li Y, Sidore C, Kang HM, et al. Low-coverage sequencing: Implications for design of
575 complex trait association studies. Genome Res. 2011; 21:940–951
576 7. Li Y, Willer CJ, Ding J, et al. MaCH: using sequence and genotype data to estimate
577 haplotypes and unobserved genotypes. Genet. Epidemiol. 2010; 34:816–834
578 8. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat.
579 Rev. Genet. 2010; 11:499–511
580 9. Huang J, Howie B, McCarthy S, et al. Improved imputation of low-frequency and rare
581 variants using the UK10K haplotype reference panel. Nat. Commun. 2015; 6:8111
582 10. Das S, Abecasis GR, Browning BL. Genotype Imputation from Large Reference
583 Panels. Annu. Rev. Genomics Hum. Genet. 2018; 19:73–96

584 11. Frazer KA, Ballinger DG, Cox DR, et al. A second generation human haplotype map
585 of over 3.1 million SNPs. Nature 2007; 449:851–861
586 12. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic
587 variation. Nature 2015; 526:68–74
588 13. Huang L, Li Y, Singleton AB, et al. Genotype-imputation accuracy across worldwide
589 human populations. Am. J. Hum. Genet. 2008; 84:235–250
590 14. Liu Q, Cirulli ET, Han Y, et al. Systematic assessment of imputation performance
591 using the 1000 Genomes reference panels. Brief. Bioinform. 2015; 16:549–562
592 15. Deelen P, Menelaou A, Van Leeuwen EM, et al. Improved imputation quality of low-
593 frequency and rare variants in European samples using the 'Genome of the
594 Netherlands'. Eur. J. Hum. Genet. 2014; 22:1321–1326
595 16. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes
596 for genotype imputation. Nat. Genet. 2016; 48:1279–1283
597 17. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from
598 the NHLBI TOPMed Program. Nature 2021; 590:290–299
599 18. Hanks SC, Forer L, Schönherr S, et al. Extent to which array genotyping and
600 imputation with large reference panels approximate deep whole-genome sequencing.
601 Am. J. Hum. Genet. 2022; 109:1653–1666
602 19. Zhou W, Fritsche LG, Das S, et al. Improving power of association tests using
603 multiple sets of imputed genotypes from distributed reference panels. Genet. Epidemiol.
604 2017; 41:744–755

605 20. Mitt M, Kals M, Pärn K, et al. Improved imputation accuracy of rare and low-
606 frequency variants using population-specific high-coverage WGS-based imputation
607 reference panel. Eur. J. Hum. Genet. 2017; 25:869–876
608 21. Lin Y, Liu L, Yang S, et al. Genotype imputation for Han Chinese population using
609 Haplotype Reference Consortium as reference. Hum. Genet. 2018; 137:431–436
610 22. Bai WY, Zhu XW, Cong PK, et al. Genotype imputation and reference panel: a
611 systematic evaluation on haplotype size and diversity. Brief. Bioinform. 2020; 21:1806–
612 1817
613 23. Akiyama M, Ishigaki K, Sakaue S, et al. Characterizing rare and low-frequency
614 height-associated variants in the Japanese population. Nat. Commun. 2019; 10:4393
615 24. Jeon S, Bhak Y, Choi Y, et al. Korean Genome Project: 1094 Korean personal
616 genomes with clinical information. Sci. Adv. 2020; 6:eaaz7835
617 25. Yoo S-K, Kim C-U, Kim HL, et al. NARD: whole-genome reference panel of 1779
618 Northeast Asians improves imputation accuracy of rare and low-frequency variants.
619 Genome Med. 2019; 11:64
620 26. Cong P-K, Bai W-Y, Li J-C, et al. Genomic analyses of 10,376 individuals in the
621 Westlake BioBank for Chinese (WBBC) pilot project. Nat. Commun. 2022; 13:2939
622 27. Zhang P, Luo H, Li Y, et al. NyuWa Genome resource: A deep whole-genome
623 sequencing-based variation profile and reference panel for the Chinese population. Cell
624 Rep. 2021; 37:110017
625 28. Kawai Y, Mimori T, Kojima K, et al. Japonica array: improved genotype imputation
626 by designing a population-specific SNP array with 1070 Japanese individuals. J. Hum.
627 Genet. 2015; 60:581–587

628 29. Pistis G, Porcu E, Vrieze SI, et al. Rare variant genotype imputation with thousands
629 of study-specific whole-genome sequences: implications for cost-effective study
630 designs. Eur. J. Hum. Genet. 2015; 23:975–983
631 30. Roshyara NR, Horn K, Kirsten H, et al. Comparing performance of modern genotype
632 imputation methods in different ethnicities. Sci. Rep. 2016; 6:34386
633 31. Kent WJ, Sugnet CW, Furey TS, et al. The Human Genome Browser at UCSC.
634 Genome Res. 2002; 12:996–1006
635 32. Loh P-R, Danecek P, Palamara PF, et al. Reference-based phasing using the
636 Haplotype Reference Consortium panel. Nat. Genet. 2016; 48:1443–1448
637 33. Das S. Next Generation of Genotype Imputation Methods. 2017;
638 34. Liu EY, Buyske S, Aragaki AK, et al. Genotype Imputation of Metabochip SNPs
639 Using a Study-Specific Reference Panel of ~4,000 haplotypes in African Americans
640 From the women's health initiative. Genet. Epidemiol. 2012; 36:107–117
641 35. Yu K, Das S, LeFaive J, et al. Meta-imputation: An efficient method to combine
642 genotype data after imputation with multiple reference panels. Am. J. Hum. Genet.
643 2022; 109:1007–1015
644 36. Li H. A statistical framework for SNP calling, mutation discovery, association
645 mapping and population genetical parameter estimation from sequencing data.
646 Bioinformatics 2011; 27:2987–2993
647 37. Marchini J, Howie B, Myers S, et al. A new multipoint method for genome-wide
648 association studies by imputation of genotypes. Nat. Genet. 2007; 39:906–913
649 38. Sun Q, Yang Y, Rosen JD, et al. MagicalRsq: Machine-learning-based genotype
650 imputation quality calibration. Am. J. Hum. Genet. 2022; 109:1986–1997

651 39. Ferwerda B, Abdellaoui A, Nieuwdorp M, et al. A Genetic Map of the Modern Urban
652 Society of Amsterdam. Front. Genet. 2021; 12:
653 40. Hassan S, Surakka I, Taskinen M-R, et al. High-resolution population-specific
654 recombination rates and their effect on phasing and genotype imputation. Eur. J. Hum.
655 Genet. 2021; 29:615–624
656 41. Sun Q, Liu W, Rosen JD, et al. Leveraging TOPMed imputation server and
657 constructing a cohort-specific imputation reference panel to enhance genotype
658 imputation among cystic fibrosis patients. Hum. Genet. Genomics Adv. 2022; 3:100090
659
660

Figures

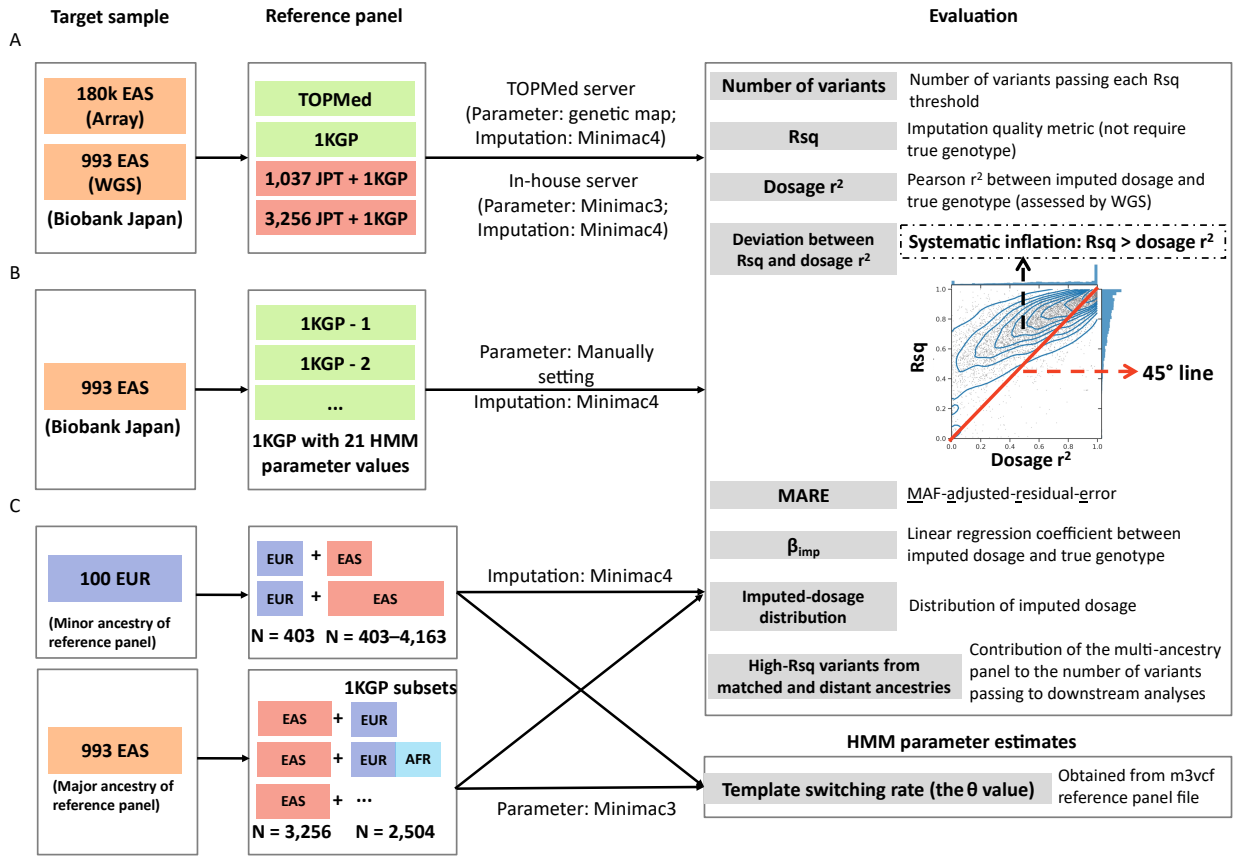


Figure 1. Overview of this study. (A) Imputation using the TOPMed, 1KGP, BBJ1k (1,037 Japanese (JPT) WGS + 1KGP), and JEWEL3k (3,256 JPT WGS + 1KGP) reference panels. The 180k East Asian (EAS) samples in the Biobank Japan (BBJ) first cohort were imputed against the four panels. Samples with whole genome sequencing (WGS_{993}) were used to empirically evaluate the imputation performance, including a count of imputed variants and estimated (Rsq) and true (dosage r^2) imputation accuracy. The deviation between Rsq and dosage r^2 was examined and quantified analytically. Two novel metrics, MAF-adjusted-residual-error (MARE) and β_{imp}, were introduced to indicate the different imputed-dosage distributions underlying the

673 deviation. (B) Simulations to evaluate the relationship between the hidden Markov
674 model (HMM) parameter, Rsq, dosage r^2, the deviation between Rsq and dosage r^2,
675 and imputed dosage. The reference panel and target sample were fixed to the 1KGP
676 and WGS_{993}, and 21 scaled template switching rates (the θ value) were used for
677 running imputations. Evaluations were the same as in (A). (C) Simulations to evaluate
678 the relationship between the reference panel, HMM parameter, imputation metrics, and
679 imputation result. Two scenarios were simulated to represent the imputations of the
680 major and minor ancestry in a multi-ancestry panel. The 3,256 JPT WGS and 1KGP
681 were used to simulate reference panels with different sizes and ancestral diversities.
682 European (EUR) and JPT samples were used as the target. Imputation results were
683 evaluated the same as in (B). In addition, we evaluated how the Minimac3's parameter
684 estimates change with the panel composition.

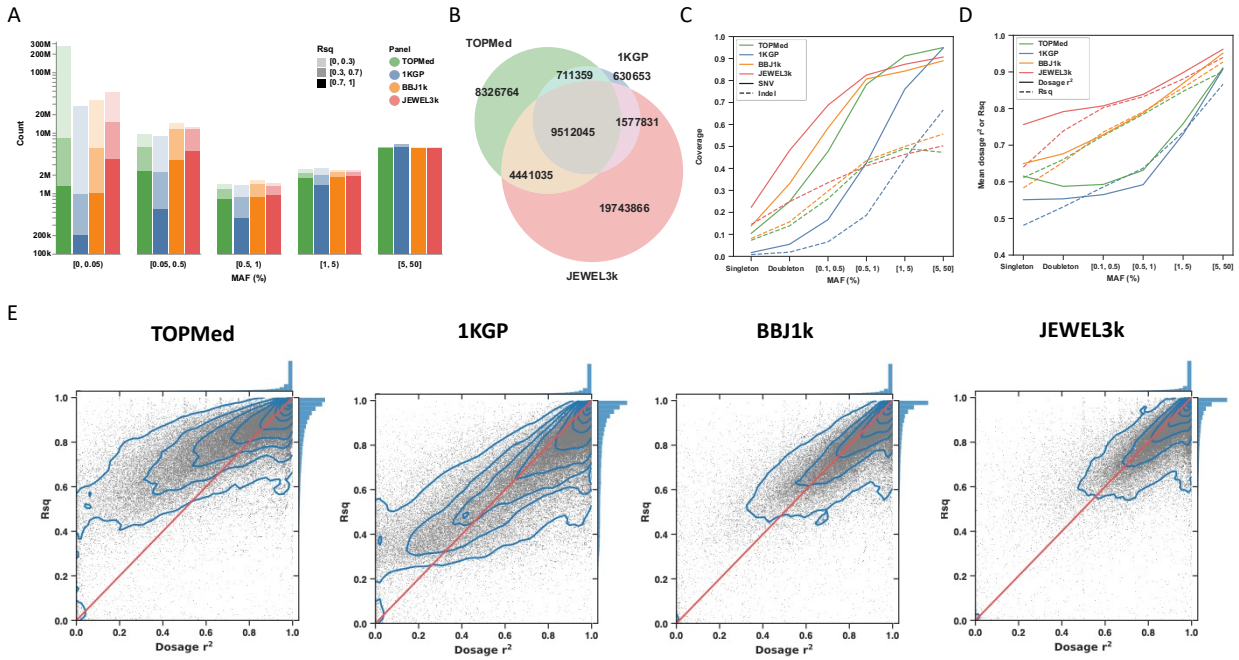


Figure 2. Imputation results using the TOPMed, 1KGP, BBJ1k, and JEWEL3k reference panels. (A) Count of imputed variants using the four panels, stratified by minor allele frequency (MAF) and Rsq in each imputation result. (B) Count of unique and shared variants (with Rsq ≥ 0.3) using the TOPMed, 1KGP, and JEWEL3k panels. (C) Fraction of SNV and indel in WGS_{993} that could be imputed with Rsq ≥ 0.3 (coverage) using each panel, stratified by MAF of WGS_{993}. (D) The mean dosage r^2 and Rsq, stratified by MAF of WGS_{993}. (E) Comparison of dosage r^2 and Rsq in the imputation results using the four panels. The scatter represents the variants, with the blue lines showing the density. The red line shows that Rsq equals dosage r^2, and the histograms on the side show the distribution. In (C), (D), and (E), variants on chromosome 19 were used. In (D) and (E), overlapping variants imputed by the four panels (Rsq ≥ 0.3 using all panels) were used.

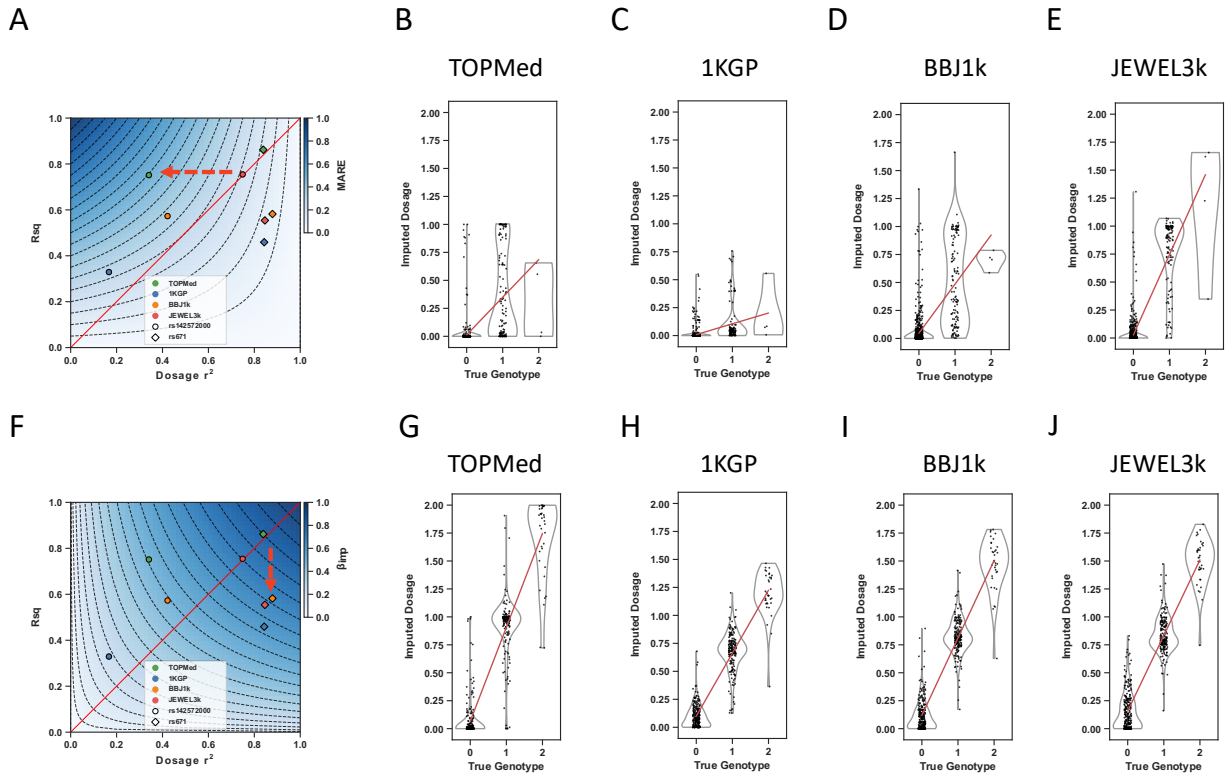


Figure 3. MARE, β_{imp}, Rsq, dosage r^2, and imputed dosage of rs142572000 and rs671. (A) The plot shows Rsq and dosage r^2 of rs142572000 and rs671, and the theoretical value of MARE on the Rsq ~ dosage r^2 plot. (B–E) The plot shows the imputed dosage of rs142572000 in the imputation result using the (B) TOPMed, (C) 1KGP, (D) BBJ1k, and (E) JEWEL3k reference panels. (F) The plot shows the theoretical value of β_{imp} on the Rsq ~ dosage r^2 plot. (G–J) The plot shows the imputed dosage of rs671 in the imputation result using the (G) TOPMed, (H) 1KGP, (I) BBJ1k, and (J) JEWEL3k panels. In (A) and (F), the red line shows that Rsq equals dosage r^2, and the color keys show the theoretical MARE or β_{imp} values. The dashed lines represent the contours of MARE or β_{imp}, and adjacent contours have a value difference of 0.05. The red dashed arrow highlights the difference in MARE and β_{imp} (crossing how many contour lines) for variants showing a similar Rsq but different dosage r^2

711 (rs142572000), or similar dosage r^2 but different Rsq (rs671). In (B–E) and (G–J), the
712 strip plot shows the individual and the violin plot shows the distribution. The red line
713 shows the linear regression line between the imputed dosage and true genotype. In (B–
714 E), MARE is 0.473, 0.274, 0.328, and 0.187, and β_{imp} is 0.336, 0.097, 0.439, and 0.717.
715 In (G–J), MARE is 0.139, 0.072, 0.071, and 0.085, and β_{imp} is 0.843, 0.565, 0.694, and
716 0.675.

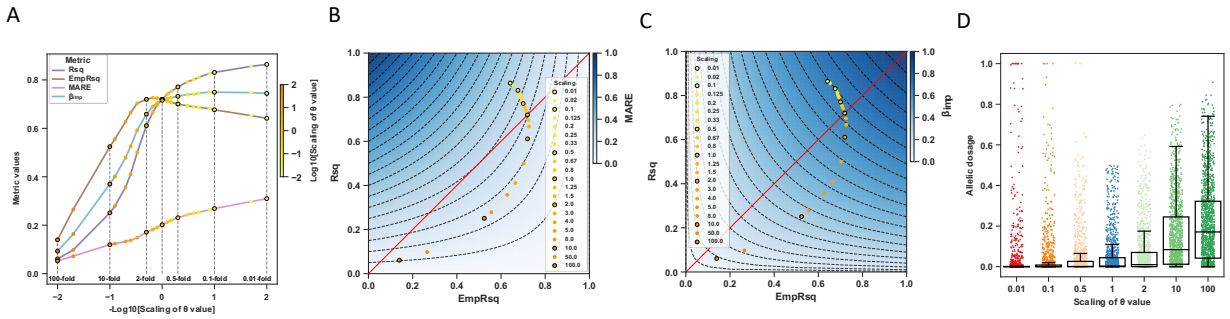
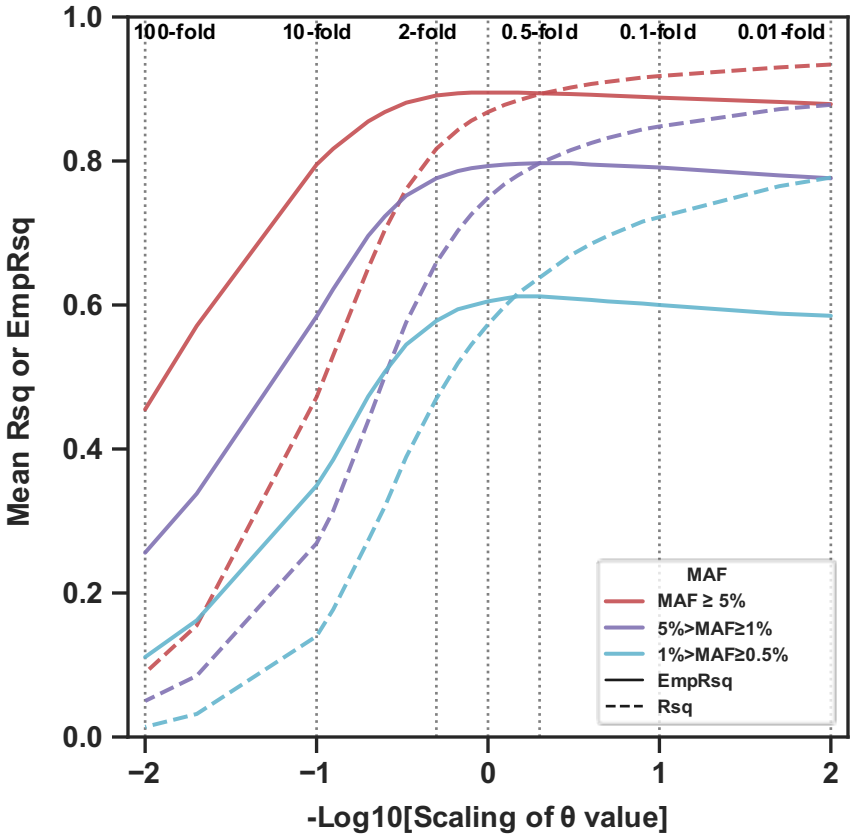


Figure 4. Changes in the metrics value and the imputed allelic dosage with different scalings of the θ value. (A) Rs10410162 is used as an example to visualize the changes in Rsq, EmpRsq, MARE, and β_{imp} (y-axis) with the 21 scalings of the θ value (x-axis). To have a better illustration, the x-axis is in the minus log scale. The scalings of 100, 10, 2, 1, 0.5, 0.1, and 0.01-fold are circled, marked with vertical dashed lines, and labeled. The other scaling folds are shown as the scatters, with the color key denoting the fold in a log scale. (B–C) Rs10410162 is used as an example to visualize the deviation between Rsq, EmpRsq, and MARE (B) or β_{imp} (C). The red line shows that Rsq equals dosage r^2, and the color keys show the theoretical MARE or β_{imp} values. The dashed lines represent the contours of MARE or β_{imp}, and adjacent contours have a value difference of 0.05. The 21 scatters are the same as (A). (D) The plot shows the imputed allelic dosage of a randomly selected target haplotype (chr19) with the 7 scalings of the θ value. The scatters show variants. The boxes show the median, upper (75%), and lower (25%) quartiles. The whiskers show the 1.5-fold of the interquartile range (IQR) extended from the upper or lower quartile if a value exceeds them; otherwise, they show the maximum or minimum value. Only variants with imputed allelic dosage < 0.5 and 0.3 < EmpRsq < 0.8 when using the 1-fold θ value are shown.



735

736 **Figure 5. Rsq and EmpRsq using the 1KGP panel with different scalings of the θ**

737 **value.** The plot shows the mean Rsq and EmpRsq of each imputation result, stratified

738 by the MAF of the array. The scalings of 100, 10, 2, 1, 0.5, 0.1, and 0.01-fold are

739 marked with vertical dashed lines and labeled.

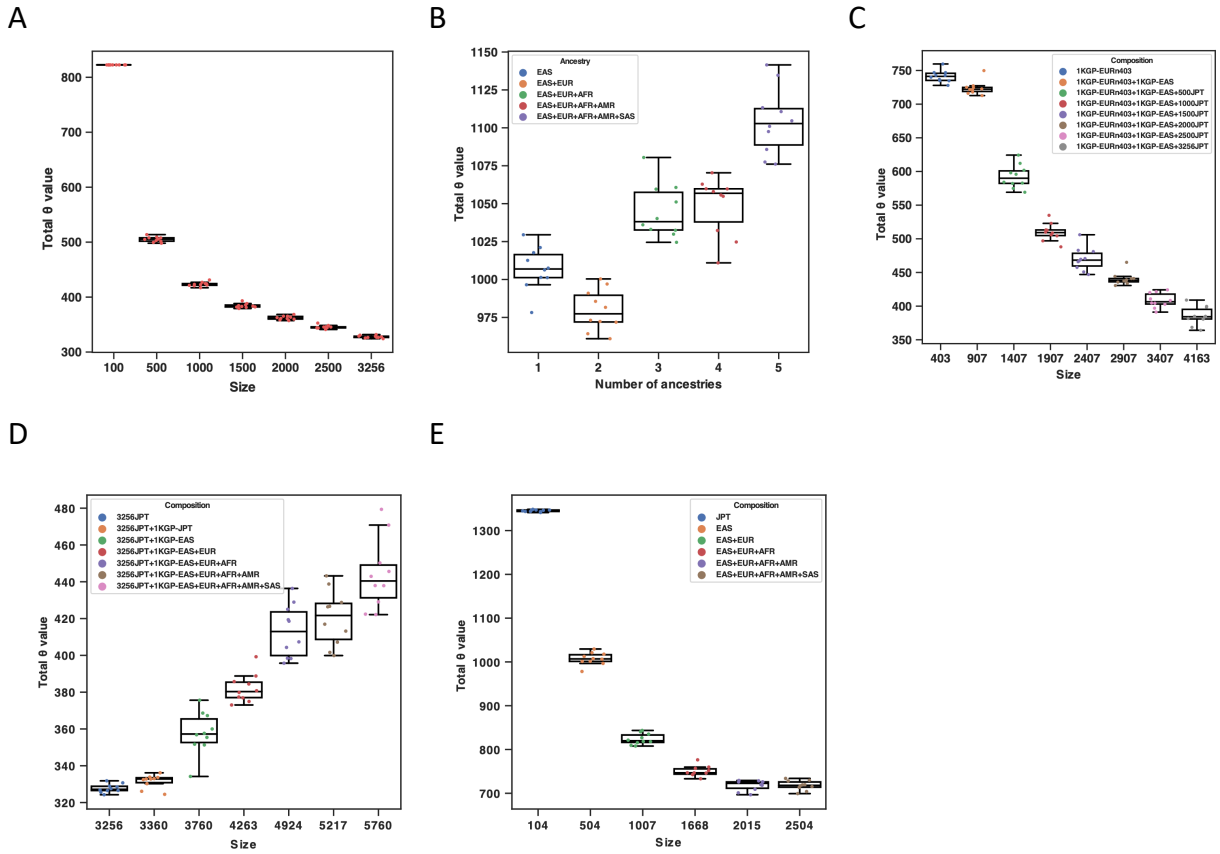


Figure 6. Total θ value of the simulated reference panels. The Supplementary Figure hows the total θ value of simulated reference panels, estimated by Minimac3. (A) JPT-only panels with different sample sizes. (B) Fixed-size panels (504) with 1–5 levels of ancestral diversity, as shown in the legend. (C) Adding large-size (504–3,760) EAS samples to a small-size (403) EUR panel. (D) Adding 1KGP samples to a large-size (3,256) JPT panel. (E) Simultaneously increasing the size and ancestral diversity using 1–5 ancestries in the 1KGP. In each plot, the scatters show the values of 10 runs. The boxes show the median, upper (75%), and lower (25%) quartiles. The whiskers show the 1.5-fold interquartile range (IQR) extended from the upper or lower quartile if a value exceeds them; otherwise, they show the maximum or minimum value. In (C),

751 EURn403 represents the 403 EUR. In (C) and (D), 500–3256JPT represents the

752 number of JPT samples in the panel.

753

754 **Tables**

755

756 **Table 1. Characteristics of the reference panels and target samples used in this**

757 **study, and the number of variants imputed.**

Reference panel	Reference panel size	Size of East Asian subjects	Percentage of EAS subjects	Sequencing depth	Target sample size	Target sample selection criteria	Number of imputed variants on autosomes
1KGP	2,504	504	20.13%	4–8 ×	180,882	All EAS in BBJ -180k passed QC	47,109,465
BBJ1k	3,541	1,541	43.52%	1KGP + 30 ×	179,947	Same as above. Exclude samples in the reference panel	59,387,070
JEWEL3k	5,760	3,760	65.28%	1KGP + 15–30 ×	177,930	Same as above. Exclude samples in the reference panel	69,076,783
TOPMed	97,256	1,184	1.22%	38 ×	180,882	All EAS in BBJ-180k passed QC	291,214,407

758

759 **Table 2. Reference panel description, the total θ value estimated by Minimac3, the**

760 **mean EmpRsq, and the mean Rsq using the simulated EUR-EAS reference panels**

761 **with different sizes and EUR proportions, and 100 EUR as the target sample.**

Reference panel	Reference panel size	EUR (%)	Total θ value	MAF ≥ 5%		5% ≥ MAF > 1%	
				EmpRsq	Rsq	EmpRsq	Rsq
1KGP-EURn403	403	100	740.0	0.849	0.797	0.75	0.679
1KGP-EURn403+1KGP-EAS	907	44.43	722.7	0.841	0.795	0.743	0.7
1KGP-EURn403+1KGP-EAS+500JPT	1,407	28.64	595.9	0.839	0.814	0.744	0.736
1KGP-EURn403+1KGP-EAS+1000JPT	1,907	21.13	534.9	0.836	0.822	0.74	0.752
1KGP-EURn403+1KGP-EAS+1500JPT	2,407	16.74	466.0	0.834	0.831	0.739	0.767
1KGP-EURn403+1KGP-EAS+2000JPT	2,907	13.86	437.5	0.833	0.835	0.737	0.774
1KGP-EURn403+1KGP-EAS+2500JPT	3,407	11.83	408.4	0.831	0.839	0.734	0.78
1KGP-EURn403+1KGP-EAS+3256JPT	4,163	9.68	364.1	0.829	0.844	0.732	0.788

762 EURn403 represents the 403 EUR; 1KGP-EAS represents the 504 EAS; 500–3256JPT

763 represents the number of JPT samples in the reference panel. Minor allele frequency

764 (MAF) is determined by the genotyping array.

765 **Table 3. Reference panel description, the total θ value estimated by Minimac3, the**

766 **mean EmpRsq, and the mean Rsq using the simulated JPT-1KGP reference**

767 **panels with different sizes and ancestral diversity, and WGS_{993} as the target**

768 **sample.**

Reference panel	Reference panel size	EAS (%)	Total θ value	MAF ≥ 5%		5% > MAF ≥ 1%		1% > MAF ≥ 0.5%	
				EmpRsq	Rsq	EmpRsq	Rsq	EmpRsq	Rsq
3256JPT	3,256	100	326.9	0.959	0.949	0.921	0.91	0.861	0.841
3256JPT+1KGP-JPT	3,360	100	336.2	0.959	0.949	0.921	0.909	0.861	0.839
3256JPT+1KGP-EAS	3,760	100	334.2	0.959	0.948	0.923	0.908	0.859	0.834
3256JPT+1KGP-EAS+EUR	4,263	88.2	377.4	0.959	0.945	0.921	0.899	0.855	0.816
3256JPT+1KGP-EAS+EUR+AFR	4,924	76.36	395.7	0.958	0.944	0.92	0.892	0.857	0.8
3256JPT+1KGP-EAS+EUR+AFR+AMR	5,217	72.07	426.4	0.958	0.942	0.919	0.887	0.857	0.793
3256JPT+1KGP-All	5,760	65.28	422.4	0.958	0.942	0.919	0.886	0.854	0.793

769 The 3,256 WGS samples of BBJ are included in all panels. 1KGP- followed by the

770 ancestry represents the 1KGP subset. Minor allele frequency (MAF) is determined

771 by the genotyping array.