Comparison of HapMap and 1000 Genomes reference panels in a large-scale genome-wide association study.

Paul S. de Vries[1,2], Maria Sabater-Lleal[3], Daniel I. Chasman[4,5], Stella Trompet[6,7], Tarunveer S. Ahluwalia[8,9], Alexander Teumer[10], Marcus E. Kleber[11], Ming-Huei Chen[12,13], Jie Jin Wang[14], John R. Attia[15,16], Riccardo E. Marioni[17,18,19], Maristella Steri[20], Lu-Chen Weng[21], Rene Pool[22,23], Vera Grossmann[24], Jennifer A. Brody[25], Cristina Venturini[26,27], Toshiko Tanaka[28], Lynda M. Rose[4], Christopher Oldmeadow[15,16], Johanna Mazur[29], Saonli Basu[30], Mattias Frånberg[3,31], Qiong Yang[13,32], Symen Ligthart[1], Jouke J. Hottenga[22], Ann Rumley[33], Antonella Mulas[20], Anton J. M. de Craen[7], Anne Grotevendt[34], Kent D. Taylor[35,36], Graciela E. Delgado[11], Annette Kifley[14], Lorna M. Lopez[17,37,38], Tina L. Berentzen[39], Massimo Mangino[27,40], Stefania Bandinelli[41], Alanna C. Morrison[1], Anders Hamsten[3], Geoffrey Tofler[42], Moniek P. M. de Maat[43], Harmen H. M. Draisma[22,44], Gordon D. Lowe[45], Magdalena Zoledziewska[20], Naveed Sattar[46], Karl J. Lackner[47], Uwe Völker[48], Barbara McKnight[49], Jie Huang[50], Elizabeth G. Holliday[51], Mark A. McEvoy[16], John M. Starr[17,52], Pirro G. Hysi[27], Dena G. Hernandez[53], Weihua Guan[30], Fernando Rivadeneira[1,54], Wendy L. McArdle[55], P. Eline Slagboom[56], Tanja Zeller[57,58], Bruce M. Psaty[59,60], André G. Uitterlinden[1,54], Eco J. C. de Geus[22,23], David J. Stott[61], Harald Binder[62], Albert Hofman[1,63], Oscar H. Franco[1], Jerome I. Rotter[64,65], Luigi Ferrucci[28], Tim D. Spector[27], Ian J. Deary[17,66], Winfried März[11,67, 68], Andreas Greinacher[69], Philipp S. Wild[70,71,72], Francesco Cucca[20], Dorret I. Boomsma[22], Hugh Watkins[73], Weihong Tang[21], Paul M. Ridker[4,5], Jan W. Jukema[6,74,75], Rodney J. Scott[76,77], Paul Mitchell[14], Torben Hansen[78], Christopher J. O'Donnell[13,79], Nicholas L. Smith[60,80, 81], David P. Strachan[82], and Abbas Dehghan[1, 83*]

[1]Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands

[2]University of Texas Health Science Center at Houston, Houston, TX, USA

25 [3]Cardiovascular Medicine Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden

26 [4]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA, USA

27 [5]Harvard Medical School, Boston, MA, USA

28 [6]Department of Cardiology, Leiden University Medical Center, Leiden, the Netherlands

29 [7]Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, the Netherlands

30 [8]Steno Diabetes Center Copenhagen, Gentofte, Denmark

31 [9]Novo Nordisk Foundation Center For Basic Metabolic Research, Section of Metabolic Genetics, Faculty
32 of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

33 [10]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany

34 [11]Vth Department of Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

35 [12]Department of Neurology, Boston University School of Medicine, Boston, MA, USA

36 [13]Framingham Heart Study, Population Sciences Branch, Division of Intramural Research National Heart
37 Lung and Blood Institute, National Institutes of Health, Framingham, MA, USA

38 [14]Centre for Vision Research, Department of Ophthalmology, and Westmead Institute for Medical
39 Research, University of Sydney, Sydney, Australia

40 [15]Public Health Stream, Hunter Medical Research Institute, University of Newcastle, Newcastle, Australia

41 [16]School of Medicine and Public Health, University of Newcastle, Newcastle, Australia

42 [17]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK

43 [18]Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, UK

44 [19]Queensland Brain Institute, University of Queensland, Brisbane, Australia

45 [20]Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Monserrato, Cagliari,
46 Italy

47 [21]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN, USA

48 [22]Department of Biological Psychology, Netherlands Twin Register, VU University, Amsterdam, the
49 Netherlands

50 [23]EMGO+ institute, VU University & VU medical center, Amsterdam, the Netherlands

51 [24]Center for Thrombosis and Hemostasis (CTH), University Medical Center of the Johannes Gutenberg
52 University Mainz, Mainz, Germany

53 [25]Department of Medicine, University of Washington, Seattle WA, USA

54 [26]Division of Infection and Immunology, UCL, London, UK

55 [27]Department of Twin Research and Genetic Epidemiology, Kings College London, London, UK

56 [28]Translational Gerontology Branch, National Institute on Aging, Baltimore, MD, USA

57 [29]Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center of the
58 Johannes Gutenberg University Mainz, Mainz, Germany

59 [30]Division of Biostatistics, University of Minnesota, Minneapolis, MN, USA

60 [31]Department of Numerical Analysis and Computer Science, Stockholm University, Stockholm, Sweden

61 [32]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

62 [33]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK

63 [34]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald,
64 Germany

65 [35]Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research
66 Institute at Harbor/UCLA Medical Center, Torrance, CA, USA

67 [36]Division of Genomic Outcomes, Department of Pediatrics, Harbor-UCLA Medical Center, Torrance,
68 CA, USA

69 [37]Royal College of Surgeons in Ireland, Department of Psychiatry, Education and Research Centre,
70 Beaumont Hospital, Dublin, Ireland

71 [38]University College Dublin, UCD Conway Institute, Centre for Proteome Research, UCD, Belfield,
72 Dublin, Ireland

73 [39]Institute of Preventive Medicine, Bispebjerg and Frederiksberg Hospital, The Capital Region,
74 Copenhagen, Denmark

75 [40]NIHR Biomedical Research Centre at Guy's and St. Thomas' Foundation Trust, London, UK

76 [41]Geriatric Unit, Azienda Sanitaria Firenze (ASF), Florence, Italy

77 [42]Royal North Shore Hospital, Sydney University, Sydney, Australia

78 [43]Department of Hematology, Erasmus MC, Rotterdam, the Netherlands

79 [44]Neuroscience Campus Amsterdam, Amsterdam, the Netherlands

80 [45]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK

81 [46]BHF Glasgow Cardiovascular Research Centre, Faculty of Medicine, Glasgow, UK

82 [47]Institute of Clinical Chemistry and Laboratory Medicine, University Medical Center, Johannes
83 Gutenberg University Mainz, Mainz, Germany

84 [48]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald,
85 Greifswald, Germany

86 [49]Department of Biostatistics, University of Washington, Seattle, WA, USA

87 [50]Department of Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK

88 [51]Public Health Stream, Hunter Medical Research Institute, and School of Medicine and Public Health,
89 University of Newcastle, Newcastle, Australia

90 [52]Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, UK

91 [53]Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA

92 [54]Department of Internal Medicine, Erasmus MC, Rotterdam, the Netherlands

93 [55]School of Social and Community Medicine, University of Bristol, Bristol, UK

94 [56]Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, the Netherlands

95 [57]Department of General and Interventional Cardiology, University Heart Centre, University Medical
96 Center Hamburg-Eppendorf, Hamburg, Germany

97 [58]German Center for Cardiovascular Research (DZHK), Partner Site Hamburg, Lübeck, Kiel, Hamburg,
98 Germany

99 [59]Department of Medicine, Epidemiology, and Health Services, University of Washington, Seattle WA,
100 USA

101 [60]Group Health Research Institute, Group Health Cooperative, Seattle WA, USA

102 [61]Institute of Cardiovascular and Medical Sciences, Faculty of Medicine, University of Glasgow,
103 Glasgow, UK

104 [62]Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Center of the
105 Johannes Gutenberg University Mainz, Mainz, Germany

106 [63]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MS, USA

107 [64]Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Institute for Translational
108 Genomics and Population Sciences, Torrance, CA, USA

109 [65]Division of Genomic Outcomes, Departments of Pediatrics & Medicine, Harbor-UCLA Medical Center,
110 Torrance, CA, USA

111 [66]Department of Psychology, University of Edinburgh, Edinburgh, UK

112 [67]Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim, Germany

113 [68]Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Graz,
114 Austria

115 [69]Institute for Immunology and Transfusion Medicine, University Medicine Greifswald, Greifswald,
116 Germany

117 [70]Preventive Cardiology and Preventive Medicine, Center for Cardiology, University Medical Center of
118 the Johannes Gutenberg-University Mainz, Mainz, Germany

119 [71]Center for Thrombosis and Hemostasis (CTH), University Medical Center of the Johannes Gutenberg-
120 University Mainz, Mainz, Germany

121 [72]German Center for Cardiovascular Research (DZHK), Partner Site RhineMain, Mainz, Germany

[73]Cardiovascular Medicine Dept/Radcliffe Dept of Medicine, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK

[74]Durrer Center for Cardiogenetic Research, Amsterdam, the Netherlands

[75] Interuniversity Cardiology Institute of the Netherlands, Utrecht, the Netherlands

[76]Information based Medicine Program, Hunter Medical Research Institute, University of Newcastle, Newcastle, Australia

[77]School of Biomedical Sciences and Pharmacy, University of Newcastle, Newcastle, Australia

[78]Novo Nordisk Foundation Centre for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

[79]Cardiology Division, Massachusetts General Hospital, Boston, MA, USA

[80]Department of Epidemiology, University of Washington, Seattle WA, USA

[81]Seattle Epidemiologic Research and Information Center, Department of Veteran Affairs Office of Research and Development, Seattle, WA, USA

[82]Population Health Research Institute, St George's, University of London, London, UK

[83]Department of Epidemiology and Biostatistics, Imperial College London, London, UK

*Corresponding author

E-mail: a.dehghan@imperial.ac.uk (AD)

141 # Abstract

142 An increasing number of genome-wide association (GWA) studies are now using the higher resolution
143 1000 Genomes Project reference panel (1000G) for imputation, with the expectation that 1000G
144 imputation will lead to the discovery of additional associated loci when compared to HapMap imputation.
145 In order to assess the improvement of 1000G over HapMap imputation in identifying associated loci, we
146 compared the results of GWA studies of circulating fibrinogen based on the two reference panels. Using
147 both HapMap and 1000G imputation we performed a meta-analysis of 22 studies comprising the same
148 91,953 individuals. We identified six additional signals using 1000G imputation, while 29 loci were
149 associated using both HapMap and 1000G imputation. One locus identified using HapMap imputation
150 was not significant using 1000G imputation. The genome-wide significance threshold of 5×10^{-8} is based
151 on the number of independent statistical tests using HapMap imputation, and 1000G imputation may lead
152 to further independent tests that should be corrected for. When using a stricter Bonferroni correction for
153 the 1000G GWA study (P-value $< 2.5\times10^{-8}$), the number of loci significant only using HapMap
154 imputation increased to 4 while the number of loci significant only using 1000G decreased to 5. In
155 conclusion, 1000G imputation enabled the identification of 20% more loci than HapMap imputation,
156 although the advantage of 1000G imputation became less clear when a stricter Bonferroni correction was
157 used. More generally, our results provide insights that are applicable to the implementation of other dense
158 reference panels that are under development.

159

Introduction

Most genome-wide association (GWA) studies to date have used their genotyped single nucleotide polymorphisms (SNPs) to impute about 2.5 million SNPs detected in the Phase 2 version of the HapMap Project (HapMap) [1-13], including mostly common SNPs with a minor allele frequency (MAF) of over 5%. HapMap imputation enabled the interrogation of most common SNPs possible, even while meta-analyzing studies that used different genotyping arrays with low overlap [1]. However, low-frequency and rare variants are not well covered in the HapMap panel [14]. In addition, genetic variants other than SNPs, such as small insertion/deletions (indels) and large structural variants, are not included in HapMap-based imputed projects, and may be possible sources of missing explained heritability.

In contrast, the more recently released Phase 1 version 3 of the 1000 Genomes Project (1000G) is based on a larger set of individuals [15], and comprises nearly 40 million variants, including 1.4 million indels. 1000G allows the interrogation of most common and low-frequency variants (MAF > 1%), and rare variants (MAF < 1%) that were previously not covered [16]. In general, improving reference panels can lead to the identification of additional significant loci both through the addition of new variants and the improved imputation of known variants. 1000G imputation may thus have several advantages, but given that the denser 1000G imputation comes at the cost of an increased computational and analytical burden, it is important to estimate the observed benefits in practice. Furthermore, such empirical data is needed to make informed decisions in the future on the use of newer reference panels such as UK10K, and the Haplotype Reference Consortium [17, 18]. While several GWA studies using 1000G imputation have been published or are in progress, their sample size differs from the previous GWA studies using HapMap imputation, making comparison difficult. Therefore, with the aim of evaluating the benefits of using 1000G imputation in GWA studies compared to HapMap imputation, we carried out meta-analyses of GWA studies of circulating fibrinogen concentration (a quantitative trait), using both HapMap and 1000G imputed data on the same set of 91,953 individuals.

184

185 ## Results

186 Baseline characteristics of the participants for each of the included studies are shown in S1 Table, and
187 genomic inflation factors are shown separately for the HapMap and 1000G GWA studies in S2 Table.
188 The HapMap GWA study included 2,749,429 SNPs, and the 1000G GWA study included 10,883,314
189 variants. Summary statistics for all variants in the HapMap and 1000G GWA studies are available via the
190 dbGAP CHARGE Summary Results site [19]. Using a genome-wide significance threshold of 5×10^{-8}, a
191 total of 1,210 SNPs across 30 loci were associated with circulating fibrinogen concentration in the
192 HapMap imputed GWA study compared with 4,096 variants across 35 loci in the 1000G imputed GWA
193 study (S1 Fig and S2 Fig). These loci are described in further detail in S3 Table. Of these loci, six were
194 associated only in the 1000G GWA study and one was associated only in the HapMap GWA study, while
195 29 were overlapping (Fig 1A). The HapMap and 1000G lead variants of non-overlapping loci are
196 described in Table 1, and leads variants of overlapping loci are described in Table 2. Among significant
197 loci, the correlation coefficient across cohorts of the beta coefficients, *P*-values, and imputation quality
198 scores of HapMap and 1000G lead variants were 0.925, 0.998, and 0.435 respectively (S3 Fig).

199 **Fig 1. Venn diagram of the number of loci significant using HapMap (left circle) and 1000G (right**
200 **circle) imputation in A) the main analysis, B) the sensitivity analysis applying a significance**
201 **threshold of 2.5×10^{-8} to the 1000G GWA analysis, C) the sensitivity analysis without using genomic**
202 **control corrections, and D) the sensitivity analysis excluding studies that used different imputation**
203 **software, analysis software, or covariates in the HapMap and 1000G GWA analyses.**

204 ### Non-overlapping loci

205 The lead variants for the seven non-overlapping loci always differed between the HapMap and 1000G
206 GWA studies, and all *P*-value differences were greater than one order of magnitude (for example: from
207 5×10^{-8} to 5×10^{-9} or less). Differences between HapMap and 1000G imputation for the seven non-
208 overlapping loci are summarized in Fig 2.

209 **Fig 2. Summary of the differences between HapMap and 1000G imputation for the seven non-**
210 **overlapping loci.**

211

212 Regional plots of the six loci significant only in the 1000G GWA study are shown in Fig 3. For
213 four of these six loci, the correlation r^2 between allelic dosages of the most associated variants imputed
214 using HapMap and 1000G was less than 0.8 (S4 Table). None of the 1000G lead variants among these
215 four loci were included in the HapMap GWA study, and neither were any good proxies (S5 Table).

216 **Fig 3. Regional plots of non-overlapping loci that were more significantly associated with fibrinogen**
217 **in the 1000G GWA study, including variants from both the HapMap (red) and 1000G (green)**
218 **GWA studies.**

219

220 A regional plot of the 6p21.3 locus, which was significant only in the HapMap GWA study, is
221 shown in Fig 4. The most significant *P*-value at the locus was 8.5×10^{-9} in the HapMap GWA study
222 compared to 7.9×10^{-6} in the 1000G GWA study. The correlation r^2 between imputed dosages of the
223 HapMap and 1000G lead variants was low (0.07). The HapMap lead SNP was included in the 1000G
224 GWA study under a different name, rs114339898, but the imputation quality was only sufficient for
225 inclusion in seven of the studies (S5 Table).

226 **Fig 4. Regional plot of 6p21.3, a non-overlapping locus that was more significantly associated with**
227 **fibrinogen in the HapMap GWA study, including variants from both the HapMap (red) and 1000G**
228 **(green) GWA studies.**

229

230 Overlapping loci

231 Regional plots of the 29 overlapping loci are shown in S4 Table. The lead variants of eight of the 29
232 overlapping loci were the same for the HapMap and 1000G GWA studies. *P*-value differences between
233 the HapMap and 1000G GWA studies were often small: they were smaller than or equal to one order of
234 magnitude for 22 loci. *P*-values differed by more than one order of magnitude for seven loci. Five of these

235 loci were more significant in the 1000G GWA study (2q37.3, 4q31.3, 10q21.3, 12q24.12, and 21q22.2),
236 while two of these loci were more significant in the HapMap GWA study (5q31.1 and 8q24.3).

237 Among the five overlapping loci with lower *P*-values in the 1000G GWA study, the correlation r^2
238 between imputed dosages of lead variants from HapMap and 1000G was higher than 0.8 for 4 loci, but
239 was 0.68 for the 12q24.12 locus (S4 Table). There was no good proxy of the 1000G lead variant at the
240 12q24.12 locus included in the HapMap GWA study.

241 The 5q31.1 and 8q24.3 loci had lower *P*-values in the HapMap GWA study. The correlation r^2
242 between imputed dosages from HapMap and 1000G was almost perfect for 5q31.1, but was 0.75 for
243 8q24.3. The HapMap lead variant of the 8q24.3 locus was also included in the 1000G GWA study. These
244 differences between HapMap and 1000G imputation for the 29 overlapping loci are summarized in Fig 5.

245 **Fig 5. Summary of the differences between HapMap and 1000G imputation for the 29 overlapping**
246 **loci.**

247

248 Sensitivity analyses

249 Because more independent variants are included in the 1000G GWA study [20, 21] , using the
250 conventional genome-wide significance threshold of 5×10^{-8} may result in an increased type I error rate.
251 When we used a more stringent genome-wide significance threshold of $2.5\times10^{-8}$for the 1000G GWA
252 study as suggested by Huang et al. [20], there were 4 loci significant only in the HapMap GWA study, 5
253 loci significant only in the 1000G GWA study, and 26 overlapping loci (Fig 1B). Three loci that were
254 significant using both HapMap and 1000G imputation thus became non-significant when the stricter
255 significance threshold was applied to the 1000G results.

256 Genomic inflation factors to correct for genomic control were calculated separately for the
257 HapMap and 1000G analyses of each study. Thus, differences in the genomic inflation factors could
258 explain some of the differences between the HapMap and 1000G results. When we repeated the HapMap

259 and 1000G GWA study without applying genomic control corrections, 2 loci were associated only with
260 circulating fibrinogen concentration in the HapMap GWA study, 6 were only associated in the 1000G
261 GWA study, and 30 were associated in both GWA studies (Fig 1C and S6 Table).

262 For practical reasons, not all of the studies used the same imputation software, analysis software,
263 or covariates for the HapMap and 1000G analyses. Specifically, fewer studies used principal components
264 in the HapMap GWA study. When we restricted the analysis to those studies that used the same
265 imputation software, analysis software, and covariates in the HapMap and 1000G GWA studies (S7 Table
266 and S8 Table), 3 loci were associated only in the 1000G GWA study, and 6 were associated in both the
267 HapMap and the 1000G GWA studies (Fig 1D and S9 Table). No loci were associated only in the
268 HapMap GWA study.

269

270 ## Discussion

271 In our fibrinogen GWA study of 91,953 individuals, using 1000G instead of HapMap imputation led to
272 the identification of six additional fibrinogen loci, suggesting an improvement in the detection of
273 associated signals. Nevertheless, there was also one locus that was only identified when using HapMap
274 imputation, and the advantage of 1000G imputation was attenuated when using a more stringent
275 Bonferroni correction for the 1000G GWA study. The inclusion of indels in the 1000G GWA study did
276 not lead to the identification of any new loci. Only one locus in our 1000G GWA study was led by an
277 indel, and it was in strong linkage disequilibrium with a SNP present in HapMap.

278 While this is the first study of the impact of HapMap and 1000G imputation on genome-wide
279 associations using exactly the same individuals in a large-scale consortium setting, four previous studies
280 have addressed this question on a smaller scale. In the Wellcome Trust Case Control Consortium,
281 consisting of 2000 for seven diseases (bipolar disorder, coronary artery disease, Crohn's disease,
282 hypertension, rheumatoid arthritis, type 1 and 2 diabetes) and 3000 shared controls, Huang et al. re-

283 analyzed GWA studies of these seven diseases with 1000G imputation, and found two novel loci: one for
284 type 1 diabetes and one for type 2 diabetes [20]. A more conservative genome-wide significance threshold
285 of 2.5×10^{-8} was used in the 1000G GWA studies, while the MAF inclusion threshold was the same at 1%.
286 The second study was a 1000G imputed GWA study of around 2000 cases of venous thrombosis and
287 2400 controls [22]. Using a conservative P-value threshold of 7.4×10^{-9}, but no MAF threshold, Germain
288 et al. identified an uncommon variant at a novel locus that was not identified in the HapMap GWA study
289 [22]. Third, the National Cancer Institute Breast and Prostate Cancer Cohort Consortium found no new
290 loci by applying 1000G imputation to their existing dataset of 2800 cases and 4500 controls [23, 24]. The
291 conventional genome-wide significance threshold of 5×10^{-8} was used, but no MAF threshold was used.
292 Fourthly, Wood et al. compared HapMap and 1000G imputation for a total of 93 quantitative traits in
293 1210 individuals from the InCHIANTI study [25]. Using a significance threshold of 5×10^{-8} for both the
294 HapMap and 1000G GWA studies, they found 20 overlapping associations, 13 associations that were
295 only significant using 1000G imputation, and one association that was only significant using HapMap
296 imputation. For the association significant only in HapMap, the *P*-value difference between HapMap and
297 1000G lead variants was less than one order of magnitude. When the authors lowered their significance
298 threshold to 5×10^{-11} to reflect the number of tests being done in analyzing multiple traits, 9 associations
299 remained significant based on HapMap imputation and 11 associations remained significant based on
300 1000G imputation.

301 All four of these comparison studies used an earlier 1000 genomes reference panel. The present
302 study adds to the literature as it is based on the widely implemented Phase 1 Version 3 of 1000G.
303 Crucially, the large sample size allowed us to examine differences at many non-overlapping and
304 overlapping loci, and improved the generalizability of our results, as ongoing GWA studies are often
305 conducted in large consortia.

306 Two further studies with different approaches also provide insights. First, Springelkamp et al.
307 found a novel locus using 1000G imputation even though the sample size was smaller than the previous

308 HapMap GWA study [26, 27]. The same genome-wide significance (5×10^{-8}) and MAF (1%) thresholds
309 were used. The lowest *P*-value at the locus was 1.9×10^{-8}. Because different individuals were included in
310 these GWA studies, the difference between HapMap and 1000G may partially be explained by sampling
311 variability. Second, Shin et al. identified 299 SNP-metabolite associations based on HapMap imputation,
312 and reexamined the associated loci using 1000G imputation in the same individuals [28]. They found that
313 HapMap and 1000G imputation yielded similar *P*-values and variance explained for all but one loci. For
314 that locus, the 1000G imputation based association was considerably stronger: the explained variance
315 increased from 10% to 16%, and the *P*-value decreased from 8.8×10^{-113} to 7.7×10^{-244}. Although Shin et al.
316 did not compare loci identified using HapMap and 1000G, their results do support our finding that large
317 differences in association strengths are possible, albeit not at every locus. All these studies, along with the
318 current study, suggest that additional signals not previously identified in HapMap GWA studies can be
319 found using the 1000G GWA study, with the same sample size.

320 In the current study we demonstrate that, although 1000G imputation was overall more effective
321 at identifying associated loci, HapMap imputation may outperform 1000G imputation for specific loci.
322 The 6p21.3 locus, corresponding to the major histocompatibility complex (MHC), was significant in the
323 HapMap GWA study but not in the 1000G GWA study. The MHC locus is highly polymorphic and hosts
324 many repetitive sequences, rendering it difficult to genotype and sequence [29-31]. The HapMap
325 reference panel was based largely on the genotyping of variants that were known at that time, whereas the
326 1000G reference panel is based entirely on low-coverage sequencing. This may explain the rather large
327 discrepancy between HapMap and 1000G at this locus.

328 Differences in associations when GWA studies are based on different participants can be
329 explained by sampling variability, even with the same sample size. Hence, by using exactly the same
330 participants in the HapMap and 1000G comparisons in the present project, we rule out both statistical
331 power and sampling variability as possible explanations for differences between the HapMap and 1000G
332 GWA studies. Several real differences between the HapMap and 1000G reference panels may underlie

333 the net benefit of 1000G imputation. The HapMap reference panel was largely based on genotypes of
334 known variants, whereas the 1000G reference panel was primarily based on low-pass whole genome
335 sequencing, enhancing the inclusion of novel variants. Additionally, most studies used only a small
336 number of European-ancestry participants for HapMap imputation, whereas they used a larger number of
337 participants of all available ancestries for 1000G imputation, introducing further haplotypes into the
338 imputation process.

339 Nevertheless, some analytical differences between the HapMap and 1000G analyses were not
340 controlled for in our main analysis and therefore remain as potential alternative explanations. First,
341 genomic control corrections were applied to the results of each of the studies before meta-analysis,
342 separately for the HapMap and 1000G GWA studies. As a result, for any given study, there could be
343 differences between the correction applied to the HapMap GWA analysis and to the 1000G GWA
344 analysis. As these differences do not appear to differ systematically between the HapMap and 1000G
345 GWA analyses in our study, the genomic control corrections are unlikely to explain our results. The
346 results from our sensitivity analysis were concordant with this interpretation: when no genomic control
347 corrections were applied there were 6 loci only significant in the 1000G GWA study compared to 2 loci
348 only significant in the HapMap GWA study.

349 The second difference between the HapMap and 1000G GWA studies that may explain our
350 findings is that in the 1000G GWA study more studies were adjusted for ancestry-informative principal
351 components. This difference reflects common practice, as population stratification is suspected to have a
352 stronger influence on variants with lower MAF, and 1000G includes more of these [32]. However, the
353 adjustments are applied to variants across the spectrum of minor allele frequencies, which may have
354 influenced our results.

355 Thirdly, some studies used different software for HapMap and 1000G imputation (S1 Table). The
356 imputation quality metrics used by IMPUTE and MACH differ, and this has traditionally been dealt with

by applying different imputation quality thresholds: > 0.3 for MACH and > 0.4 for IMPUTE [5, 33]. In studies that used different imputation software for the HapMap and 1000G GWA studies, the filtering of variants can therefore differ. There may, additionally, be real differences in imputation quality. Finally, some studies used different analysis software (S3 Table). When we restricted our analysis to only those studies that used the same covariates, analysis software, and imputation software for the HapMap and 1000G GWA studies, 3 loci were only significant in the 1000G GWA study, while all loci significant in the HapMap GWA study were also significant in the 1000G GWA study. This suggests that differences in imputation software, analysis software, and covariates do not fully explain the observed difference between the HapMap and 1000G GWA studies, and that there are real differences resulting from choice of reference panel.

1000G GWA studies include more independent statistical tests than HapMap GWA studies [20, 21]. Thus, while a *P*-value threshold of 5×10^{-8}, correcting for 1 million independent tests, maintains the type I error rate at 5% for HapMap GWA studies, this may not be the case for 1000G GWA studies. Using 1000G pilot data, Huang et al. estimated that 2 million independent tests were being done, and thus suggested a *P*-value threshold of 2.5×10^{-8} [20]. In our study we used a *P*-value threshold of 5×10^{-8} for both the HapMap and 1000G GWA studies, in accordance with the majority of published 1000G GWA studies [26, 34-37]. When we used the threshold of 2.5×10^{-8} in the 1000G imputed GWA study, the difference between the HapMap and 1000G GWA studies became smaller. Thus, while we expect applying 1000G imputation may lead to novel findings using the conventional genome-wide significance threshold, this expectation may not be met when using stricter, and perhaps more appropriate thresholds. In other words, using the traditional significance threshold for 1000G may increase the type 1 error rate, which may account for some additional significant loci detected in 1000G GWA studies.

In this study we only examined variants with a MAF of greater than 1%. This restriction was common practice for HapMap GWA studies, but given the improved coverage of rare variants in 1000G, this may not remain the case for 1000G GWA studies. Different MAF thresholds have been used in

382 published 1000G GWA studies, although many have used 1% [20, 22, 23, 26, 27, 34-40]. Therefore, an
383 advantage of 1000G not illustrated by this study may be the identification of rare variants, at new loci or
384 as secondary signals at known loci. The advantage of 1000G imputation will then in part depend on the
385 importance and impact of rare variants in the trait being studied, as well as the distribution of these
386 variants. Rare and uncommon variants are often clustered in genes with previously associated common
387 variants, limiting the new biology revealed through their identification [41, 42]. This appears to be the
388 case for fibrinogen concentration as well [43, 44].

389 In conclusion, we show that the reference panel used in GWA studies can have an impact on the
390 identification of common variants, although our results do not support the expectation that 1000G
391 imputation always outperforms HapMap imputation, as we found one locus that appeared to be better
392 covered in HapMap. This suggests that GWA studies will continue to be more successful as newer
393 reference panels such as the Haplotype Reference Consortium are adopted. Nevertheless, our results also
394 suggest that the benefits of 1000G are considerably reduced when the additional independent tests
395 introduced by 1000G imputation are corrected for. Given that the bulk of the new information provided
396 by 1000G imputation relates to low-frequency variants, we expect the penalty increased multiple testing
397 burden to become less relevant in future studies as the power to examine these low-frequency variants
398 increases with larger sample sizes and enhanced imputation quality. Imputation using the Haplotype
399 Reference Consortium reference panel improves the imputation quality of low-frequency variants when
400 compared to 1000G, and future reference panels based on the wealth of whole-genome sequencing data
401 currently being generates by efforts such as TOPMed are likely to continue this trend [45].

402

403 # Methods

404 **Population**

405 The sample for both the HapMap and 1000G GWA studies consists of 22 studies including the same
406 91,953 European-ancestry participants. The sample is largely a subset of the sample used in our previous
407 work, and when possible the same analyses were used in this project [44, 46]. However, to ensure that
408 only the same individuals were used, one or both of the analyses was rerun using only overlapping
409 individuals when necessary. All studies were approved by appropriate research ethics committees and all
410 respondents signed informed consent prior to participation. The ARIC study was approved by the
411 University of Mississippi Medical Center IRB, Wake Forest University Health Sciences IRB, University
412 of Minnesota IRB, and John Hopkins University IRB. The B58C study was approved by the South East
413 England Multi-Centre Research Ethics Committee and the London & South East Committee of the
414 National Research Ethics Service. The BMES was approved by the University of Sydney and the Western
415 Sydney Area Health Service Human Research Ethics Committees. The CHS was approved by the Wake
416 Forest University Health Sciences IRB, University of California, Davis IRB, John Hopkins University
417 IRB, and University of Pittsburgh IRB, and University of Washington IRB. The FHS was approved by
418 the Bostin University IRB. The GHS was approved by the Ethics Committee of the Landesärztekammer
419 Rheinland-Pfalz (State Chamber of Physicians of Rhineland-Palatinate, Germany). The GOYA-Male
420 study was approved by the regional scientific ethics committee of Copenhagen, Denmark, and the Danish
421 data protection board. The HCS was approved by the University of Newcastle and Hunter New
422 England Human Research Ethics Committee. The InCHIANTI study was approved by the Italian National
423 Institute of Research and Care of Aging Institutional Review and Medstar Research Institute (Baltimore,
424 MD). The LBC1921 study was approved by the Lothian Research Ethics Committee and the Scotland A
425 Research Ethics Committee. The LBC1936 study was approved by the Multi-Centre Research Ethics
426 Committee for Scotland and the Lothian Research Ethics Committee and the Scotland A Research Ethics
427 Committee. The LURIC study was approved by the Ethics Committee at the Ärztekammer Rheinland-
428 Pfalz. The NTR study was approved by the Medical Ethical Committee of the VU University Medical
429 Center Amsterdam, and the Central Committee on Research Involving Human Subjects of the VU
430 University Medical Center Amsterdam. The PROCARDIS study was approved by the Ethics Committee

431 of the Karolinska Institutet. The PROSPER-PHASE study was approved by the Greater Glasgow
432 Community/Primary Care Local Research Ethics Committee, Dumfries and Galloway Health Board Local
433 Research Ethics Committee, Argyll and Clyde Health Board Local Research Ethics Committee,
434 Lanarkshire Research Ethics Committee, Research Ethics Committee of the Cork Teaching Hospitals, and
435 the Medical Ethical Committee of the Leiden University Medical Center. The RS was approved by the
436 Medical Ethics Committee of the Erasmus MC and the Dutch Ministry of Health, Welfare and Sport. The
437 SardiNIA study was approved by the Ethics Committee at Azienda Sanitaria Locale (ASL) n°1 of Sassari,
438 Sardinia, Italy. The SHIP was approved by the Medical Ethics Committee of the University of
439 Greifswald. The TwinsUK study was approved by the NRES Committee London-Westminster (formerly
440 St Thomas' Ethics Committee). The WGHS was approved by Brigham and Women's Hostpital IRB.

441

442 **Genotyping and Imputation**

443 Genotyping and pre-imputation quality control methods for each study are shown in S7 Table. Studies
444 imputed dosages of genetic variants using reference panels from the 1000 genomes project with MACH
445 [47, 48] or IMPUTE [49]. Studies imputed variant dosages using Phase 2 reference panels from the
446 HapMap project with MACH [47, 48], IMPUTE [49], or BIMBAM [50]. We excluded variants with
447 MACH imputation quality < 0.3, IMPUTE/BIMBAM imputation quality < 0.4, or MAF < 0.01 from each
448 study.

449

450 **Fibrinogen measurement**

451 Fibrinogen concentration was measured in citrated or EDTA plasma samples using a variety of methods
452 including the Clauss method, immunonephelometric methods, immunoturbidimetric methods, and other
453 functional methods. Fibrinogen concentration was measured in g/L and natural log transformed. Details
454 about the fibrinogen measurement are shown in S10 Table.

455

456 **Genome-wide association analysis**

457 All analyses were adjusted for age and sex, and study specific covariates such as center or case/control
458 status. In family studies, linear mixed models were used to account for family structure. Some studies
459 adjusted the analysis for principle components to account for population structure and cryptic relatedness.
460 Some studies used a different number of principle components in the HapMap and 1000G analyses. The
461 adjustments and analysis software used by each study are shown in S8 Table. We applied a genomic
462 control correction to the results of each of the studies before meta-analysis to remove any remaining
463 genomic inflation. The genomic inflation factor used in this correction was calculated separately in the
464 HapMap and 1000G analyses for each study. We meta-analyzed the results using an inverse-variance
465 model with fixed effects implemented in METAL [51]. Loci were defined as the 500 Kb area on either
466 side of lead variants (the variant with the smallest *P*-value). Build 36 positions of HapMap SNPs were
467 converted to build 37 using the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgLiftOver).
468 Variants were annotated to genes using ANNOVAR version 2013Mar07. At the meta-analysis level, the
469 imputation quality of each variant was defined as the sample-size weighted mean imputation quality
470 across the studies, not including studies where the variant was filtered out.

471

472 **Comparison of HapMap and 1000G**

473 When a locus was significant in both the HapMap and 1000G GWA studies we defined it as an
474 overlapping locus. When a locus was significant in only one of the two analyses we defined it as a non-
475 overlapping locus. To compare the strength of association in the HapMap and 1000G GWA studies, we
476 identified loci with *P*-value differences of 1 order of magnitude or greater (for example: from 5×10^{-8}
477 compared to 5×10^{-9} or less).

478 For each significant locus we used two approaches to assess the relationship between lead
479 variants from HapMap and 1000G. First, we determined whether or not the more significant of the two
480 lead variants or a good proxy (linkage disequilibrium $r^2 > 0.8$) was included in the analysis of the other

reference panel. If so, we examined its association in the other reference panel. Thus, if a locus was more significant in the 1000G GWA study, we checked whether the 1000G lead variant or a proxy was included in the HapMap GWA study. Second, we examined the correlation R^2 between HapMap and 1000G lead variants in the form of imputed genotype dosages. This was performed for 5966 individuals from the Rotterdam Study (see study description in S1 Text) [52].

Sensitivity analysis

First, we compared the results of the HapMap and 1000G GWA studies when applying a stricter Bonferroni-corrected *P*-value threshold of 2.5×10^{-8} to the 1000G GWA study. This threshold was suggested by Huang et al. to keep the type 1 error rate at 5% when using 1000G data [20]. Second, we repeated the analysis without using genomic control corrections. Third, we repeated the analysis in 34,098 participants using only the 10 studies that used the same imputation and analysis software as well as the same covariates for the HapMap and 1000G GWA studies.

Acknowledgements

The authors acknowledge the essential role of the Cohorts for Heart and Aging Research in Genome Epidemiology (CHARGE) Consortium in development and support of this manuscript. The authors thank the staff and participants of the ARIC study for their important contributions. We would like to thank the University of Minnesota Supercomputing Institute for use of the calhoun supercomputers. A full list of principal CHS investigators and institutions can be found at CHS-NHLBI.org. The analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. The authors would like to thank the men and women participating in the HCS as well as The University of Newcastle, Vincent Fairfax Family Foundation and The Hunter Medical Research Institute. We thank the LBC1936 and LBC1921 participants and research team members. We thank the nurses and staff at the Wellcome Trust Clinical

506 Research Facility, where subjects were tested and the genotyping was performed. We thank the LURIC
507 study team who were either temporarily or permanently involved in patient recruitment as well as sample
508 and data handling, in addition to the laboratory staff at the Ludwigshafen General Hospital and the
509 Universities of Freiburg and Ulm, Germany. This work was performed as part of an ongoing
510 collaboration of the PROSPER study group in the universities of Leiden, Glasgow and Cork. The authors
511 are grateful to the study participants, the staff from the Rotterdam Study and the participating general
512 practitioners and pharmacists. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera,
513 Marjolein Peters and Carolina Medina-Gomez for their help in creating the GWAS database, and Karol
514 Estrada and Carolina Medina-Gomez for the creation and analysis of imputed data. We thank the many
515 individuals who generously participated in this study, the Mayors and citizens of the Sardinian towns
516 involved, the head of the Public Health Unit ASL4, and the province of Ogliastra for their volunteerism
517 and cooperation. In addition, we are grateful to the Mayor and the administration in Lanusei for providing
518 and furnishing the clinic site. We are grateful to the physicians Angelo Scuteri, Marco Orrù, Maria Grazia
519 Pilia, Liana Ferreli, Francesco Loi, nurses Paola Loi, Monica Lai and Anna Cau who carried out
520 participant physical exams; the recruitment personnel Susanna Murino; Mariano Dei, Sandra Lai, Andrea
521 Maschio, Fabio Busonero for genotyping; Maria Grazia Piras and Monia Lobina for fibrinogen
522 phenotyping.

523 Steno Diabetes Center and Synlab Holding Deutschland GmbH provided support in the form of salaries
524 for authors T.S.A. and W.M. respectively, but did not have any additional role in the study design, data
525 collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these
526 authors are articulated in the 'author contributions' section. Infrastructure for the CHARGE Consortium is
527 supported in part by the National Heart, Lung, and Blood Institute grant R01HL105756. **ARIC** is carried
528 out as a collaborative study supported by National Heart, Lung, and Blood Institute (NHLBI) contracts
529 HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C,
530 HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C,

531 R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract
532 U01HG004402; and National Institutes of Health contract HHSN268200625226C. Infrastructure was
533 partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and
534 NIH Roadmap for Medical Research. LITE is supported by HL0597367 from the NHLBI. **B58C**
535 acknowledges use of phenotype and genotype data from the British 1958 Birth Cohort DNA collection,
536 funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02.
537 Genotyping for the B58C-WTCCC subset was funded by the Wellcome Trust grant 076113/B/04/Z. The
538 B58C-T1DGC genotyping utilized resources provided by the Type 1 Diabetes Genetics Consortium, a
539 collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney
540 Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human
541 Genome Research Institute (NHGRI), National Institute of Child Health and Human Development
542 (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supported by U01
543 DK062418. B58C-T1DGC GWAS data were deposited by the Diabetes and Inflammation Laboratory,
544 Cambridge Institute for Medical Research (CIMR), University of Cambridge, which is funded by Juvenile
545 Diabetes Research Foundation International, the Wellcome Trust and the National Institute for Health
546 Research Cambridge Biomedical Research Centre; the CIMR is in receipt of a Wellcome Trust Strategic
547 Award (079895). The B58C-GABRIEL genotyping was supported by a contract from the European
548 Commission Framework Programme 6 (018996) and grants from the French Ministry of Research. The
549 **BMES** has been supported by the Australian National Health & Medical Research Council, Canberra
550 Australia (Grant Numbers 974159, 211069, 457349, 512423, 475604, 529912, and the funding for Centre
551 for Clinical Research Excellence in Translational Clinical Research in Eye Diseases, CCRE in TCR-Eye,
552 grant ID 529923); In addition, funding by the Wellcome Trust, UK (to A Viswanathan, P McGuffin, P
553 Mitchell, F Topouzis, P Foster) has supported the genotyping costs of the entire BMES population. This
554 **CHS** research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C,
555 N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083,
556 N01HC85086; and NHLBI grants U01HL080295, R01HL087652, R01HL105756, R01HL103612, and

557 R01HL120393 with additional contribution from the National Institute of Neurological Disorders and
558 Stroke (NINDS). Additional support was provided through R01AG023629 from the National Institute on
559 Aging (NIA). The provision of genotyping data was supported in part by the National Center for
560 Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and
561 Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern
562 California Diabetes Endocrinology Research Center. The **FHS** was partially supported by the National
563 Heart, Lung, and Blood Institute's (NHLBI's) Framingham Heart Study (Contract No. N01-HC-25195)
564 and its contract with Affymetrix, Inc. for genotyping services (Contract No. N02-HL-6-4278). A portion
565 of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson
566 Evans Endowment of the Department of Medicine at Boston University School of Medicine and Boston
567 Medical Center. Partial investigator support was provided by the National Institute of Diabetes and
568 Digestive and Kidney Diseases K24 DK080140 (JB Meigs), the National Institute on Aging and National
569 Institute for Neurological Disorders and Stroke R01 AG033193, NS017950 (S Seshadri). **The GOYA**
570 **Male** study was conducted as part of the activities of the Gene-diet Interactions in Obesity project
571 (GENDINOB, www.gendinob.dk) and the MRC centre for Causal Analyses in Translational
572 Epidemiology (MRC CAiTE). We thank the staff of the Copenhagen City Heart Study for their skillful
573 examination of the study subjects in collection of baseline and follow-up data. Tarunveer Singh
574 Ahluwalia received his Postdoctoral Research funding from GENDINOB project and acknowledges the
575 same. The **Gutenberg Health Study** is funded through the government of Rhineland-Palatinate
576 („Stiftung Rheinland-Pfalz für Innovation", contract AZ 961-386261/733), the research programs
577 "Wissen schafft Zukunft" and "Center for Translational Vascular Biology (CTVB)" of the Johannes
578 Gutenberg-University of Mainz, and its contract with Boehringer Ingelheim and PHILIPS Medical
579 Systems, including an unrestricted grant for the Gutenberg Health Study. VG, PSW are funded by the
580 Federal Ministry of Education and Research (BMBF 01EO1003). **The InCHIANTI study** baseline
581 (1998-2000) was supported as a "targeted project" (ICS110.1/RF97.71) by the Italian Ministry of Health
582 and in part by the U.S. National Institute on Aging (Contracts: 263 MD 9164 and 263 MD 821336); This

583 research was supported in part by the Intramural Research Program of the NIH, National Institute on
584 Aging. The whole genome association study in **LBC1936 and LBC1921** was funded by the
585 Biotechnology and Biological Sciences Research Council (BBSRC; Ref. BB/F019394/1). The LBC1936
586 research was supported by Age UK. The LBC1921 data collection was funded by the BBSRC. The work
587 was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive
588 Epidemiology (MR/K026992/1), part of the cross council Lifelong Health and Wellbeing Initiative.
589 Funding from the BBSRC, and MRC is gratefully acknowledged. **LURIC** has received funding from the
590 6th Framework Program (integrated project Bloodomics, grant LSHM-CT-2004-503485) and from the
591 7th Framework Program (Atheroremo, grant agreement number 201668 and RiskyCAD, grant agreement
592 number 305739) of the European Union as well as from the INTERREG IV Oberrhein Program (Project
593 A28, Genetic mechanisms of cardiovascular diseases) with support from the European Regional
594 Development Fund (ERDF) and the Wissenschaftsoffensive TMO. **NTR:** Funding was obtained from the
595 Netherlands Organization for Scientific Research (NWO) and MagW/ZonMW grants 904-61-090, 985-
596 10-002, 904-61-193,480-04-004, 400-05-717, Addiction-31160008, Middelgroot-911-09-032,
597 Spinozapremie 56-464-14192, Center for Medical Systems Biology (CSMB, NWO Genomics),
598 NBIC/BioAssist/RK(2008.024), Biobanking and Biomolecular Resources Research Infrastructure
599 (BBMRI –NL, 184.021.007). VU University's Institute for Health and Care Research (EMGO+)
600 and Neuroscience Campus Amsterdam (NCA); the European Science Foundation (ESF, EU/QLRT-
601 2001-01254), the European Community's Seventh Framework Program (FP7/2007-2013), ENGAGE
602 (HEALTH-F4-2007-201413); the European Science Council (ERC Advanced, 230374), Rutgers
603 University Cell and DNA Repository (NIMH U24 MH068457-06), the Avera Institute, Sioux Falls,
604 South Dakota (USA) and the National Institutes of Health (NIH, R01D0042157-01A, MH081802, Grand
605 Opportunity grants 1RC2 MH089951). Part of the genotyping and analyses were funded by the Genetic
606 Association Information Network (GAIN) of the Foundation for the National Institutes of Health.
607 Computing was supported by BiG Grid, the Dutch e-Science Grid, which is financially supported by
608 NWO. **PROCARDIS** was supported by the European Community Sixth Framework Program (LSHM-

609 CT- 2007-037273), AstraZeneca, the British Heart Foundation, the Wellcome Trust (Contract No.
610 075491/Z/04), the Swedish Research Council, the Knut and Alice Wallenberg Foundation, the Swedish
611 Heart-Lung Foundation, the Torsten and Ragnar Söderberg Foundation, the Strategic Cardiovascular and
612 Diabetes Programs of Karolinska Institutet and Stockholm County Council, the Foundation for Strategic
613 Research and the Stockholm County Council. Jemma C Hopewell and Robert Clarke acknowledge
614 support from the BHF Centre of Research Excellence, Oxford. M.Sabater-Lleal is supported by the
615 Swedish Heart-Lung Foundation (20130399), and acknowledges funding from Åke Wiberg and Tore
616 Nilssons foundations. B.Sennblad acknowledges funding from the Magnus Bergvall Foundation and the
617 Foundation for Old Servants. **PROSPER** received funding from the European Union's Seventh
618 Framework Programme (FP7/2007-2013) under grant agreement n° HEALTH-F2-2009-223004. For a
619 part of the genotyping we received funding from the Netherlands Consortium of Healthy Aging (NGI:
620 05060810). Measurement of serum fibrinogen was supported by a grant from the Scottish Executive
621 Chief Scientist Office, Health Services Research Committee grant number CZG/4/306. Prof. Dr. J.W.
622 Jukema is an Established Clinical Investigator of the Netherlands Heart Foundation (2001 D 032). The
623 generation and management of GWAS genotype data for the **Rotterdam Study** is supported by the
624 Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012).
625 This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the
626 Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project
627 nr. 050-060-810. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University,
628 Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research
629 Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry
630 for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam.
631 Abbas Dehghan is supported by NWO grant (veni, 916.12.154) and the EUR Fellowship. The **SardiNIA**
632 (‘‘ProgeNIA’’) team was supported by Contract NO1-AG-1–2109 from the NIA. This research was
633 supported by the Intramural Research Program of the NIH, National Institute on Aging, by Sardinian
634 Autonomous Region (L.R. no. 7/2009) grant cRP3-154, and by grant FaReBio2011 “Farmaci e Reti

635 Biotecnologiche di Qualità". **SHIP** is part of the Community Medicine Research net of the University of
636 Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no.
637 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs as well as the Social Ministry of
638 the Federal State of Mecklenburg - West Pomerania. Genome- wide data have been supported by the
639 Federal Ministry of Education and Research (grant no. 03ZIK012) and a joint grant from Siemens
640 Healthcare, Erlangen, Germany and the Federal State of Mecklenburg West Pomerania. Computing
641 resources have been made available by the Leibniz Supercomputing Centre of the Bavarian Academy of
642 Sciences and Humanities (HLRB project h1231). The University of Greifswald is a member of the 'Center
643 of Knowledge Interchange' program of the Siemens AG and the Caché Campus program of the
644 InterSystems GmbH. This work is also part of the research project Greifswald Approach to Individualized
645 Medicine (GANI_MED). The GANI_MED consortium is funded by the Federal Ministry of Education
646 and Research and the Ministry of Cultural Affairs of the Federal State of Mecklenburg – West Pomerania
647 (03IS2061A). **TwinsUK**. The study was funded by the Wellcome Trust; European Community's Seventh
648 Framework Programme (FP7/2007-2013). The study also receives support from the National Institute for
649 Health Research (NIHR) Clinical Research Facility at Guy's & St Thomas' NHS Foundation Trust and
650 NIHR Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's
651 College London. Tim Spector is an NIHR senior Investigator and is holder of an ERC Advanced Principal
652 Investigator award. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and
653 National Eye Institute via NIH/CIDR. The **WGHS** is supported by HL043851 and HL080467 from the
654 National Heart, Lung, and Blood Institute and CA047988 from the National Cancer Institute, the Donald
655 W. Reynolds Foundation and the Fondation Leducq, with collaborative scientific support and funding for
656 genotyping provided by Amgen.

657

658

659 **References**

660 1. International HapMap Consortium. The International HapMap Project. Nature.
661 2003;426(6968):789-96. doi: 10.1038/nature02168. PubMed PMID: 14685227.
662 2. Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, et al. Defining the role of common
663 variation in the genomic and biological architecture of adult human height. Nat Genet.
664 2014;46(11):1173-86. Epub 2014/10/06. doi: ng.3097 [pii]

665 10.1038/ng.3097. PubMed PMID: 25282103; PubMed Central PMCID: PMC4250049.
666 3. Global Lipids Genetics C, Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, et al.
667 Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013;45(11):1274-83. Epub
668 2013/10/08. doi: ng.2797 [pii]

669 10.1038/ng.2797. PubMed PMID: 24097068; PubMed Central PMCID: PMC3838666.
670 4. Smith NL, Huffman JE, Strachan DP, Huang J, Dehghan A, Trompet S, et al. Genetic predictors of
671 fibrin D-dimer levels in healthy adults. Circulation. 2011;123(17):1864-72. Epub 2011/04/20. doi:
672 CIRCULATIONAHA.110.009480 [pii]

673 10.1161/CIRCULATIONAHA.110.009480. PubMed PMID: 21502573; PubMed Central PMCID:
674 PMC3095913.
675 5. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass
676 index yield new insights for obesity biology. Nature. 2015;518(7538):197-206. Epub 2015/02/13. doi:
677 nature14177 [pii]

678 10.1038/nature14177. PubMed PMID: 25673413; PubMed Central PMCID: PMC4382211.
679 6. International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret GB,
680 Munroe PB, Rice KM, Bochud M, Johnson AD, et al. Genetic variants in novel pathways influence blood
681 pressure and cardiovascular disease risk. Nature. 2011;478(7367):103-9. Epub 2011/09/13. doi:
682 nature10405 [pii]

683 10.1038/nature10405. PubMed PMID: 21909115; PubMed Central PMCID: PMC3340926.
684 7. Huang J, Sabater-Lleal M, Asselbergs FW, Tregouet D, Shin SY, Ding J, et al. Genome-wide
685 association study for circulating levels of PAI-1 provides novel insights into its regulation. Blood.
686 2012;120(24):4873-81. Epub 2012/09/20. doi: blood-2012-06-436188 [pii]

687 10.1182/blood-2012-06-436188. PubMed PMID: 22990020; PubMed Central PMCID: PMC3520624.
688 8. Huang J, Huffman JE, Yamakuchi M, Trompet S, Asselbergs FW, Sabater-Lleal M, et al. Genome-
689 wide association study for circulating tissue plasminogen activator levels and functional follow-up
690 implicates endothelial STXBP5 and STX2. Arterioscler Thromb Vasc Biol. 2014;34(5):1093-101. Epub
691 2014/03/01. doi: ATVBAHA.113.302088 [pii]

692 10.1161/ATVBAHA.113.302088. PubMed PMID: 24578379; PubMed Central PMCID: PMC4009733.
693 9. Estrada K, Styrkarsdottir U, Evangelou E, Hsu YH, Duncan EL, Ntzani EE, et al. Genome-wide
694 meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture.
695 Nat Genet. 2012;44(5):491-501. Epub 2012/04/17. doi: ng.2249 [pii]

696 10.1038/ng.2249. PubMed PMID: 22504420; PubMed Central PMCID: PMC3338864.
697 10. CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, et
698 al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet.
699 2013;45(1):25-33. Epub 2012/12/04. doi: ng.2480 [pii]

700 10.1038/ng.2480. PubMed PMID: 23202125; PubMed Central PMCID: PMC3679547.
701 11. Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Lu C, et al. Meta-analysis of genome-wide
702 association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. Circulation.
703 2011;123(7):731-8. Epub 2011/02/09. doi: CIRCULATIONAHA.110.948570 [pii]

704 10.1161/CIRCULATIONAHA.110.948570. PubMed PMID: 21300955; PubMed Central PMCID:
705 PMC3147232.
706 12. Smith NL, Chen MH, Dehghan A, Strachan DP, Basu S, Soranzo N, et al. Novel associations of
707 multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE
708 (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium. Circulation.
709 2010;121(12):1382-92. Epub 2010/03/17. doi: CIRCULATIONAHA.109.869156 [pii]

710 10.1161/CIRCULATIONAHA.109.869156. PubMed PMID: 20231535; PubMed Central PMCID:
711 PMC2861278.
712 13. Loth DW, Artigas MS, Gharib SA, Wain LV, Franceschini N, Koch B, et al. Genome-wide
713 association analysis identifies six new loci associated with forced vital capacity. Nat Genet.
714 2014;46(7):669-77. Epub 2014/06/16. doi: ng.3011 [pii]

715 10.1038/ng.3011. PubMed PMID: 24929828; PubMed Central PMCID: PMC4140093.
716 14. Zuk O, Schaffner SF, Samocha K, Do R, Hechter E, Kathiresan S, et al. Searching for missing
717 heritability: designing rare variant association studies. Proc Natl Acad Sci U S A. 2014;111(4):E455-64.
718 Epub 2014/01/21. doi: 1322563111 [pii]

719 10.1073/pnas.1322563111. PubMed PMID: 24443550; PubMed Central PMCID: PMC3910587.
720 15. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM,
721 et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56-
722 65. doi: 10.1038/nature11632. PubMed PMID: 23128226; PubMed Central PMCID: PMC3498066.
723 16. Zheng HF, Rong JJ, Liu M, Han F, Zhang XW, Richards JB, et al. Performance of genotype
724 imputation for low frequency and rare variants from the 1000 genomes. PLoS One.
725 2015;10(1):e0116487. doi: 10.1371/journal.pone.0116487. PubMed PMID: 25621886; PubMed Central
726 PMCID: PMC4306552.
727 17. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Y AR, H KF, et al. Reference-based phasing using
728 the Haplotype Reference Consortium panel. Nat Genet. 2016;48(11):1443-8. doi: 10.1038/ng.3679.
729 PubMed PMID: 27694958; PubMed Central PMCID: PMCPMC5096458.
730 18. Consortium UK, Walter K, Min JL, Huang J, Crooks L, Memari Y, et al. The UK10K project
731 identifies rare variants in health and disease. Nature. 2015;526(7571):82-90. doi: 10.1038/nature14962.
732 PubMed PMID: 26367797; PubMed Central PMCID: PMCPMC4773891.
733 19. Rich SS, Wang ZY, Sturcke A, Ziyabari L, Feolo M, O'Donnell CJ, et al. Rapid evaluation of
734 phenotypes, SNPs and results through the dbGaP CHARGE Summary Results site. Nat Genet.
735 2016;48(7):702-3. doi: 10.1038/ng.3582. PubMed PMID: 27350599.
736 20. Huang J, Ellinghaus D, Franke A, Howie B, Li Y. 1000 Genomes-based imputation identifies novel
737 and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. Eur J Hum
738 Genet. 2012;20(7):801-5. Epub 2012/02/02. doi: ejhg20123 [pii]

739 10.1038/ejhg.2012.3. PubMed PMID: 22293688; PubMed Central PMCID: PMC3376268.
740 21. Li MX, Yeung JM, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests
741 and significant p-value thresholds in commercial genotyping arrays and public imputation reference
742 datasets. Hum Genet. 2012;131(5):747-56. Epub 2011/12/07. doi: 10.1007/s00439-011-1118-2. PubMed
743 PMID: 22143225; PubMed Central PMCID: PMC3325408.

744 22. Germain M, Saut N, Oudot-Mellakh T, Letenneur L, Dupuy AM, Bertrand M, et al. Caution in
745 interpreting results from imputation analysis when linkage disequilibrium extends over a large distance:
746 a case study on venous thrombosis. PLoS One. 2012;7(6):e38538. Epub 2012/06/08. doi:
747 10.1371/journal.pone.0038538

748 PONE-D-12-01409 [pii]. PubMed PMID: 22675575; PubMed Central PMCID: PMC3366937.
749 23. Machiela MJ, Chen C, Liang L, Diver WR, Stevens VL, Tsilidis KK, et al. One thousand genomes
750 imputation in the National Cancer Institute Breast and Prostate Cancer Cohort Consortium aggressive
751 prostate cancer genome-wide association study. Prostate. 2013;73(7):677-89. Epub 2012/12/21. doi:
752 10.1002/pros.22608. PubMed PMID: 23255287; PubMed Central PMCID: PMC3962143.
753 24. Schumacher FR, Berndt SI, Siddiq A, Jacobs KB, Wang Z, Lindstrom S, et al. Genome-wide
754 association study identifies new prostate cancer susceptibility loci. Hum Mol Genet. 2011;20(19):3867-
755 75. Epub 2011/07/12. doi: ddr295 [pii]

756 10.1093/hmg/ddr295. PubMed PMID: 21743057; PubMed Central PMCID: PMC3168287.
757 25. Wood AR, Perry JR, Tanaka T, Hernandez DG, Zheng HF, Melzer D, et al. Imputation of variants
758 from the 1000 Genomes Project modestly improves known associations and can identify low-frequency
759 variant-phenotype associations undetected by HapMap based imputation. PLoS One. 2013;8(5):e64343.
760 doi: 10.1371/journal.pone.0064343. PubMed PMID: 23696881; PubMed Central PMCID: PMC3655956.
761 26. Springelkamp H, Iglesias AI, Cuellar-Partida G, Amin N, Burdon KP, van Leeuwen EM, et al.
762 ARHGEF12 influences the risk of glaucoma by increasing intraocular pressure. Hum Mol Genet. 2015.
763 Epub 2015/02/01. doi: ddv027 [pii]

764 10.1093/hmg/ddv027. PubMed PMID: 25637523.
765 27. Hysi PG, Cheng CY, Springelkamp H, Macgregor S, Bailey JN, Wojciechowski R, et al. Genome-
766 wide analysis of multi-ancestry cohorts identifies new loci influencing intraocular pressure and
767 susceptibility to glaucoma. Nat Genet. 2014;46(10):1126-30. Epub 2014/09/01. doi: ng.3087 [pii]

768 10.1038/ng.3087. PubMed PMID: 25173106; PubMed Central PMCID: PMC4177225.
769 28. Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic
770 influences on human blood metabolites. Nat Genet. 2014;46(6):543-50. Epub 2014/05/13. doi: ng.2982
771 [pii]

772 10.1038/ng.2982. PubMed PMID: 24816252; PubMed Central PMCID: PMC4064254.
773 29. Major E, Rigo K, Hague T, Berces A, Juhos S. HLA typing from 1000 genomes whole genome and
774 whole exome illumina data. PLoS One. 2013;8(11):e78410. Epub 2013/11/14. doi:
775 10.1371/journal.pone.0078410

776 PONE-D-13-23280 [pii]. PubMed PMID: 24223151; PubMed Central PMCID: PMC3819389.
777 30. Hosomichi K, Jinam TA, Mitsunaga S, Nakaoka H, Inoue I. Phase-defined complete sequencing of
778 the HLA genes by next-generation sequencing. BMC Genomics. 2013;14:355. Epub 2013/05/30. doi:
779 1471-2164-14-355 [pii]

780 10.1186/1471-2164-14-355. PubMed PMID: 23714642; PubMed Central PMCID: PMC3671147.
781 31. de Bakker PI, Raychaudhuri S. Interrogating the major histocompatibility complex with high-
782 throughput genomics. Hum Mol Genet. 2012;21(R1):R29-36. Epub 2012/09/15. doi: dds384 [pii]

783 10.1093/hmg/dds384. PubMed PMID: 22976473; PubMed Central PMCID: PMC3459647.
784 32. Mathieson I, McVean G. Differential confounding of rare and common variants in spatially
785 structured populations. Nat Genet. 2012;44(3):243-6. Epub 2012/02/07. doi: ng.1074 [pii]

786 10.1038/ng.1074. PubMed PMID: 22306651; PubMed Central PMCID: PMC3303124.
787 33. Shungin D, Winkler TW, Croteau-Chonka DC, Ferreira T, Locke AE, Magi R, et al. New genetic loci
788 link adipose and insulin biology to body fat distribution. Nature. 2015;518(7538):187-96. Epub
789 2015/02/13. doi: nature14132 [pii]

790 10.1038/nature14132. PubMed PMID: 25673412; PubMed Central PMCID: PMC4338562.
791 34. Verhaaren BF, Debette S, Bis JC, Smith JA, Ikram MK, Adams HH, et al. Multi-Ethnic Genome-
792 Wide Association Study of Cerebral White Matter Hyperintensities on MRI. Circ Cardiovasc Genet. 2015.
793 Epub 2015/02/11. doi: CIRCGENETICS.114.000858 [pii]

794 10.1161/CIRCGENETICS.114.000858. PubMed PMID: 25663218.
795 35. Geller F, Feenstra B, Carstensen L, Pers TH, van Rooij IA, Korberg IB, et al. Genome-wide
796 association analyses identify variants in developmental genes associated with hypospadias. Nat Genet.
797 2014;46(9):957-63. Epub 2014/08/12. doi: ng.3063 [pii]

798 10.1038/ng.3063. PubMed PMID: 25108383.
799 36. Feenstra B, Pasternak B, Geller F, Carstensen L, Wang T, Huang F, et al. Common variants
800 associated with general and MMR vaccine-related febrile seizures. Nat Genet. 2014;46(12):1274-82.
801 Epub 2014/10/27. doi: ng.3129 [pii]

802 10.1038/ng.3129. PubMed PMID: 25344690; PubMed Central PMCID: PMC4244308.
803 37. Germain M, Chasman DI, de Haan H, Tang W, Lindstrom S, Weng LC, et al. Meta-analysis of
804 65,734 Individuals Identifies TSPAN15 and SLC44A2 as Two Susceptibility Loci for Venous
805 Thromboembolism. Am J Hum Genet. 2015;96(4):532-42. Epub 2015/03/17. doi: S0002-9297(15)00051-
806 8 [pii]

807 10.1016/j.ajhg.2015.01.019. PubMed PMID: 25772935; PubMed Central PMCID: PMC4385184.
808 38. Surakka I, Horikoshi M, Magi R, Sarin AP, Mahajan A, Lagou V, et al. The impact of low-frequency
809 and rare variants on lipid levels. Nat Genet. 2015. doi: 10.1038/ng.3300. PubMed PMID: 25961943.
810 39. Horikoshi M, Mgi R, van de Bunt M, Surakka I, Sarin AP, Mahajan A, et al. Discovery and Fine-
811 Mapping of Glycaemic and Obesity-Related Trait Loci Using High-Density Imputation. PLoS Genet.
812 2015;11(7):e1005230. Epub 2015/07/02. doi: 10.1371/journal.pgen.1005230

813 PGENETICS-D-14-03245 [pii]. PubMed PMID: 26132169; PubMed Central PMCID: PMC4488845.
814 40. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000
815 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet.
816 2015;47(10):1121-30. doi: 10.1038/ng.3396. PubMed PMID: 26343387; PubMed Central PMCID:
817 PMCPMC4589895.
818 41. Panagiotou OA, Evangelou E, Ioannidis JP. Genome-wide significant associations for variants
819 with minor allele frequency of 5% or less--an overview: A HuGE review. Am J Epidemiol.
820 2010;172(8):869-89. Epub 2010/09/30. doi: kwq234 [pii]

821 10.1093/aje/kwq234. PubMed PMID: 20876667.
822 42. de Vries PS, Boender J, Sonneveld MA, Rivadeneira F, Ikram MA, Rottensteiner H, et al. Genetic
823 variants in the ADAMTS13 and SUPT3H genes are associated with ADAMTS13 activity. Blood.
824 2015;125(25):3949-55. doi: 10.1182/blood-2015-02-629865. PubMed PMID: 25934476.
825 43. Huffman JE, de Vries PS, Morrison AC, Sabater-Lleal M, Kacprowski T, Auer PL, et al. Rare and
826 low-frequency variants and their association with plasma levels of fibrinogen, FVII, FVIII, and vWF.
827 Blood. 2015. Epub 2015/06/25. doi: blood-2015-02-624551 [pii]

828 10.1182/blood-2015-02-624551. PubMed PMID: 26105150.
829 44. de Vries PS, Chasman DI, Sabater-Lleal M, Chen MH, Huffman JE, Steri M, et al. A meta-analysis
830 of 120,246 individuals identifies 18 new loci for fibrinogen concentration. Hum Mol Genet. 2015. Epub
831 2015/11/13. doi: ddv454 [pii]

832 10.1093/hmg/ddv454. PubMed PMID: 26561523.
833 45. Abecasis G, Natarajan P, Peloso G, Lee S, NHLBI Trans-Omics for Precision Medicine and
834 TOPMed Anthropometry and Lipids Working Groups. TOPMed: Early insights from sequencing and
835 analysis of 45,934 deep human genomes. . 66th Annual Meeting of The American Society of Human
836 Genetics; Vancouver2016.
837 46. Sabater-Lleal M, Huang J, Chasman D, Naitza S, Dehghan A, Johnson AD, et al. Multiethnic meta-
838 analysis of genome-wide association studies in >100 000 subjects identifies 23 fibrinogen-associated Loci
839 but no strong evidence of a causal association between circulating fibrinogen and cardiovascular
840 disease. Circulation. 2013;128(12):1310-24. Epub 2013/08/24. doi: CIRCULATIONAHA.113.002251 [pii]

841 10.1161/CIRCULATIONAHA.113.002251. PubMed PMID: 23969696; PubMed Central PMCID:
842 PMC3842025.
843 47. Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to
844 estimate haplotypes and unobserved genotypes. Genet Epidemiol. 2010;34(8):816-34. Epub
845 2010/11/09. doi: 10.1002/gepi.20533. PubMed PMID: 21058334; PubMed Central PMCID: PMC3175618.
846 48. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. Annu Rev Genomics Hum Genet.
847 2009;10:387-406. Epub 2009/09/01. doi: 10.1146/annurev.genom.9.081307.164242. PubMed PMID:
848 19715440; PubMed Central PMCID: PMC2925172.
849 49. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the
850 next generation of genome-wide association studies. PLoS Genet. 2009;5(6):e1000529. Epub
851 2009/06/23. doi: 10.1371/journal.pgen.1000529. PubMed PMID: 19543373; PubMed Central PMCID:
852 PMC2689936.
853 50. Servin B, Stephens M. Imputation-based analysis of association studies: candidate regions and
854 quantitative traits. PLoS Genet. 2007;3(7):e114. Epub 2007/08/07. doi: 06-PLGE-RA-0550 [pii]

855 10.1371/journal.pgen.0030114. PubMed PMID: 17676998; PubMed Central PMCID: PMC1934390.
856 51. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association
857 scans. Bioinformatics. 2010;26(17):2190-1. Epub 2010/07/10. doi: btq340 [pii]

858 10.1093/bioinformatics/btq340. PubMed PMID: 20616382; PubMed Central PMCID: PMC2922887.
859 52. Hofman A, Brusselle GG, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, et al. The
860 Rotterdam Study: 2016 objectives and design update. Eur J Epidemiol. 2015;30(8):661-708. Epub
861 2015/09/21. doi: 10.1007/s10654-015-0082-x

862 10.1007/s10654-015-0082-x [pii]. PubMed PMID: 26386597; PubMed Central PMCID: PMC4579264.

863

864

Table 1. Non-overlapping loci that were significant in either the HapMap or 1000G GWA studies.

	HapMap					1000G				
Locus	**Lead Variant**	**Beta**	***P*-value**	**MAF**	**Imputation Quality**	**Lead Variant**	**Beta**	***P*-value**	**MAF**	**Imputation Quality**
Significant in 1000G										
1q42.13	rs10489615	0.0052	8.3×10^{-07}	0.38	0.97	rs10864726	0.0059	1.1×10^{-08}	0.40	0.96
3q21.1	rs16834024	0.0173	1.4×10^{-07}	0.03	0.79	rs1976714	0.0064	7.5×10^{-09}	0.35	0.89
4p16.3	rs2699429	0.0060	1.3×10^{-07}	0.43	0.87	rs59950280	0.0080	2.5×10^{-11}	0.34	0.80
7p15.3	rs1029738	0.0057	3.2×10^{-07}	0.30	1.00	rs61542988	0.0065	3.1×10^{-08}	0.25	0.98
8p23.1	rs7004769	0.0062	1.4×10^{-06}	0.20	1.00	rs7012814	0.0061	8.0×10^{-09}	0.47	0.91
11q12.2	rs7935829	0.0056	5.6×10^{-08}	0.40	0.99	rs11230201	0.0060	3.0×10^{-09}	0.41	0.99
Significant in HapMap										
6p21.3	rs12528797	0.0095	8.5×10^{-09}	0.11	0.98	rs116134220	0.0082	7.9×10^{-06}	0.49	0.89

Further detail about these loci and the lead variants is provided in S3 Table.

Abbreviations: HapMap refers to the GWA study using imputation based on the HapMap project. 1000G refers to the GWA study using imputation based on the 1000 Genomes Project. Variants were coded according to the fibrinogen increasing allele. MAF refers to minor allele frequency.

Table 2. Overlapping loci that were significant in both the HapMap and 1000G GWA studies.

	HapMap					1000G				
Locus	Lead Variant	Beta	*P*-value	MAF	Imputation Quality	Lead Variant	Beta	*P*-value	MAF	Imputation Quality
1p31.3	rs4655582	0.0069	4.8×10^{-11}	0.38	0.98	rs2376015	0.0075	5.1×10^{-12}	0.35	0.91
1q21.3	rs8192284	0.0115	8.9×10^{-29}	0.40	0.97	rs61812598	0.0114	1.8×10^{-28}	0.39	0.99
1q44	rs12239046	0.0103	9.7×10^{-21}	0.38	0.99	rs12239046	0.0102	9.8×10^{-22}	0.38	0.99
2q12	rs1558643	0.0066	5.8×10^{-10}	0.40	0.99	rs1558643	0.0063	6.0×10^{-10}	0.40	0.98
2q13	rs6734238	0.0106	1.7×10^{-23}	0.41	0.99	rs6734238	0.0106	3.7×10^{-24}	0.41	1.00
2q34	rs715	0.0092	9.1×10^{-14}	0.32	0.92	rs715	0.0082	1.7×10^{-13}	0.32	0.89
2q37.3	rs1476698	0.0075	4.2×10^{-12}	0.36	1.00	rs59104589	0.0081	2.4×10^{-14}	0.34	0.98
3q22.2	rs548288	0.0113	6.6×10^{-21}	0.24	0.99	rs150213942	0.0117	3.1×10^{-21}	0.23	0.95
4q31.3	rs2227401	0.0311	4.7×10^{-134}	0.21	0.95	rs72681211	0.0313	1.3×10^{-142}	0.20	0.99
5q31.1	rs1012793	0.0208	4.4×10^{-60}	0.21	0.98	rs1012793	0.0207	1.0×10^{-58}	0.20	0.98
7p21.1	rs10950690	0.0071	9.9×10^{-12}	0.48	0.94	rs12699921	0.0071	1.3×10^{-12}	0.47	0.98
7q14.2	rs2710804	0.0061	9.3×10^{-09}	0.38	0.98	rs2710804	0.0057	4.3×10^{-08}	0.38	0.99
7q36.1	rs13226190	0.008	2.2×10^{-10}	0.21	0.99	rs13234724	0.0076	1.6×10^{-09}	0.21	0.99
8q24.3	rs7464572	0.0066	2.4×10^{-09}	0.40	0.98	rs11136252	0.0056	4.6×10^{-08}	0.42	0.96
9q22.2	rs7873907	0.006	5.4×10^{-09}	0.50	0.96	rs3138493	0.006	3.5×10^{-09}	0.48	0.98
10q21.3	rs10761756	0.0093	5.4×10^{-20}	0.48	1.00	rs7916868	0.0097	1.2×10^{-21}	0.49	0.97
11p12	rs7937127	0.0083	2.3×10^{-10}	0.18	0.99	rs7934094	0.0081	2.9×10^{-10}	0.22	0.90
12q13.12	rs1521516	0.0072	3.0×10^{-11}	0.36	1.00	12:51042486	0.0073	4.9×10^{-12}	0.36	0.98
12q24.12	rs3184504	0.0066	1.1×10^{-10}	0.49	0.97	rs4766897	0.009	3.8×10^{-12}	0.34	0.64
14q24.1	rs194741	0.0092	8.3×10^{-14}	0.25	0.95	rs194714	0.0086	3.7×10^{-13}	0.25	0.97
15q15.1	rs1703755	0.0088	1.8×10^{-09}	0.14	0.96	rs8026198	0.009	5.9×10^{-10}	0.15	0.93
15q21.2	rs12915052	0.0069	2.4×10^{-10}	0.31	1.00	rs11630054	0.0067	3.3×10^{-10}	0.34	0.99
16q12.2	rs12598049	0.0074	3.0×10^{-11}	0.32	0.99	rs6499550	0.007	8.2×10^{-11}	0.32	0.98
16q22.2	rs11864453	0.0057	4.6×10^{-08}	0.40	0.99	rs1035560	0.0058	1.2×10^{-08}	0.40	0.99
17q21.2	rs7224737	0.0073	2.2×10^{-09}	0.23	0.99	rs7224737	0.0068	5.2×10^{-09}	0.24	1.00

17q25.1	rs10512597	0.0078	2.2×10^{-08}	0.18	0.94	rs35489971	0.0077	1.6×10^{-08}	0.18	0.94
20q13.12	rs1800961	0.0183	6.8×10^{-09}	0.03	0.95	rs1800961	0.0178	1.7×10^{-09}	0.03	0.99
21q22.2	rs4817986	0.0091	1.9×10^{-14}	0.28	0.95	rs9808651	0.0093	5.4×10^{-16}	0.28	0.94
22q13.33	rs6010044	0.0074	2.5×10^{-08}	0.20	0.89	rs75347843	0.0082	4.3×10^{-08}	0.19	0.76

Further detail about these loci and the lead variants is provided in S3 Table.

Abbreviations: HapMap refers to the GWA study using imputation based on the HapMap project. 1000G refers to the GWA study using imputation based on the 1000 Genomes Project. Variants were coded according to the fibrinogen increasing allele. MAF refers to minor allele frequency.

Figure 1

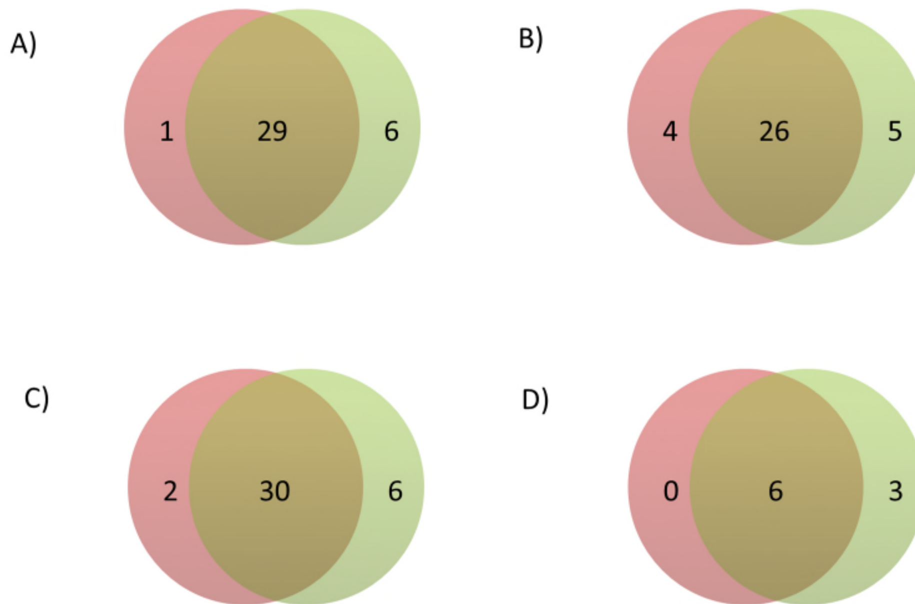

Figure 2

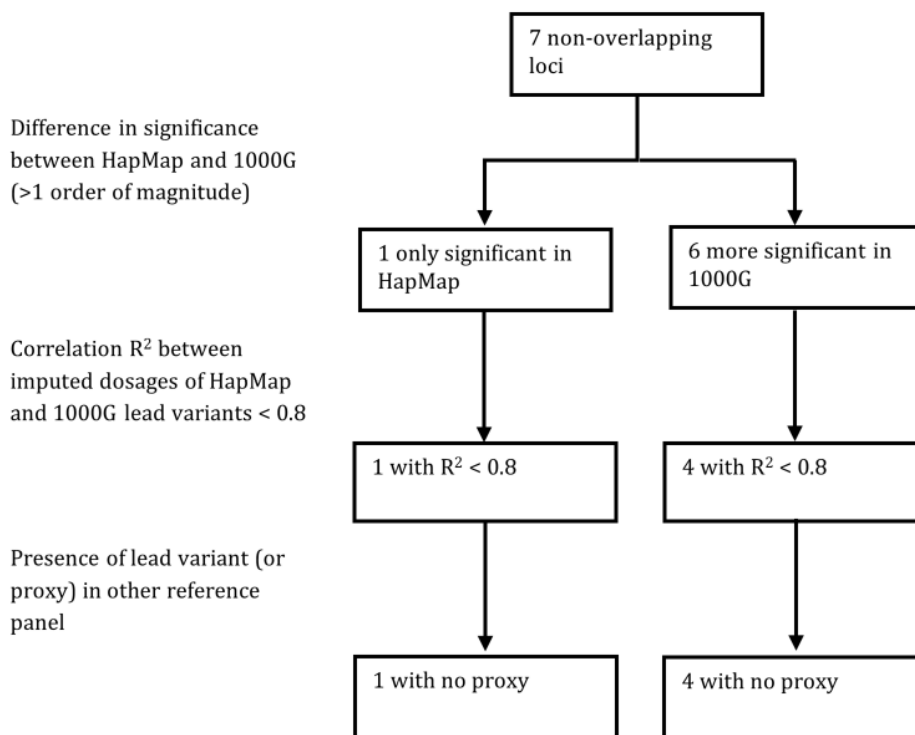

Figure 3

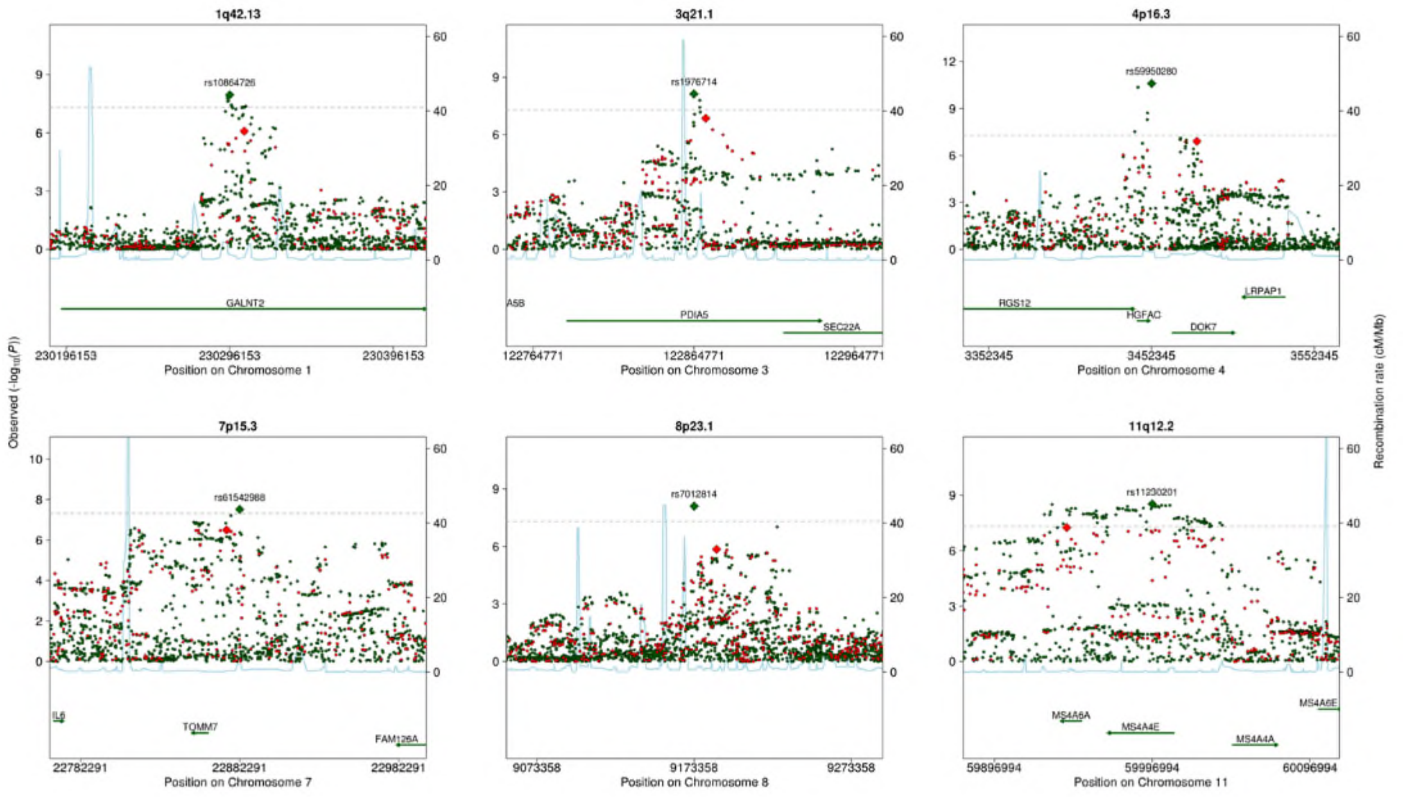

Figure 4

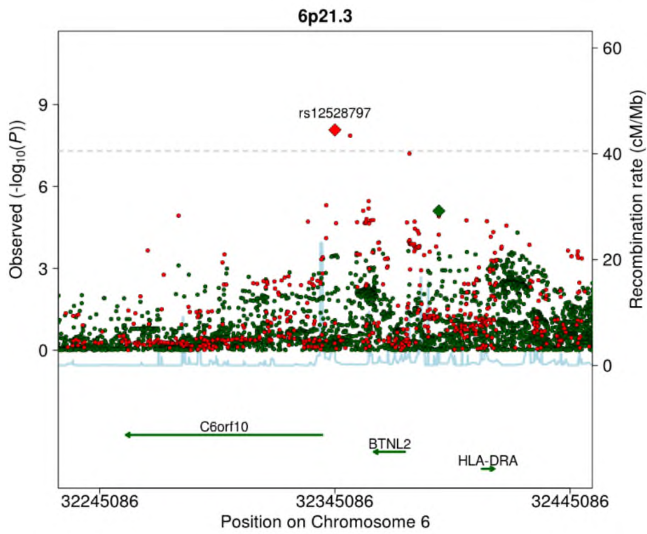

Figure 5