Characterizing the quality metric in genotype imputation

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, 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)

Key Words

Genotype imputation, Rsq, East Asian, Under-represented population, TOPMed

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 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 obtaining the accurate imputed dosage and unbiased Rsq.

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-

68 Omics for Precision Medicine (TOPMed),[16,17] greatly improved the imputation accuracy
69 and increased the number of high-Rsq variants in European (EUR), African (AFR), and
70 Admixed American (AMR).[10,18]
71
72 However, constructing large reference panels by combining diverse ancestries is not
73 always beneficial. Small-size population-specific panels, such as Norwegian and
74 Estonian panels,[19,20] achieve similar performance to the HRC panel in EUR. Moreover,
75 even though the HRC panel includes all 1KGP samples, the imputation accuracy in non-
76 EUR populations can be inferior to that of the 1KGP panel alone.[10,21] Bai et al. found
77 that adding 27 samples from a different ancestry to the Han Chinese (1KGP-CHB, size
78 = 103) resulted in better imputation accuracy than using the 1KGP-CHB or 1KGP
79 panel.[22] In studies of Asian, combining population-specific WGS datasets with the 1KGP
80 had improved the imputation accuracy in some instances,[23–25] but not in others.[26–28] So
81 far, the reason for these discrepancies is unclear.
82
83 Two sets of parameters are used in the HMM-based imputation algorithms. Transition
84 probability is governed by the template switching rate (θ) between adjacent markers,
85 which models the recombination and relatedness, and emission probability is
86 determined by the error rate (ε) for each marker, which models genotyping error, gene
87 conversion, and recurrent mutation.[3,7] These parameters are crucial to the HMM for
88 calculating matching probabilities between the template and target during the imputation
89 process, but dosage r^2 is robust to these parameters.[3,5] Many studies have suggested

90 different Rsq or INFO thresholds to achieve a similar dosage r^2,[29,30] making it difficult to
91 ascertain the relationship between the imputation process, Rsq, and dosage r^2.
92
93 In this study, we imputed the East Asian (EAS) using the TOPMed reference panel
94 (EAS comprised 1.22% of the total samples) and found that Rsq substantially
95 overestimated dosage r^2. We introduced novel variance component analysis for Rsq
96 and analytically investigated why Rsq was overestimated and characterized the
97 relationship between the template switching rate used in the HMM, quality metrics, and
98 imputed dosage. Further, we evaluated the θ value, Rsq, dosage r^2, the deviation
99 between them, and the number of high-Rsq variants from matched or distant ancestry in
100 cases where the target ancestry was the major or minor component of the multi-
101 ancestry panels.
102

103 **Material and methods**

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

105 Cohort specifications of Biobank Japan (BBJ-180k), genotyping, and QC are provided in
106 **Method S1**. For imputation using the TOPMed panel, we lifted the coordinates of the
107 genotyping array from hg19 to hg38 using LiftOver.[31] Palindromic variants
108 (reference/alternative (ref/alt) alleles were G/C or A/T) and variants not in the TOPMed
109 freeze5b (https://bravo.sph.umich.edu/freeze5/hg38/) were excluded. The ref/alt alleles
110 were swapped and/or reverse-complemented according to the hg38 reference
111 sequence. A total of 515,587 autosomal variants remained.

112

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

114 EAGLE v2.4.1 was used for pre-phasing without an external reference.[32] The 1KGP
115 reference panel was downloaded from https://genome.sph.umich.edu/wiki/Minimac3.
116 Methods used to construct the population-specific panels (named the BBJ1k and
117 JEWEL3k) are provided in **Method S2**. We used an in-house server to perform
118 imputation using the 1KGP, BBJ1k, and JEWEL3k panels. Minimac3 v2.0.3 was used to
119 estimate the HMM parameters and prepare m3vcf files.[3] Minimac4 v1.0.2 was used for
120 imputation.[33] Imputation using the TOPMed and HRC panel were performed using the
121 TOPMed and Michigan imputation servers,[3,17] respectively.

122

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

124 Rsq from the Minimac4's info file was extracted for the subsequent analyses. Variants
125 with Rsq < 0.3 were removed.[34] Imputation performance was empirically evaluated
126 using 993 samples with WGS (named WGS_{993}). QC and processing are elaborated in
127 **Method S3**. Coverage of a reference panel was the fraction of variants in WGS_{993} that
128 could be imputed with Rsq ≥ 0.3. Dosage r^2 was the squared Pearson correlation
129 coefficient between the imputed dosage and true genotype (encoded as 0, 1, or 2).
130 Minor allele frequency (MAF) and minor allele count (MAC) of WGS_{993} were used to
131 stratify variants into bins.

132

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.

Rsq is the ratio between Var(***y***) and $p(1-p)$, where *p* is the alternative allele frequency (AAF) in the imputed dataset.[3,7] Then,

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

(Equation 3)

where SS_{res} is the residual sum of squares that follows $SS_{tot} = SS_{reg} + SS_{res}$, and *n* is the number of imputed haplotypes. Hence, Rsq comprises two parts: regression-related and residual-related. We define $\frac{SS_{res}/n}{p(1-p)}$ as MAF-Adjusted-Residual-Error (MARE). Then:

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

(Equation 4)

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

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

(Equation 5)

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

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

(Equation 6)

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

(Equation 7)

In Equation 7, $\hat{\beta}_{imp}$ is negative if ***x*** and ***y*** are negatively correlated. We do not consider that situation. Hence, each combination of Rsq and EmpRsq indicates specific values of *MARE* and $\hat{\beta}_{imp}$. Hereafter, we refer to $\hat{\beta}_{imp}$ as the β_{imp} metric and *MARE* as the MARE metric.

Because the WGS dataset is unphased diploid data, dosage r^2 calculated from that will be slightly different from EmpRsq. **Note S1** provides the detailed methods to calculate

175 MARE and β_{imp} from haploid or diploid data, and the concordance between values
176 obtained using Equations 6–7 and calculated from the imputed dosage in both haploid
177 and diploid cases.
178

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

180 The 1KGP vcf files were downloaded from the Minimac3 website (see above).
181 Parameters were estimated using Minimac3. The θ value was extracted (denoted as
182 "Recom" in the m3vcf file),[3] scaled by 21 folds manually (0.01–100), and replaced in the
183 original file. A brief explanation of the θ value is provided in **Note S2**. The array data of
184 WGS_{993}, a subset of BBJ-180k, was used as the target sample. The imputation
185 procedure was the same as described above.
186

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

188 We obtained the imputed allelic dosage (LooDosage) from Minimac4's empiricalDose
189 file (by turning the "--meta" option on).[35] It was derived from the leave-one-out (LOO)
190 method by hiding the markers on the array during the imputation. Rsq, EmpRsq, MARE,
191 and β_{imp} were calculated from the LooDosage and array data. MAF was obtained from
192 the array data.
193

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

195 We sampled subsets from the 1KGP and JEWEL3k, shuffled the sample order 10 times,
196 and created new vcf files using bcftools v1.14 (https://samtools.github.io/bcftools/).[36] We
197 estimated the parameters using Minimac3. **Method S4** provides the detailed methods.
198 The total θ value along chr19 (by summing the θ values between adjacent markers) was
199 used to evaluate the reference panel size and ancestral diversity's impact. A discussion
200 of the θ value qualification is provided in **Note S3**.

201

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

203 We randomly sampled 403 individuals (named 1KGP-EUR_{403}) from the 1KGP-EUR and
204 combined them with the 1KGP-EAS and 6 subsets (size = 500, 1,000, 1,500, 2,000,
205 2,500, and 3,256) of the 3,256 JPT WGS samples (named as JPT_{3256}) in JEWEL3k.
206 The remaining 100 individuals in the 1KGP-EUR were used as the target sample. We
207 extracted the 10,375 polymorphic variants (chr19) on the Illumina Global Screening
208 Array v3.0 (https://support.illumina.com/content/dam/illumina-
209 support/documents/downloads/productfiles/global-screening-array-24/v3-0/infinium-
210 global-screening-array-24-v3-0-a1-b151-rsids.zip) to simulate the genotyping array.
211 Parameter estimation and imputation were the same as above.

212

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

214 We sampled 7 subsets (size = 100, 500, 1,000, 1,500, 2,000, 2,500, and 3,256) from the
215 JPT_{3256} and then combined the JPT_{3256} with 6 subsets (1KGP-JPT and 1–5 ancestries)
216 of the 1KGP. WGS_{993} was used as the target. The other processing methods were the
217 same as above.
218

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

220 Haploid dosage (HDS) is the imputed alternative allele's allelic dosage at the haploid
221 level, which could be obtained by the "--format HDS" option in Minimac4. Briefly, it is in
222 a per-variant and per-individual manner and a higher value represents an imputed
223 alternative allele is more certain. We quantified the θ value's impact on the imputed
224 dosages by the number of confident alleles (HDS > 0.9). Rsq was used to judge how
225 many high-Rsq (Rsq > 0.7) variants could be passed to the downstream analyses.
226

227 Furthermore, we determined how many confident alleles and high-Rsq variants could be
228 obtained only from the additional distant ancestries in the multi-ancestry reference
229 panel. We defined ancestry-specific variants as follows: EUR-only variants only existed
230 in the $1KGP\text{-}EUR_{403}$, and it has no non-EUR variants. JPT_{3256}-only and 1KGP-EAS-only
231 variants only existed in the JPT_{3256} and 1KGP-EAS, respectively. Non-EAS variants
232 were not found in the JPT_{3256} or 1KGP-EAS.
233

234 **Replication of the TOPMed imputation pipeline**

235 We followed the TOPMed imputation pipeline (Accessed on Oct. 15, 2022,
236 https://topmedimpute.readthedocs.io/) and used the 1KGP and JEWEL3k reference
237 panels. Specifically, the HapMap2 genetic map was used as a reference for the θ value
238 instead of estimating it using Minimac3. WGS_{993} was used as the target. As the θ value
239 was not outputted by default, we modified the Minimac4's source code to obtain the
240 transformed θ value (**Method S5**).

241

242 **Results**

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

244 The BBJ-180k was imputed using the TOPMed, 1KGP, BBJ1k, and JEWEL3k reference
245 panels (**Figure 1A**). Characteristics of each panel and the target sample are listed in
246 **Table 1**. We categorized the imputed variants by MAF and Rsq in each imputed
247 dataset. With more EAS samples in the panel, more variants with low MAF passed each
248 Rsq threshold (**Figure 2A** and **Table S1**). In addition to the absolute number, unique
249 variants were imputed from each panel (**Figure 2B**). These results reproduced the
250 benefits of using large and different reference panels.[19]

251

252 We then used WGS_{993} to empirically evaluate the imputation performance. MAF and
253 MAC of WGS_{993} were used to categorize the variants into six bins: common (MAF ≥

254 5%), low-frequency (5% > MAF ≥ 1% and 1% > MAF ≥ 0.5%), rare (0.5% > MAF >
255 0.1%), doubleton (MAC = 2), and singleton (MAC = 1). For common single nucleotide
256 polymorphisms (SNVs), the TOPMed imputation result showed the highest coverage
257 (0.950), whereas BBJ1k (0.890) and JEWEL3k (0.907) were lower than 1KGP (0.948).
258 This loss of variants was possibly due to the additional QC steps in combining the
259 1KGP with Japanese WGS (**Method S2**). As MAF decreased, the coverage also
260 decreased; however, more EAS samples in the reference panel mitigated this decrease,
261 as expected (**Figure 2C**).

262

263 The 155,297 shared SNVs (on chr19) imputed by the four panels were used to compare
264 imputation accuracy. The mean dosage r^2 using the TOPMed panel was between those
265 of 1KGP and BBJ1k (**Figure 2D**). However, Rsq was substantially upward biased only
266 in the TOPMed imputation result (**Figure 2E**). Using all SNVs and short insertions and
267 deletions (indels) with Rsq ≥ 0.3, although with higher coverage, the TOPMed
268 imputation result showed an inferior dosage r^2 than 1KGP (**Figure S1**).

269

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

271 The deviation between Rsq and dosage r^2 was persistent in all MAF bins only when
272 using the TOPMed panel (**Figure S2**), which indicated a potential systematic bias from
273 the imputation pipeline or the reference panel. To justify our observation, we analytically
274 derived the relationship between Rsq and dosage r^2 (**Methods**). Two novel metrics,

MARE and β_{imp}, were introduced to quantify the deviation. MARE, a MAF-adjusted form of residual error, takes a value between 0 and 1, and increases with SS_{res}. β_{imp} describes the distinguishability between the mean imputed dosage of each true genotype group. Rsq is the ratio between observed and expected variance and dosage r^2 shows the correlation. Under the assumption of "well-calibration" (the posterior allele probability from imputation equals the expected true allele dose), Rsq equals dosage r^2.[10] Our analytical derivation of the relationship between Rsq and dosage r^2 did not assume the "well-calibration" (**Figure S3–7** and **Note S1**). Any two of Rsq, dosage r^2, MARE, and β_{imp} could entirely quantify the relationship between the imputed dosage and true genotype and determine the other two metrics (Equations 6–7), whereas Rsq or dosage r^2 alone could not.

To exhibit the relationships, we plotted the theoretical values of MARE and β_{imp} on the coordinates of Rsq and dosage r^2 (**Figure 3A** and **3F**). It clearly showed that the overestimated Rsq was accompanied by a higher MARE (**Figure 3A**). We used 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)[37] (**Figure 3B**), compared to the other three panels (**Figure 3C–E**). The high certainty increased Var(***y***) and Rsq. In **Note S4**, 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

296 heterozygotes were incorrectly imputed with a dosage of approximately 0 in the
297 TOPMed imputation, causing higher MARE and Rsq, and an even low dosage r^2.

298

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

308

309 We categorized MARE into Rsq bins and β_{imp} into dosage r^2 bins to compare them
310 between reference panels and to the expected values when assuming Rsq equals
311 dosage r^2. The mean MARE of the TOPMed result was above the expected for Rsq bins
312 0.35–0.9, and the mean β_{imp} of the 1KGP result was below those of the other panels
313 (**Figure S8**). These results suggested that the TOPMed result was more certain and
314 might contain more wrongly imputed genotypes (**Figure 3B** and **3G**), while the 1KGP
315 result might possess higher shrinkage in the imputed dosage, as shown above (**Figure**
316 **3C** and **3H**).

317

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

319 The high certainty suggested an overconfident matching between the reference panel
320 and target sample. We investigated how the template switching rate (θ) used in the
321 HMM affects the imputed-allele certainty using the 1KGP reference panel and 21
322 scalings of the θ value (0.01–100-fold of that estimated by Minimac3; **Methods**; **Figure**
323 **1B**). Low θ values caused Rsq > EmpRsq (EmpRsq is the alternate of dosage r^2;
324 **Methods**) (**Figure S9A–C**), while high θ values caused Rsq < EmpRsq (**Figure S9E–**
325 **G**). Using rs10410162 as an example, Rsq and MARE increased as the θ value
326 decreased (**Figure 4A**), with increasing certainty of the imputed alleles (**Figure S10A–**
327 **C**) and the deviation toward Rsq > EmpRsq (**Figure 4B**). As the θ value increased, the
328 imputed allelic dosages were shrunk to the AAF (0.351) (**Figure S10E–G)**, resulting in a
329 more drastic decrease in Rsq than EmpRsq (**Figure 4A**), thereby causing EmpRsq >
330 Rsq. Notably, EmpRsq and β_{imp} were roughly maintained unless the θ value was scaled
331 up by 2-fold or higher (**Figure 4A** and **4C**), suggesting EmpRsq was insensitive to
332 altering the θ value, particularly the downscaling. In contrast, Rsq was sensitive to the θ
333 value and consistently increased with downscaling of the θ value, leading to the
334 deviation between Rsq and EmpRsq.

335

336 We also showed that low θ values increased the imputed-allele certainty using a
337 randomly selected target haplotype **(Figure 4D** and **Figure S11).** High certainty might

338 cause some variants to be wrongly imputed as the opposite allele (**Figure S12**). An
339 explanation is provided in **Note S5**. Such imputed-dosage properties were also revealed
340 by the mean MARE and β_{imp} across all variants on chr19 (**Figure S13**). Taken together,
341 the simulated imputation using an extremely low θ value mimics the high certainty and
342 Rsq overestimation in the TOPMed imputation result in EAS (**Figure 3**).

343

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

345 Using the modified 1KGP panels, although the mean EmpRsq was insensitive to the
346 scaling of the θ value (maximum difference of 0.034 for the θ value scaling between
347 0.01 and 2-fold; **Figure 5**), low θ value increased Rsq (**Figure 5**). We evaluated the
348 number of confident alleles (HDS > 0.9) and high-Rsq variants (Rsq > 0.7) (**Methods**).
349 Downscaling the θ value increased the number of confident alleles and high-Rsq
350 variants (**Table S2**). When the θ value was 0.5-fold, the number of confident alleles and
351 high-Rsq variants increased by 5.46% and 8.89%, respectively, and if the θ value was
352 2-fold, these numbers decreased by 11.8% and 13.8%, respectively. These results
353 indicated that the θ value shaped the imputed allelic dosage and changed Rsq and the
354 number of high-Rsq variants while leaving EmpRsq almost the same.

355

356 **Reference panel and θ estimates**

357 We evaluated how Minimac3's parameter estimation changed with the composition of
358 the reference panel (**Method S4**, **Figure S14**, and **Note S6**). The θ estimates
359 decreased with the sample size of a single ancestry and increased with ancestral
360 diversity when size was fixed (**Figure 6A** and **6B**). On pooling samples from different
361 ancestries together at a small panel size, the θ estimates still decreased as the panel
362 size increased (**Figure 6C**); however, at a larger panel size, it increased with
363 simultaneously increasing the panel size and ancestral diversity (**Figure 6D**),
364 suggesting that the θ estimates were in a trade-off between the panel size and ancestral
365 diversity (**Figure 6E**). **Note S6** explains these effects.

366

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

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

371

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

373 We simulated 8 EUR-EAS reference panels and used 100 EUR samples as the target
374 (**Methods**). As the panel size increased and θ value decreased (**Table 2**), Rsq and
375 MARE were upwardly biased, as expected (**Figure S15** and **S16**). The mean EmpRsq

376 decreased consistently with the addition of EAS samples to the reference panel (**Table**
377 **2**). However, only a marginal difference was observed (maximum difference of 0.020).

378

379 There were 114,606 EUR-only and 696,017 non-EUR variants (**Methods**). As the panel
380 size increased and θ value decreased, the number of confident alleles increased from
381 84,797 to 101,116 and from 0 to 458 for EUR-only and non-EUR variants, respectively
382 (**Table S3**). The number of high-Rsq variants increased from 163,468 to 200,161,
383 18,646 to 27,941, and 0 to 521 for all, EUR-only, and non-EUR variants, respectively
384 (**Table S3**). These results revealed that a lower θ value increases the number of
385 confident alleles and high-Rsq variants, without improving the EmpRsq (**Table S3**).
386 Meanwhile, even if the reference panel comprised 90.3% EAS samples, non-EUR
387 variants only comprised 0.26% of the total variants in the imputation result (when setting
388 a cutoff of Rsq > 0.7). The majority of variants gaining from the larger panel when
389 imputing the under-represented ancestry was because of the lower θ value.

390

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

392 We simulated 13 JPT-1KGP reference panels and used WGS_{993} as the target
393 (**Methods**). When only JPT samples were in the panel, the mean EmpRsq and Rsq
394 increased with the panel size, as expected (**Figure S17**). When combined with the
395 1KGP subsets, the mean EmpRsq was highest when using JPT_{3256}+1KGP-EAS for
396 variants with MAF ≥ 1% and JPT_{3256}+1KGP-JPT for variants with 1% > MAF ≥ 0.5%

397 (**Table 3**). Adding other ancestries decreased the mean EmpRsq marginally (maximum
398 difference < 0.01 for all MAF categories). The mean Rsq was the highest when using
399 JPT_{3256} and decreased with the addition of more ancestries, with maximum differences
400 of 0.007, 0.024, and 0.048 for variants with MAF ≥ 5%, 5% > MAF ≥ 1%, and 1% > MAF
401 ≥ 0.5%, respectively (**Table 3**). None of the imputation results showed a noticeable
402 deviation in the MARE and β_{imp}. However, the MARE and β_{imp} of the JPT_{3256},
403 JPT_{3256}+1KGP-JPT, and JPT_{3256}+1KGP-EAS results were closer to the expected values
404 (**Figure S18**), while adding other ancestries made MARE and β_{imp} correspond to the
405 condition of using a higher θ value.

406

407 There were 74,490 JPT_{3256}-only, 7,627 1KGP-EAS-only, and 495,489 non-EAS variants
408 (**Methods**). From JPT_{3256} to JPT_{3256}+1KGP, the number of confident alleles decreased
409 from 24,599 to 23,666, increased from 0 to 131, and increased from 0 to 254 for these
410 three groups of variants (**Table S4**). The number of high-Rsq variants decreased from
411 274,343 to 264,221 (a decrease of 3.69%) when using JPT_{3256} or JPT_{3256}+1KGP (**Table**
412 **S4**). Only 265 non-EAS variants reached an Rsq > 0.7 when using JPT_{3256}+1KGP,
413 which was 0.10% of the total variants. Hence, combining with the 1KGP caused fewer
414 confident alleles and high-Rsq variants to remain in the imputation results, as expected
415 from using higher θ values. Both scenarios suggest that the distant ancestry in the
416 reference panel affects the θ estimates and the number of high-Rsq variants. However,
417 the expanded haplotypes and variant sets from distant ancestries provided only a few
418 additional variants.

419

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

421 We used the TOPMed imputation pipeline (**Methods**) and observed an upward
422 deviation of Rsq, particularly when using the 1KGP panel (**Figure S19**). The θ value
423 transformed from the genetic map was 0.15- and 0.27-fold of that estimated by
424 Minimac3 when using the 1KGP and JEWEL3k panels, respectively (**Table 4**).

425

426 As the θ estimates decreased with the size of the major ancestry (**Figure 6C**), it would
427 be underestimated for the minor ancestry in the multi-ancestry panel regardless of using
428 the genetic map or estimating it by the Mimimac3. One example was the HRC panel,
429 which did not use the genetic map but induced Rsq overestimation in the BBJ-180k
430 (**Figure S20**). Thus, caution may be required when imputing the under-represented
431 population using large ancestry-imbalanced panels under the current framework.

432

433 **Discussion**

434 The deviation between Rsq and dosage r^2 in the imputation result is raised from the
435 imputed dosage, and has been widely observed using different reference panels and
436 software.[9,15,38] One reason was that the θ value used in the HMM does not fit the panel
437 size, ancestral components in the reference panel, and the target population. When
438 using the multi-ancestry reference panel, distant ancestries affect the θ estimates (in

439 Minimac3/4), then the imputed dosage and Rsq. Our simulations indicated that the
440 lower θ value used by the larger panel could increase the imputed-genotype certainty
441 and the number of high-Rsq variants but not dosage r^2. It may cause confusing
442 benchmarking results and increase the chance of false positives in association tests.
443 Ferwerda et al. reported height and body mass index association signals in an ethnically
444 diverse cohort only when imputing against the TOPMed panel.[39] Bai et al. reported the
445 highest Rsq but lowest dosage r^2 when imputing the Han Chinese using the HRC panel,
446 compared to the 1KGP and a population-specific panel,[22] similar to what we had
447 observed in Japanese. In addition to simply comparing dosage r^2 or Rsq, we provided
448 the changes in those metrics, the θ estimates, imputed-dosage certainty, and the
449 number of high-Rsq variants passing to downstream analyses. We provided the script
450 (https://github.com/shimaomao26/impumetric) to allow users to check their results using
451 the leave-one-out imputation of Minimac4 (external WGS not required) or additional
452 WGS data.

453

454 The TOPMed imputation pipeline used the θ value transformed from the HapMap2
455 genetic map by assuming 0.01 centimorgan (cM) corresponds 1% switching rate. The
456 switching rate would thereby be fixed given the variant's base pair position and cM. Two
457 reasons may explain why it works well. First, the genetic map did not significantly
458 change the imputation accuracy, as verified in a Finnish study.[40] We had also shown
459 that dosage r^2 was insensitive to the θ value. Second, the θ value was only
460 underestimated for the EAS (size = 1,184), but the sizes of EUR, AFR, and AMR (size =

461 17,085–47,159) might fit this value. Previous studies reported that Rsq in the TOPMed
462 imputation result was sometimes misleadingly high in AFR [38] and even EUR.[41] A recent
463 paper reported that the TOPMed panel was more robust to the low-density genotyping
464 array than the HRC and 1KGP panels.[18] These results also suggested that a fixed low θ
465 value might be used by the TOPMed pipeline, as we have inferred. Further studies are
466 warranted to comprehensively elucidate the impacts on genetic studies.

467

468 Our results suggest that avoiding ancestral diversity is best when more than 3,000 WGS
469 samples are available to construct a JPT reference panel. Increased diversity would
470 then only have a marginal impact on dosage r^2 but would affect the individual's imputed
471 dosage and cause fewer variants to pass a predetermined Rsq filter. Zhang et al. and
472 Cong et al. similarly observed that adding the 1KGP to about 3,000 Chinese WGS
473 samples would neither benefit nor harm dosage r^2.[26,27] We focused on explaining the
474 reason in this work. Further studies are warranted to take advantage of combining
475 population-specific and public WGS datasets while avoiding our identified problems.

476

477 Our study has several limitations. First, our simulations of the reference panel were
478 study-specific. As discussed, the θ estimates decreased with the panel size and
479 increased with the ancestral diversity. However, a larger panel typically increases the
480 diversity simultaneously. Therefore, the θ value and imputation result need to be
481 discussed case by case. This limitation also implies that the general experience may not
482 work well for a new reference panel and target dataset. Second, in the simulations of
483 multi-ancestry reference panels (scenarios 1 and 2), WGS datasets were merged using

484 IMPUTE2.[9] We did not investigate the impact of IMPUTE2 but treated the merged
485 dataset similarly to the WGS dataset. It might underestimate the number of high-Rsq
486 variants from the distant ancestry (**Note S7)**. Our results using the 1KGP (not modified
487 by IMPUTE2) also revealed that only a limited number of confident alleles and high-Rsq
488 variants in the imputation result were from the distant ancestry (**Note S7** and **Table S2**).
489 Therefore, the conclusion that distant ancestry in the reference panel affects the θ
490 estimates rather than providing high-Rsq variants would be valid. However, further
491 studies should be conducted to determine the impact of the panel-merging method and
492 the net gain of variants from distant ancestries.
493
494 In summary, we explained that the HMM parameter could be a potential reason for
495 inaccurate Rsq and inferior dosage r^2 when using large multi-ancestry reference panels.
496 This is also the first study of the relationship between the template switching rate,
497 imputed-genotype certainty, Rsq, and dosage r^2. We envision that our methods and
498 conclusions could provide insights into benchmarking studies, construction of reference
499 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

522 **Data and code availability**

523 We provided the tool to replicate the metrics used in this work
524 (https://github.com/shimaomao26/impumetric). Other scripts were also deposited there.
525 Genotype data for BBJ and the TOPMed imputation results were deposited at NBDC
526 Human Database (research ID: hum0014 and hum0311, respectively).
527

528 **Appendix**

529 **Imputation quality metrics**

530 Rsq: The standard quality metric in Minimac. The ratio between observed variance and
531 expected variance of imputed allelic dosage.
532 INFO: The standard quality metric in IMPUTE. The ratio between observed information
533 and complete information of imputed genotype distribution.
534 Dosage r^2: The squared Pearson correlation between imputed dosage and true
535 genotype.
536 EmpRsq: The squared Pearson correlation between imputed allelic dosage and true
537 allele dose. Provided by Minimac and is only available for the genotyped markers on
538 array.
539 MARE: The minor allele frequency adjusted residual error in linear regression between
540 imputed dosage and true genotype.
541 β_{imp}: The regression coefficient between imputed dosage and true genotype.
542

543 **References**

544 1. Li, N., and Stephens, M. (2003). Modeling linkage disequilibrium and identifying
545 recombination hotspots using single-nucleotide polymorphism data. Genetics *165*,
546 2213–2233. 10.1093/genetics/165.4.2213.
547 2. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G.R.
548 (2012). Fast and accurate genotype imputation in genome-wide association
549 studies through pre-phasing. Nat Genet *44*, 955–959. 10.1038/ng.2354.
550 3. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I.,
551 Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype
552 imputation service and methods. Nat Genet *48*, 1284–1287. 10.1038/ng.3656.
553 4. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A Flexible and Accurate
554 Genotype Imputation Method for the Next Generation of Genome-Wide
555 Association Studies. PLOS Genet *5*, e1000529.
556 10.1371/JOURNAL.PGEN.1000529.
557 5. Browning, B.L., and Browning, S.R. (2016). Genotype Imputation with Millions of
558 Reference Samples. Am J Hum Genet *98*, 116–126. 10.1016/j.ajhg.2015.11.020.
559 6. Li, Y., Sidore, C., Kang, H.M., Boehnke, M., and Abecasis, G.R. (2011). Low-
560 coverage sequencing: Implications for design of complex trait association studies.
561 Genome Res *21*, 940–951. 10.1101/gr.117259.110.
562 7. Li, Y., Willer, C.J., Ding, J., Scheet, P., and Abecasis, G.R. (2010). MaCH: using
563 sequence and genotype data to estimate haplotypes and unobserved genotypes.
564 Genet Epidemiol *34*, 816–834. 10.1002/GEPI.20533.
565 8. Marchini, J., and Howie, B. (2010). Genotype imputation for genome-wide

566 association studies. Nat Rev Genet *11*, 499–511. 10.1038/nrg2796.
567 9. Huang, J., Howie, B., McCarthy, S., Memari, Y., Walter, K., Min, J.L., Danecek,
568 P., Malerba, G., Trabetti, E., Zheng, H.-F., et al. (2015). Improved imputation of
569 low-frequency and rare variants using the UK10K haplotype reference panel. Nat
570 Commun *6*, 8111. 10.1038/ncomms9111.
571 10. Das, S., Abecasis, G.R., and Browning, B.L. (2018). Genotype Imputation from
572 Large Reference Panels. Annu Rev Genomics Hum Genet *19*, 73–96.
573 10.1146/annurev-genom-083117-021602.
574 11. Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A.,
575 Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., et al. (2007). A second
576 generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851–861.
577 10.1038/nature06258.
578 12. Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Bentley, D.R.,
579 Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., Flicek, P., et al. (2015). A
580 global reference for human genetic variation. Nature *526*, 68–74.
581 10.1038/nature15393.
582 13. Huang, L., Li, Y., Singleton, A.B., Hardy, J.A., Abecasis, G., Rosenberg, N.A., and
583 Scheet, P. (2008). Genotype-imputation accuracy across worldwide human
584 populations. Am J Hum Genet *84*, 235–250. 10.1016/j.ajhg.2009.01.013.
585 14. Liu, Q., Cirulli, E.T., Han, Y., Yao, S., Liu, S., and Zhu, Q. (2015). Systematic
586 assessment of imputation performance using the 1000 Genomes reference
587 panels. Brief Bioinform *16*, 549–562. 10.1093/BIB/BBU035.
588 15. Deelen, P., Menelaou, A., Van Leeuwen, E.M., Kanterakis, A., Van Dijk, F.,

589 Medina-Gomez, C., Francioli, L.C., Hottenga, J.J., Karssen, L.C., Estrada, K., et
590 al. (2014). Improved imputation quality of low-frequency and rare variants in
591 European samples using the “Genome of the Netherlands.” Eur J Hum Genet *22*,
592 1321–1326. 10.1038/EJHG.2014.19.
593 16. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A.,
594 Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference
595 panel of 64,976 haplotypes for genotype imputation. Nat Genet *48*, 1279–1283.
596 10.1038/ng.3643.
597 17. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R.,
598 Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2021).
599 Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program.
600 Nature *590*, 290–299. 10.1038/s41586-021-03205-y.
601 18. Hanks, S.C., Forer, L., Schönherr, S., LeFaive, J., Martins, T., Welch, R.,
602 Gagliano Taliun, S.A., Braff, D., Johnsen, J.M., Kenny, E.E., et al. (2022). Extent
603 to which array genotyping and imputation with large reference panels approximate
604 deep whole-genome sequencing. Am J Hum Genet *109*, 1653–1666.
605 10.1016/j.ajhg.2022.07.012.
606 19. Zhou, W., Fritsche, L.G., Das, S., Zhang, H., Nielsen, J.B., Holmen, O.L., Chen,
607 J., Lin, M., Elvestad, M.B., Hveem, K., et al. (2017). Improving power of
608 association tests using multiple sets of imputed genotypes from distributed
609 reference panels. Genet Epidemiol *41*, 744–755. 10.1002/GEPI.22067.
610 20. Mitt, M., Kals, M., Pärn, K., Gabriel, S.B., Lander, E.S., Palotie, A., Ripatti, S.,
611 Morris, A.P., Metspalu, A., Esko, T., et al. (2017). Improved imputation accuracy

612 of rare and low-frequency variants using population-specific high-coverage WGS-
613 based imputation reference panel. Eur J Hum Genet *25*, 869–876.
614 10.1038/ejhg.2017.51.
615 21. Lin, Y., Liu, L., Yang, S., Li, Y., Lin, D., Zhang, X., and Yin, X. (2018). Genotype
616 imputation for Han Chinese population using Haplotype Reference Consortium as
617 reference. Hum Genet *137*, 431–436. 10.1007/S00439-018-1894-Z/FIGURES/3.
618 22. Bai, W.Y., Zhu, X.W., Cong, P.K., Zhang, X.J., Richards, J.B., and Zheng, H.F.
619 (2020). Genotype imputation and reference panel: a systematic evaluation on
620 haplotype size and diversity. Brief Bioinform *21*, 1806–1817.
621 10.1093/BIB/BBZ108.
622 23. Akiyama, M., Ishigaki, K., Sakaue, S., Momozawa, Y., Horikoshi, M., Hirata, M.,
623 Matsuda, K., Ikegawa, S., Takahashi, A., Kanai, M., et al. (2019). Characterizing
624 rare and low-frequency height-associated variants in the Japanese population.
625 Nat Commun *10*, 4393. 10.1038/s41467-019-12276-5.
626 24. Jeon, S., Bhak, Y., Choi, Y., Jeon, Y., Kim, S., Jang, J., Jang, J., Blazyte, A., Kim,
627 C., Kim, Y., et al. (2020). Korean Genome Project: 1094 Korean personal
628 genomes with clinical information. Sci Adv *6*, eaaz7835. 10.1126/sciadv.aaz7835.
629 25. Yoo, S.-K., Kim, C.-U., Kim, H.L., Kim, S., Shin, J.-Y., Kim, N., Yang, J.S.W., Lo,
630 K.-W., Cho, B., Matsuda, F., et al. (2019). NARD: whole-genome reference panel
631 of 1779 Northeast Asians improves imputation accuracy of rare and low-
632 frequency variants. Genome Med *11*, 64. 10.1186/s13073-019-0677-z.
633 26. Cong, P.-K., Bai, W.-Y., Li, J.-C., Yang, M.-Y., Khederzadeh, S., Gai, S.-R., Li, N.,
634 Liu, Y.-H., Yu, S.-H., Zhao, W.-W., et al. (2022). Genomic analyses of 10,376

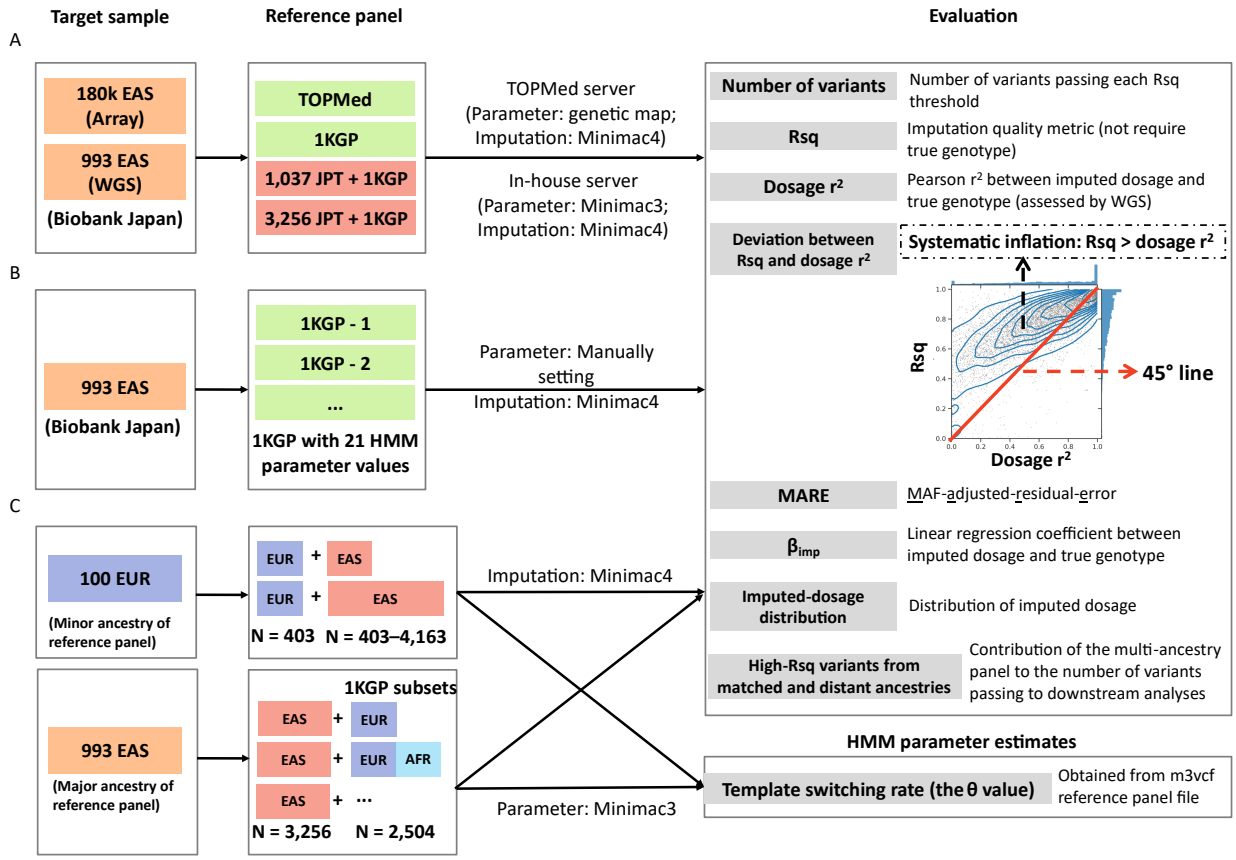
635 individuals in the Westlake BioBank for Chinese (WBBC) pilot project. Nat
636 Commun *13*, 2939. 10.1038/s41467-022-30526-x.
637 27. Zhang, P., Luo, H., Li, Y., Wang, Y., Wang, J., Zheng, Y., Niu, Y., Shi, Y., Zhou,
638 H., Song, T., et al. (2021). NyuWa Genome resource: A deep whole-genome
639 sequencing-based variation profile and reference panel for the Chinese
640 population. Cell Rep *37*, 110017. 10.1016/j.celrep.2021.110017.
641 28. Kawai, Y., Mimori, T., Kojima, K., Nariai, N., Danjoh, I., Saito, R., Yasuda, J.,
642 Yamamoto, M., and Nagasaki, M. (2015). Japonica array: improved genotype
643 imputation by designing a population-specific SNP array with 1070 Japanese
644 individuals. J Hum Genet *60*, 581–587. 10.1038/jhg.2015.68.
645 29. Pistis, G., Porcu, E., Vrieze, S.I., Sidore, C., Steri, M., Danjou, F., Busonero, F.,
646 Mulas, A., Zoledziewska, M., Maschio, A., et al. (2015). Rare variant genotype
647 imputation with thousands of study-specific whole-genome sequences:
648 implications for cost-effective study designs. Eur J Hum Genet *23*, 975–983.
649 10.1038/ejhg.2014.216.
650 30. Roshyara, N.R., Horn, K., Kirsten, H., Ahnert, P., and Scholz, M. (2016).
651 Comparing performance of modern genotype imputation methods in different
652 ethnicities. Sci Rep *6*, 34386. 10.1038/srep34386.
653 31. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M.,
654 and Haussler, and D. (2002). The Human Genome Browser at UCSC. Genome
655 Res *12*, 996–1006. 10.1101/GR.229102.
656 32. Loh, P.-R., Danecek, P., Palamara, P.F., Fuchsberger, C., A Reshef, Y., K
657 Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al.

658 (2016). Reference-based phasing using the Haplotype Reference Consortium
659 panel. Nat Genet *48*, 1443–1448. 10.1038/ng.3679.
660 33. Das, S. (2017). Next Generation of Genotype Imputation Methods.
661 34. Liu, E.Y., Buyske, S., Aragaki, A.K., Peters, U., Boerwinkle, E., Carlson, C., Carty,
662 C., Crawford, D.C., Haessler, J., Hindorff, L.A., et al. (2012). Genotype Imputation
663 of Metabochip SNPs Using a Study-Specific Reference Panel of ~4,000
664 haplotypes in African Americans From the women's health initiative. Genet
665 Epidemiol *36*, 107–117. 10.1002/gepi.21603.
666 35. Yu, K., Das, S., LeFaive, J., Kwong, A., Pleiness, J., Forer, L., Schönherr, S.,
667 Fuchsberger, C., Smith, A.V., and Abecasis, G.R. (2022). Meta-imputation: An
668 efficient method to combine genotype data after imputation with multiple reference
669 panels. Am J Hum Genet *109*, 1007–1015. 10.1016/j.ajhg.2022.04.002.
670 36. Li, H. (2011). A statistical framework for SNP calling, mutation discovery,
671 association mapping and population genetical parameter estimation from
672 sequencing data. Bioinformatics *27*, 2987–2993. 10.1093/bioinformatics/btr509.
673 37. Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P. (2007). A new
674 multipoint method for genome-wide association studies by imputation of
675 genotypes. Nat Genet *39*, 906–913. 10.1038/ng2088.
676 38. Sun, Q., Yang, Y., Rosen, J.D., Jiang, M., Chen, J., Liu, W., Wen, J., Raffield,
677 L.M., Pace, R.G., Zhou, Y.-H., et al. (2022). MagicalRsq: Machine-learning-based
678 genotype imputation quality calibration. Am J Hum Genet *109*, 1986–1997.
679 10.1016/j.ajhg.2022.09.009.
680 39. Ferwerda, B., Abdellaoui, A., Nieuwdorp, M., and Zwinderman, K. (2021). A

681 Genetic Map of the Modern Urban Society of Amsterdam. Front Genet *12*.
682 10.3389/fgene.2021.727269.
683 40. Hassan, S., Surakka, I., Taskinen, M.-R., Salomaa, V., Palotie, A., Wessman, M.,
684 Tukiainen, T., Pirinen, M., Palta, P., and Ripatti, S. (2021). High-resolution
685 population-specific recombination rates and their effect on phasing and genotype
686 imputation. Eur J Hum Genet *29*, 615–624. 10.1038/s41431-020-00768-8.
687 41. Sun, Q., Liu, W., Rosen, J.D., Huang, L., Pace, R.G., Dang, H., Gallins, P.J.,
688 Blue, E.E., Ling, H., Corvol, H., et al. (2022). Leveraging TOPMed imputation
689 server and constructing a cohort-specific imputation reference panel to enhance
690 genotype imputation among cystic fibrosis patients. Hum Genet Genomics Adv *3*,
691 100090. 10.1016/j.xhgg.2022.100090.
692
693

694 **Figures**

695



696

697 **Figure 1. Overview of this study.** (A) Imputation using the TOPMed, 1KGP, BBJ1k
698 (1,037 Japanese (JPT) WGS + 1KGP), and JEWEL3k (3,256 JPT WGS + 1KGP)
699 reference panels. The 180k East Asian (EAS) samples in the Biobank Japan (BBJ) first
700 cohort were imputed against the four panels. Samples with whole genome sequencing
701 (WGS_{993}) were used to empirically evaluate the imputation performance, including a
702 count of imputed variants and estimated (Rsq) and true (dosage r^2) imputation
703 accuracy. The deviation between Rsq and dosage r^2 was examined and quantified
704 analytically. Two novel metrics, MAF-adjusted-residual-error (MARE) and β_{imp}, were
705 introduced to indicate the different imputed-dosage distributions underlying the

706 deviation. (B) Simulations to evaluate the relationship between the hidden Markov
707 model (HMM) parameter, Rsq, dosage r^2, the deviation between Rsq and dosage r^2,
708 and imputed dosage. The reference panel and target sample were fixed to the 1KGP
709 and WGS_{993}, and 21 scaled template switching rates (the θ value) were used for
710 running imputations. Evaluations were the same as in (A). (C) Simulations to evaluate
711 the relationship between the reference panel, HMM parameter, imputation metrics, and
712 imputation result. Two scenarios were simulated to represent the imputations of the
713 major and minor ancestry in a multi-ancestry panel. The 3,256 JPT WGS and 1KGP
714 were used to simulate reference panels with different sizes and ancestral diversities.
715 European (EUR) and JPT samples were used as the target. Imputation results were
716 evaluated the same as in (B). In addition, we evaluated how the Minimac3's parameter
717 estimates change with the panel composition.

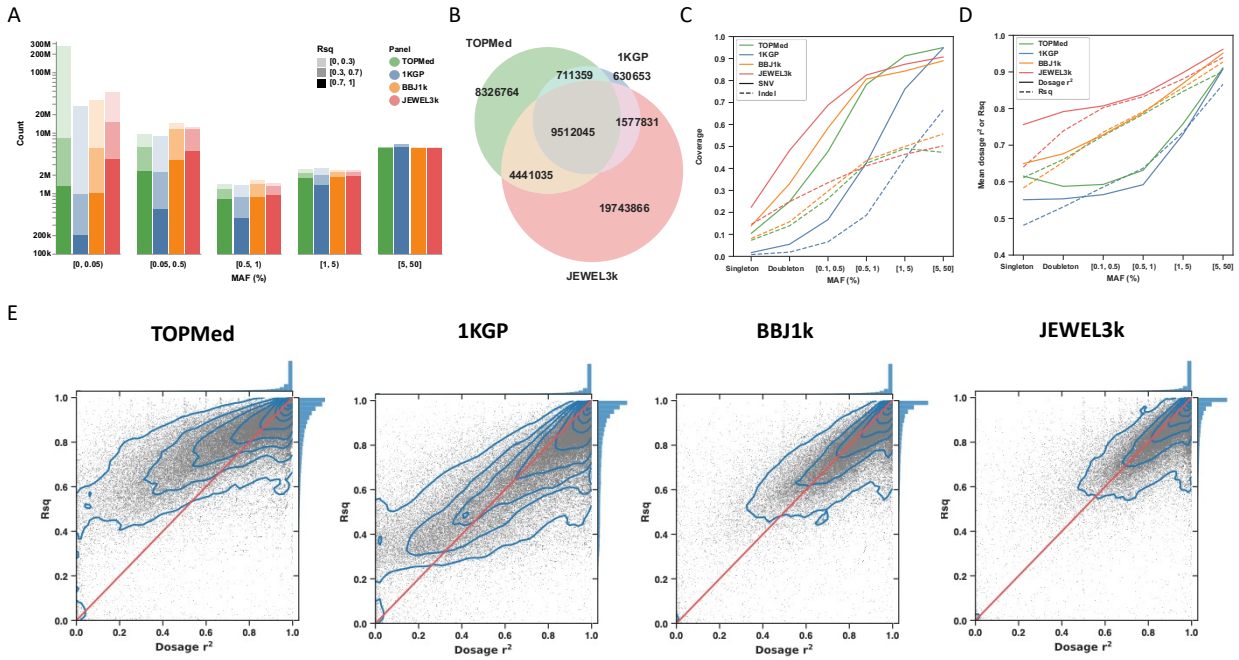
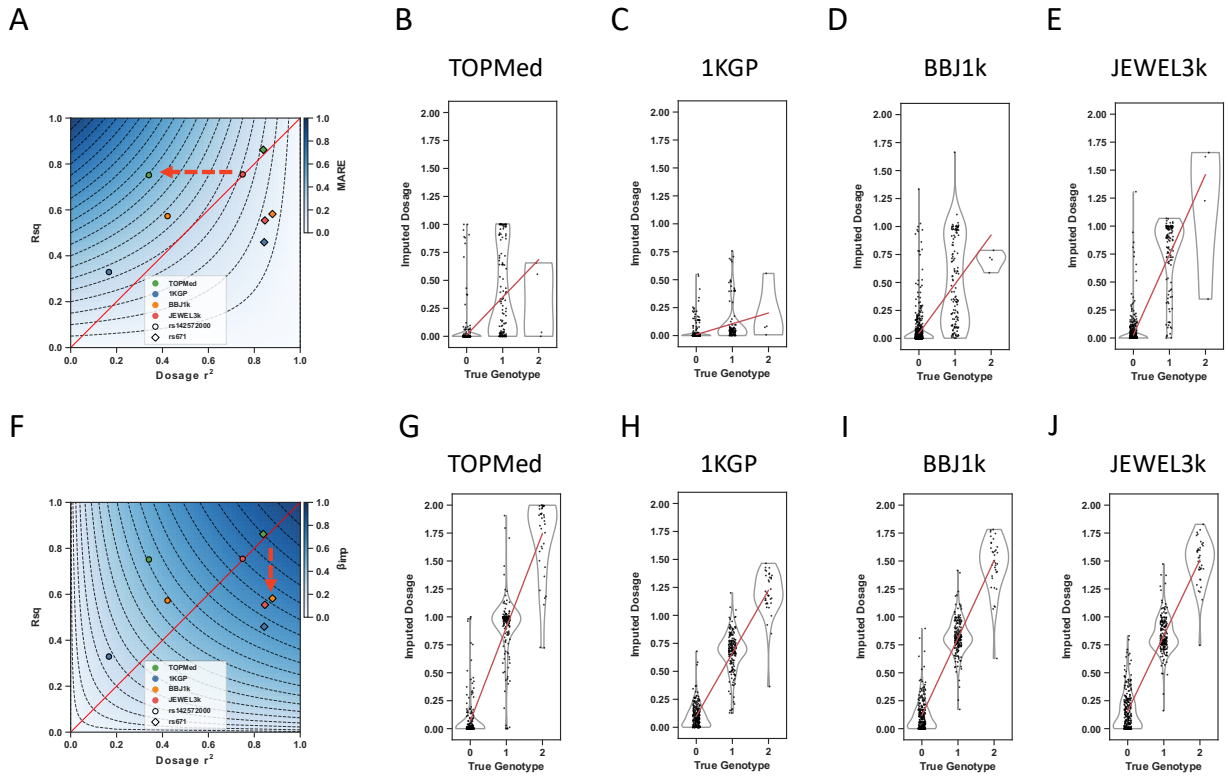


Figure 2. Imputation results using the TOPMed, 1KGP, BBJ1k, and JEWEL3k reference panels. (A) Count of imputed variants using the four panels, stratified by minor allele frequency (MAF) and Rsq in each imputation result. (B) Count of unique and shared variants (with Rsq ≥ 0.3) using the TOPMed, 1KGP, 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.



731

732 **Figure 3. MARE, β_{imp}, Rsq, dosage r^2, and imputed dosage of rs142572000 and**

733 **rs671.** (A) The plot shows Rsq and dosage r^2 of rs142572000 and rs671, and the

734 theoretical value of MARE on the Rsq ~ dosage r^2 plot. (B–E) The plot shows the

735 imputed dosage of rs142572000 in the imputation result using the (B) TOPMed, (C)

736 1KGP, (D) BBJ1k, and (E) JEWEL3k reference panels. (F) The plot shows the

737 theoretical value of β_{imp} on the Rsq ~ dosage r^2 plot. (G–J) The plot shows the imputed

738 dosage of rs671 in the imputation result using the (G) TOPMed, (H) 1KGP, (I) BBJ1k,

739 and (J) JEWEL3k panels. In (A) and (F), the red line shows that Rsq equals dosage r^2,

740 and the color keys show the theoretical MARE or β_{imp} values. The dashed lines

741 represent the contours of MARE or β_{imp}, and adjacent contours have a value difference

742 of 0.05. The red dashed arrow highlights the difference in MARE and β_{imp} (crossing how

743 many contour lines) for variants showing a similar Rsq but different dosage r^2

744 (rs142572000), or similar dosage r^2 but different Rsq (rs671). In (B–E) and (G–J), the
745 strip plot shows the individual and the violin plot shows the distribution. The red line
746 shows the linear regression line between the imputed dosage and true genotype. In (B–
747 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.
748 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
749 0.675.

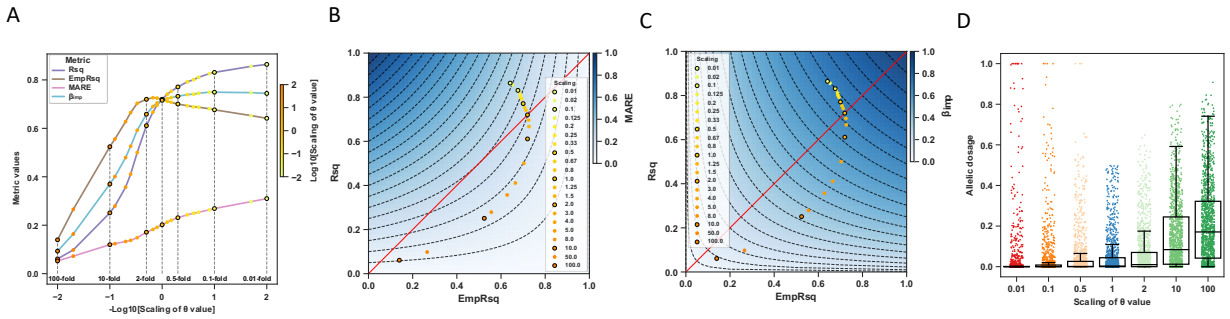


Figure 4. Changes in the metrics value and the imputed allelic dosage with different scalings of the θ value. (A) Rs10410162 is used as an example to visualize the changes in Rsq, EmpRsq, MARE, and β_{imp} (y-axis) with the 21 scalings of the θ value (x-axis). To have a better illustration, the x-axis 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.

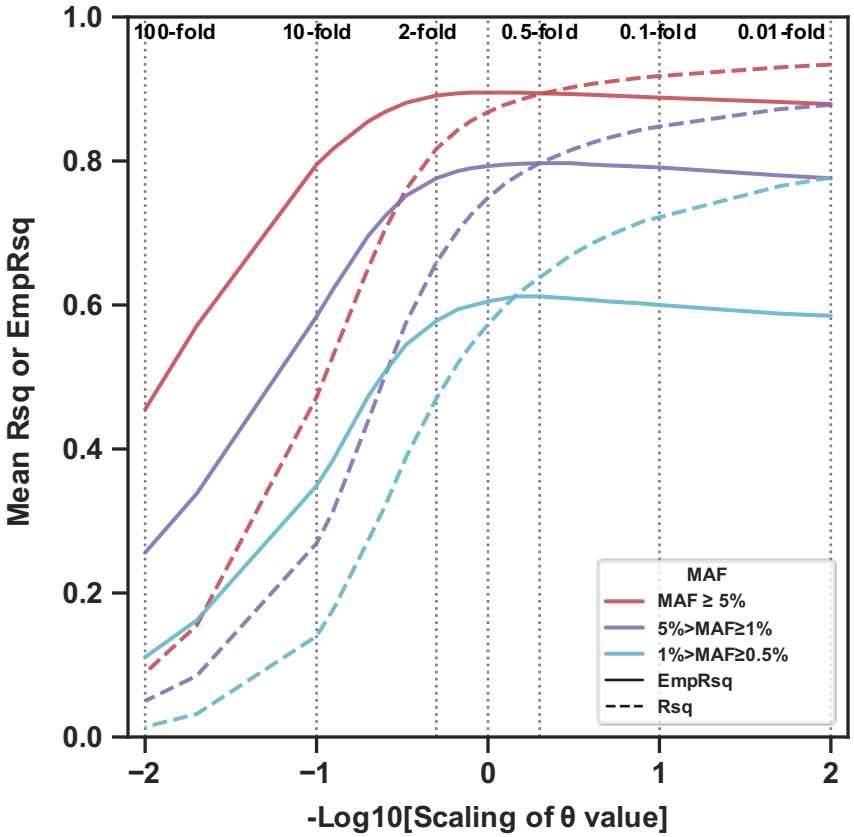


Figure 5. Rsq and EmpRsq using the 1KGP panel with different scalings of the θ value. The plot shows the mean Rsq and EmpRsq of each imputation result, stratified by the MAF of the array. The scalings of 100, 10, 2, 1, 0.5, 0.1, and 0.01-fold are marked with vertical dashed lines and labeled.

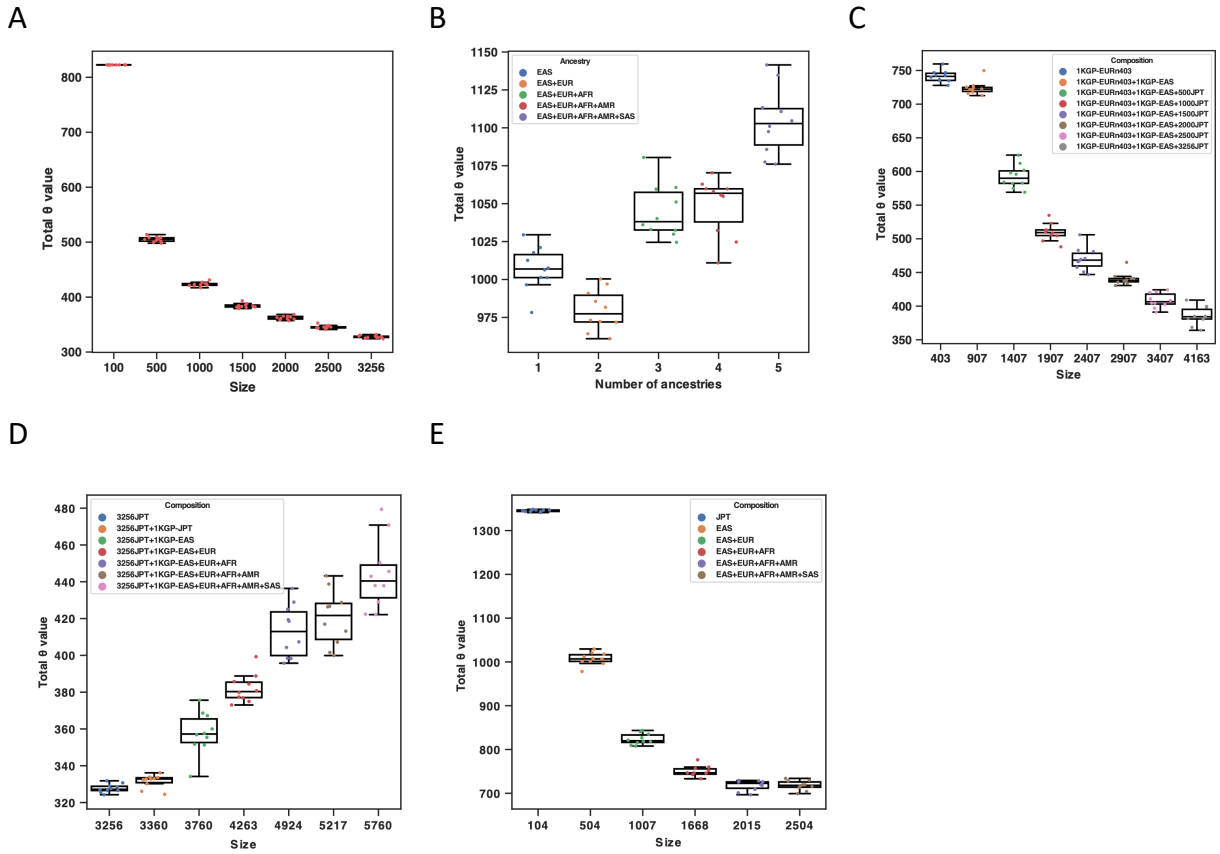


Figure 6. Total θ value of the simulated reference panels. The figure shows the total θ value of simulated reference panels, estimated by Minimac3. (A) JPT-only panels with different sample sizes. (B) Fixed-size panels (504) with 1–5 levels of ancestral diversity, as shown in the legend. (C) Adding large-size (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), EURn403

784 represents the 403 EUR. In (C) and (D), 500–3256JPT represents the number of JPT

785 samples in the panel.

786 **Tables**

787

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

789 **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

790

Table 2. Reference panel description, the total θ value estimated by Minimac3, the mean EmpRsq, and the mean Rsq using the simulated EUR-EAS reference panels 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

EURn403 represents the 403 EUR; 1KGP-EAS represents the 504 EAS; 500–3256JPT represents the number of JPT samples in the reference panel. Minor allele frequency (MAF) is determined by the genotyping array.

797 **Table 3. Reference panel description, the total θ value estimated by Minimac3, the**
798 **mean EmpRsq, and the mean Rsq using the simulated JPT-1KGP reference**
799 **panels with different sizes and ancestral diversity, and WGS_{993} as the target**
800 **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

801 The 3,256 WGS samples of BBJ are included in all panels. 1KGP- followed by the
802 ancestry represents the 1KGP subset. Minor allele frequency (MAF) is determined by
803 the genotyping array.

804 **Table 4. The total θ value estimated by Minimac3 or transformed from the**

805 **HapMap2 genetic map using Minimac4 v1.0.2.**

Reference panel	Parameter source	Background template switching rate	Number of variants with a background template switching rate	Number of variants not with the background template switching rate	Total θ value	Ratio between the transformed and estimated total θ values
1KGP	Genetic map and Minimac4	1.00E-05	493,335	591,200	110.4	0.152
JEWEL3k	Genetic map and Minimac4	1.00E-05	792,501	741,234	112.6	0.267
1KGP	Minimac3	0.00065347	1,084,368	166	728.0	*
JEWEL3k	Minimac3	0.00027504	1,533,730	4	422.4	*

806 Background template switching rate is a minimum probability of switching Minimac sets

807 for all markers. (*) denotes not available.