J Alzheimers Dis. 2018;64(1):49-54. doi: 10.3233/JAD-171104. PubMed PMID: 29865051.

Short Communication

Accuracy of inferred *APOE* genotypes, for a range of genotyping arrays and imputation reference panels.

Michelle K Lupton[a*], Sarah E Medland[a], Scott D Gordon[a], Tabatha Goncalves[a], Stuart MacGregor[a], David A Mackey[b], Terri L Young[c], David L Duffy[a], Peter M Visscher[d,e], Naomi R Wray[d,e], Dale R Nyholt[f], Lisa Bain[a], Manuel A Ferreira[a], Anjali K Henders[d], Leanne Wallace[d], Grant W Montgomery[d,e], Margaret J Wright[d,g], Nicholas G Martin[a].

a. *QIMR Berghofer Medical Research Institute, Brisbane, Australia.*
b. *University of Western Australia Centre for Ophthalmology and Visual Science, Lions Eye Institute, Perth, Australia.*
c. *Department of Ophthalmology and Visual Sciences, University of Wisconsin, Madison, Wisconsin, USA.*
d. *Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia.*
e. *Queensland Brain Institute, University of Queensland, Brisbane, Australia*.
f. *Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia.*
g. *Centre for Advanced Imaging, University of Queensland, Brisbane, Australia*

*Correspondence to: Michelle K. Lupton, PhD, QIMR Berghofer Medical Research Institute.
300 Herston Road, Herston QLD 4030, Australia.
Tel.: +61 7 3845 3947; E-mail: Michelle.Lupton@QIMRBerghofer.edu.au.

Abstract

Cohort studies investigating aging and dementia require *APOE* genotyping. We compared directly measured *APOE* genotypes to 'hard-call' genotypes derived from imputing genome-wide genotyping data from a range of platforms using several imputation panels. Older GWAS arrays imputed to 1000 Genomes Project (1KGP) phases and the Haplotype Reference Consortium (HRC) reference panels were able to achieve concordance rates of over 98% with stringent quality control (hard-call-threshold .8). However, this resulted in high levels of missingness (>12% with 1KGP and 5% with HRC). With recent GWAS arrays, concordance of 99% could be obtained with relatively lenient QC, resulting in no missingness.

Key Words: Alzheimer Disease, ApoE Receptor, Genetic Association studies, Computational Biology, Cohort Studies

Introduction

APOE is a major genetic risk factor for late-onset Alzheimer's disease (AD), explaining almost 30% of the population-attributable risk, and by far the greatest contributing genetic risk factor of all variants identified from genome-wide association studies (GWAS) [1]. *APOE* ε3 is the most common and the neutral risk allele. Heterozygotes carrying one risk allele, *APOE* ε4, have four times the risk of AD, and homozygotes carrying two *APOE* ε4 alleles have 15 times the risk compared to *APOE* ε3 homozygotes [2]. A third low frequency allele *APOE* ε2 confers decreased risk, but with a more subtle effect [2]. ApoE plays a major role in the regulation of lipid and lipoprotein levels in the blood [3].

APOE genotype is often directly genotyped. As many large cohorts carry out genome-wide genotyping using arrays, it would be cheaper and more convenient to be able to ascertain *APOE* genotype from this already available data. If not directly included on the array, the genotypes required can be estimated by imputation using a reference dataset [4]. It has previously been shown that using array-based genotyping (using the Affymetrix Kaiser Axiom array) imputed to the 1KGP Phase 1 panel compared to direct genotyping finds 90% agreement for ε2/ ε3/ε4 genotypes and 93% agreement for predicting ε4 status, but with high levels of missing data [5]. Radmanesh et al. found similar results for the Illumina HumanHap610 array imputed to 1KGP Phase 1 panel [6]. Correlations of the *APOE* SNPs rs7412 and rs429558 between imputed and genotype data were 0.9-0.94 (Kappa coefficients), but with up to 19% missing. Changing imputation algorithm parameters reduced the level of missingness to 5-9% but also reduced the accuracy. Here we extend these analyses and report the concordance between imputed and directly measured *APOE* genotypes for a range of genotyping arrays and imputation reference panels in a large sample size.

Methods

Participants in this study comprise individuals recruited for studies led by the Genetic Epidemiology group at QIMR Berghofer Medical Research Institute. The sample set is made up of Australians of European Ancestry including twins and their relatives who volunteered for studies on risk factors or biomarkers for physical or psychiatric conditions (described in [7, 8]).

APOE genotyping was performed using TaqMan SNP genotyping assays on an ABI Prism 7900HT and analyzed using SDS software (Applied Biosystems). The three main *APOE* alleles- ε2, ε3 and ε4 -differ at two residues, consisting of a two SNP haplotype. The SNPs rs429358 and rs7412 were determined by allelic discrimination assays based on fluorogenic 5' nuclease activity and the allele inferred.

Genome-wide genotyping was available for the samples on a range of arrays. 4190 individuals had GWAS data from the Illumina chips designed using the HapMap references (317K, 370K, 610K, 660K). 3385 individuals had GWAS data from the more recent Illumina arrays designed using the 1KGP data (Core+Exome, PsychArray, OmniExpress). Rs429358 is not directly genotyped on any of the arrays used. Rs7412 is genotyped on the Illumina Core+Exome and the PsychArray, but was excluded when combined with data from other chips prior to imputation. Both datasets were imputed to three different imputation panels: 1KGP Phase 1 (Version 3, Nov 2010), 1KGP Phase 3 (Release 5, May 2013) [9] and HRC Release 1 [10]. Imputation was carried out using the University of Michigan Imputation Server with standard protocols (https://imputationserver.sph.umich.edu) [11]. *APOE* SNP data was extracted from the imputed datasets. Results from imputation are in dosage format, which yields continuous values taking into account any uncertainty in the number of alleles. Dosage scores were converted to hard call allelic counts (0, 1 or 2 copies of the alternate

allele). Genotypes not above a hard-call threshold were coded as missing. The lower the hard-call-threshold, the fewer values of missing data but the higher the chance of incorrect hard-calls. Therefore a balance is required for optimum allele calls with an acceptable error rate. A modified version of DosageCoverter software (http://genome.sph.umich.edu/wiki/DosageConvertor) was used to convert the genotype probabilities to the best-guess genotype . The genotype with the highest probability in the VCF file is selected, subject to that genotype probability being above the hard-call threshold probability of 0.4. The conversion was repeated using thresholds of 0.6, 0.8 and 0.9. Where rs7412 was directly assayed on the array but excluded prior to imputation there was a good concordance with the imputed value (>99% for all thresholds).

Results

The accuracy of the *APOE* genotyping using the TaqMan SNP genotyping assays was assessed from 3576 duplicate DNA samples, where the genotyping error rate was 0.2%.

For the HapMap-based arrays, the *APOE* genotypes were well imputed using the 1KGP imputation panels (R^2 values 0.79-0.84) and the HRC panel (R^2 values 0.99). Concordances with direct genotypes are shown in Table 1 and Figure 1. Concordance and missingness rates vary between genotypes, due to the differences in imputation accuracy in the SNPs used to decipher the allele. For both the 1KGP imputation panels, concordance rates >98% could only be reached with high levels of missing data. This missingness rate was driven by the rarer ε4 and ε2 alleles. Using a hard-call-threshold of 0.8 for hard genotype calling resulted in 98% concordance for both 1KGP Phase 1v3 and 3v5 imputation panels, but yielded high levels of missingness (12 and 14%, respectively). Minor differences between the 1KGP phase 1 and 3 reference panels are likely due to an increase in sample size and the inclusion of additional world populations in the phase 3 reference, as *APOE* genotype distribution varies

by ancestry[12]. The HRC reference panel improved concordance and missingness, with 99% concordance and 5% missingness at the calling threshold of 0.8. Using a very conservative calling threshold of 0.9 did not markedly improve the concordance and greatly increased the level of missing data.

As one would expect the use of the more recent 1KGP-based arrays improves the calling accuracy of imputed *APOE* genotypes (Table 1 and Figure 1). The imputation accuracy for all panels increased to r^2 values of 0.95-0.99. The calling accuracy was also good for all panels; using a relatively lenient hard-call-threshold of 0.4 resulted in concordance rates of 98.9 to 99.3% with no missingness. Unsurprisingly, given the extremely high concordance, increasing the hard-call-threshold to 0.8 yielded little increase in accuracy over all (concordance rates 99.3 to 99.4). However, there was a marked improvement in the calling of the rarer alleles: for example, the accuracy of the ε2 homozygote from the 1KGP Phase 3v5 imputed data increased from 92.3% at a hard-call-threshold of 0.4 to 95.8 at a threshold of 0.8 (although this resulted in a corresponding increase in missingness).

We went on to use these results to provide acceptably accurate *APOE* genotype data for our cohort study PISA (Prospective Imaging Study of Aging). We had a total of 19,449 samples of European ancestry with no direct *APOE* genotype but imputed GWAS data available. Samples were genotyped on either HapMap or 1KGP-based arrays (N=10,544 and 8,905 respectively). We conservatively selected samples requiring direct *APOE* genotyping using HRC imputed data with a genotype hard call threshold of 0.9. Samples were selected if imputed *APOE* genotype were missing or if the concordance was <99.3% in the genotype group. This included all imputed *APOE* genotypes of ε2ε2, ε2ε4 and ε4ε4 for the HapMap-based arrays and genotypes of ε2ε2 for the 1KGP-based arrays. A total of 1,255 samples required genotyping from the HapMap-based arrays, and 126 from the 1KGP-based arrays. From the HapMap-based arrays 1120 had available DNA and were directly *APOE* genotyped.

Concordance between imputed and directly genotyped was close to predicted from the analysis described above (Table 1), with 97.0, 97.7 and 96.8% concordance for ε2ε2, ε2ε4 and ε4ε4, respectively. From the IKGP-based arrays, 64 had available DNA for direct genotyping, and concordance between imputed for ε2ε2 genotypes was 100%, again close to the predicted concordance (Table 1). Our final HapMap-based array dataset included *APOE* genotype accurate to ≥ 99.4% concordance for 10,409 samples where 10.8% were individually genotyped, and our final 1KGP-based arrays dataset included APOE genotype accurate to ≥99.3% concordance for 8,804 samples where 0.7% were individually genotyped.

Discussion

In agreement with previous analyses [5, 6], using a large sample of European ancestry, we have shown that use of GWAS data from HapMap-based arrays imputed to 1KGP reference panels can give reasonably accurate *APOE* genotype calls, but at the expense of increasing missingness biased towards the rarer alleles of greatest interest. We showed that use of the HRC reference panel improves the calling accuracy. Finally the newer GWAS arrays based on the 1KGP data resulted in excellent imputation of the *APOE* genotypes on all the imputation panels examined.

As the strongest known genetic risk factor for AD, *APOE* genotype is routinely required in cohort studies investigating aging and dementia-related phenotypes. In addition, *APOE* genotype is increasingly used for selecting individuals at high risk of AD for longitudinal studies investigating early stage dementia and for selecting individuals for enrolment into early intervention clinical trials. The information presented here is useful for cohorts where GWAS data is available, to aid in the decision of whether additional genotyping or new imputation analysis is required to obtain reasonably accurate *APOE* genotypes. For studies carrying out association testing where maximising power is the priority, the use of imputed

APOE genotypes could be sufficient. But where accuracy is paramount genotyping may be required. Using our own data as an example, we were able to ascertain *APOE* genotype data using historic GWAS data and minimal amount of additional genotyping.

Acknowledgements

For the GWAS datasets, we acknowledge funding from the Australian National Health and Medical Research Council (NHMRC grants 241944, 389875, 389891, 389892, 389938, 442915, 442981, 496739 and 552485), US National Institutes of Health (NIH grants AA07535, AA10248, AA014041, AA011998, AA013320, AA013321, AA017688, DA012854, NEI- 1R01EY018246-03) and the Australian Research Council (ARC grant DP0770096). MKL is supported by the Prospective Imaging Study of Aging (PISA), NHMRC grant 1095227.

References

[1] Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, Jun G, Destefano AL, Bis JC, Beecham GW, Grenier-Boley B, Russo G, Thornton-Wells TA, Jones N, Smith AV, Chouraki V, Thomas C, Ikram MA, Zelenika D, Vardarajan BN, Kamatani Y, Lin CF, Gerrish A, Schmidt H, Kunkle B, Dunstan ML, Ruiz A, Bihoreau MT, Choi SH, Reitz C, Pasquier F, Hollingworth P, Ramirez A, Hanon O, Fitzpatrick AL, Buxbaum JD, Campion D, Crane PK, Baldwin C, Becker T, Gudnason V, Cruchaga C, Craig D, Amin N, Berr C, Lopez OL, De Jager PL, Deramecourt V, Johnston JA, Evans D, Lovestone S, Letenneur L, Moron FJ, Rubinsztein DC, Eiriksdottir G, Sleegers K, Goate AM, Fievet N, Huentelman MJ, Gill M, Brown K, Kamboh MI, Keller L, Barberger-Gateau P, McGuinness B, Larson EB, Green R, Myers AJ, Dufouil C, Todd S, Wallon D, Love S, Rogaeva E, Gallacher J, St George-Hyslop P, Clarimon J, Lleo A, Bayer A, Tsuang DW, Yu L, Tsolaki M, Bossu P, Spalletta G, Proitsi P, Collinge J, Sorbi S, Sanchez-Garcia F, Fox NC, Hardy J, Naranjo MC, Bosco P, Clarke R, Brayne C, Galimberti D, Mancuso M, Matthews F, European Alzheimer's Disease I, Genetic, Environmental Risk in Alzheimer's D, Alzheimer's Disease Genetic C, Cohorts for H, Aging Research in Genomic E, Moebus S, Mecocci P, Del Zompo M, Maier W, Hampel H, Pilotto A, Bullido M, Panza F, Caffarra P, Nacmias B, Gilbert JR, Mayhaus M, Lannfelt L, Hakonarson H, Pichler S, Carrasquillo MM, Ingelsson M, Beekly D, Alvarez V, Zou F, Valladares O, Younkin SG, Coto E, Hamilton-Nelson KL, Gu W, Razquin C, Pastor P, Mateo I, Owen MJ, Faber KM, Jonsson PV, Combarros O, O'Donovan MC, Cantwell LB, Soininen H, Blacker D, Mead S, Mosley TH, Jr., Bennett DA, Harris TB, Fratiglioni L, Holmes C, de Bruijn RF, Passmore P, Montine TJ, Bettens K, Rotter JI, Brice A, Morgan K, Foroud TM, Kukull WA, Hannequin D, Powell JF, Nalls MA, Ritchie K, Lunetta KL, Kauwe JS, Boerwinkle E, Riemenschneider M, Boada M, Hiltunen M, Martin ER, Schmidt R, Rujescu D, Wang LS, Dartigues JF, Mayeux R, Tzourio C, Hofman A, Nothen MM, Graff C, Psaty BM, Jones L, Haines JL, Holmans PA, Lathrop M, Pericak-Vance MA, Launer LJ, Farrer LA, van Duijn CM, Van Broeckhoven C, Moskvina V, Seshadri S, Williams J, Schellenberg GD, Amouyel P (2013) Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452-1458.

[2] Wolf AB, Caselli RJ, Reiman EM, Valla J (2013) APOE and neuroenergetics: an emerging paradigm in Alzheimer's disease. *Neurobiol Aging* **34**, 1007-1017.

[3] Singhrao SK, Harding A, Chukkapalli S, Olsen I, Kesavalu L, Crean S (2016) Apolipoprotein E Related Co-Morbidities and Alzheimer's Disease. *J Alzheimers Dis* **51**, 935-948.

[4] Nyholt DR, Yu CE, Visscher PM (2009) On Jim Watson's APOE status: genetic information is hard to hide. *Eur J Hum Genet* **17**, 147-149.

[5] Oldmeadow C, Holliday EG, McEvoy M, Scott R, Kwok JB, Mather K, Sachdev P, Schofield P, Attia J (2014) Concordance between direct and imputed APOE genotypes using 1000 Genomes data. *J Alzheimers Dis* **42**, 391-393.

[6] Radmanesh F, Devan WJ, Anderson CD, Rosand J, Falcone GJ, Alzheimer's Disease Neuroimaging I (2014) Accuracy of imputation to infer unobserved APOE epsilon alleles in genome-wide genotyping data. *Eur J Hum Genet* **22**, 1239-1242.

[7] Benyamin B, Ferreira MA, Willemsen G, Gordon S, Middelberg RP, McEvoy BP, Hottenga JJ, Henders AK, Campbell MJ, Wallace L, Frazer IH, Heath AC, de Geus EJ, Nyholt DR, Visscher PM, Penninx BW, Boomsma DI, Martin NG, Montgomery GW, Whitfield JB (2009) Common variants in TMPRSS6 are associated with iron status and erythrocyte volume. *Nat Genet* **41**, 1173-1175.

[8] Painter JN, Anderson CA, Nyholt DR, Macgregor S, Lin J, Lee SH, Lambert A, Zhao ZZ, Roseman F, Guo Q, Gordon SD, Wallace L, Henders AK, Visscher PM, Kraft P, Martin NG, Morris AP, Treloar SA, Kennedy SH, Missmer SA, Montgomery GW, Zondervan KT (2011) Genome-wide association study identifies a locus at 7p15.2 associated with endometriosis. *Nat Genet* **43**, 51-54.

[9] Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR (2015) A global reference for human genetic variation. *Nature* **526**, 68-74.

[10] McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, Kwong A, Timpson N, Koskinen S, Vrieze S, Scott LJ, Zhang H, Mahajan A, Veldink J, Peters U, Pato C, van Duijn CM, Gillies CE, Gandin I, Mezzavilla M, Gilly A, Cocca M, Traglia M, Angius A, Barrett JC, Boomsma D, Branham K, Breen G, Brummett CM, Busonero F, Campbell H, Chan A, Chen S, Chew E, Collins FS, Corbin LJ, Smith GD, Dedoussis G, Dorr M, Farmaki AE, Ferrucci L, Forer L, Fraser RM, Gabriel S, Levy S, Groop L, Harrison T, Hattersley A, Holmen OL, Hveem K, Kretzler M, Lee JC, McGue M, Meitinger T, Melzer D, Min JL, Mohlke KL, Vincent JB, Nauck M, Nickerson D, Palotie A, Pato M, Pirastu N, McInnis M, Richards JB, Sala C, Salomaa V, Schlessinger D, Schoenherr S, Slagboom PE, Small K, Spector T, Stambolian D, Tuke M, Tuomilehto J, Van den Berg LH, Van Rheenen W, Volker U, Wijmenga C, Toniolo D, Zeggini E, Gasparini P, Sampson MG, Wilson JF, Frayling T, de Bakker PI, Swertz MA, McCarroll S, Kooperberg C, Dekker A, Altshuler D, Willer C, Iacono W, Ripatti S, Soranzo N, Walter K, Swaroop A, Cucca F, Anderson CA, Myers RM, Boehnke M, McCarthy MI, Durbin R (2016) A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-1283.

[11] Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, Fuchsberger C (2016) Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287.

[12] Singh PP, Singh M, Mastana SS (2006) APOE distribution in world populations with new data from India and the UK. *Ann Hum Biol* **33**, 279-308.

Table 1. Comparison of measured and imputed *APOE* genotypes showing concordance and missingness. A) for first-generation array (based on HapMap references) datasets; B) second-generation array datasets (based on 1KGP references)

A. HapMap-Based Arrays																							
Imput. Panel	Imput. R^2 values	Hard Call Probability Threshold*	Total (N=4190)			ε2/ε2 (N=27)			ε2/ε3 (N=459)			ε2/ε4 (N=96)			ε3/ε3 (N=2509)			ε3/ε4 (N=1008)			ε4/ε4 (N=91)		
			% Con	% Miss	ESS	% Con	% Miss	ESS	% Con	% Miss	ESS	% Con	% Miss	ESS	% Con	% Miss	ESS	% Con	% Miss	ESS	% Con	% Miss	ESS
1KGP Phase 1	rs7412: 0.79 rs429358: 0.84	0.4	91.8	0.0	3846	92.6	0.0	25	84.5	0.0	388	75.0	0.0	72	97.3	0.0	2441	85.3	0.0	860	69.2	0.0	63
		0.6	94.8	5.1	3770	96.0	7.4	24	89.7	8.9	375	84.0	15.6	68	98.5	2.3	2415	89.6	8.6	825	76.0	13.2	60
		0.8	98.1	12.3	3605	94.1	37.0	16	94.3	20.0	346	92.3	32.3	60	99.1	5.8	2342	97.8	20.4	785	93.3	34.1	56
		0.9	98.1	18.4	3354	92.3	51.2	12	94.8	36.8	275	89.8	49.0	44	98.6	10.5	2214	98.4	24.1	753	92.9	38.5	52
1KGP Phase 3	rs7412: 0.75 rs429358: 0.82	0.4	91.9	0.0	3851	92.6	0.0	25	86.1	0.0	395	77.1	0.0	74	97.3	0.0	2441	84.8	0.0	855	67.0	0.0	61
		0.6	94.1	4.4	3769	95.5	18.5	21	89.3	8.5	375	83.3	18.8	65	97.9	1.6	2417	89.0	7.2	833	70.4	11.0	57
		0.8	98.4	13.6	3562	93.3	44.4	14	96.0	29.6	310	96.0	47.9	48	99.2	6.2	2335	97.7	18.9	799	90.3	31.9	56
		0.9	98.2	21.0	3251	91.7	55.6	11	98.2	50.8	222	97.4	59.4	38	99.6	10.7	2232	98.9	26.5	733	92.3	42.3	48
HRC	rs7412: 0.99 rs429358: 0.99	0.4	96.4	0.0	4039	96.3	0.0	26	93.7	0.0	430	90.6	0.0	87	98.1	0.0	2461	94.3	0.0	951	90.1	0.0	82
		0.6	97.6	1.8	4016	96.3	0.0	26	95.9	3.5	425	93.4	5.2	85	98.9	1.2	2452	96.3	2.4	947	90.1	0.0	82
		0.8	98.9	4.9	3941	95.8	11.1	23	98.6	8.9	412	95.1	15.6	77	99.3	2.5	2429	98.8	7.4	922	93.9	9.9	77
		0.9	99.4	8.5	3811	95.5	18.5	21	99.7	19.4	369	95.8	25.0	69	99.6	4.6	2384	99.4	10.8	894	94.7	16.5	72
B. 1KGP-Based Arrays																							
Imput. Panel	Imput. R^2 values	Hard Call Probability Threshold*	Total (N=3385)			ε2/ε2 (N=26)			ε2/ε3 (N=400)			ε2/ε4 (N=84)			ε3/ε3 (N=1907)			ε3/ε4 (N=882)			ε4/ε4 (N=86)		
			% Con	% Miss	ESS	% Con	% Miss	ESS	% Con	% Miss	ESS	% Con	% Miss	ESS	% Con	% Miss	ESS	% Con	% Miss	ESS	% Con	% Miss	ESS
1KGP Phase 1	rs7412: 0.96 rs429358: 0.99	0.4	98.9	0.0	3348	92.2	0.0	24	98.5	0.0	394	97.8	0.0	82	99 .5	0.0	1897	97.9	0.0	863	98.8	0.0	85
		0.6	99.3	0.6	3341	96.0	3.9	24	99.5	0.8	395	100.0	3.6	81	99.5	0.3	1892	98.7	0.9	863	98.8	0.0	85
		0.8	99.4	1.9	3301	95.8	7.7	23	99.7	3.8	384	100.0	8.3	77	99.5	0.9	1880	99.0	2.5	851	98.8	3.5	82
		0.9	99.4	3.16	3258	95.5	15.4	21	99.7	7.0	371	100.0	13.1	73	99.6	1.7	1867	99.1	3.2	846	98.8	3.5	82
1KGP Phase 3	rs7412: 0.95 rs429358: 0.95	0.4	99.0	0.0	3351	92.3	0.0	24	98.5	0.0	394	100.0	0.0	84	99.5	0.0	1897	98.2	0.0	866	98.8	0.0	85
		0.6	99.0	0.1	3348	96.0	3.9	24	98.7	0.5	393	100.0	0.0	84	99.5	0.1	1896	98.2	0.2	864	98.8	0.0	85
		0.8	99.3	1.7	3304	95.8	7.7	23	99.7	2.8	388	100.0	4.8	80	99.5	0.6	1886	98.8	3.0	845	98.8	4.7	81
		0.9	99.5	3.8	3240	95.0	23.1	19	99.7	6.8	372	100.0	10.7	75	99.6	2.1	1859	99.2	4.5	836	98.8	5.8	80
HRC	rs7412: 0.98 rs429358: 0.99	0.4	99.3	0.0	3361	96.2	0.0	25	99.3	0.0	397	98.8	0.0	83	99.5	0.0	1897	98.9	0.0	872	98.8	0.0	85
		0.6	99.3	0.1	3358	96.2	0.0	25	99.5	0.3	397	98.8	0.0	83	99.5	0.0	1897	99.0	0.1	872	98.8	0.0	85
		0.8	99.4	0.5	3348	96.2	0.0	25	99.5	1.0	394	98.8	0.0	83	99.5	0.21	1893	99.1	1.0	865	98.8	0.0	85
		0.9	99.4	0.9	3334	95.8	7.7	23	99.7	1.5	393	98.8	0.0	83	99.5	0.31	1892	99.3	1.9	859	98.8	0.0	85

Abbreviations: %Con = percentage concordance; %Miss = percentage missing; ESS= effective sample size; 1KGP Phase 1= 1000 Genomes Phase 1 Version 3 reference panel; 1000G Phase 3 = 1000 Genomes Phase 3 version 5 reference panel; HRC = Haplotype Reference Consortium Release 1 reference panel.
*The Genotype Hard Call Probability Threshold is the threshold above which conversion of genotype call from dosage to hard call allele count format was carried out. Values below the threshold are called as missing.

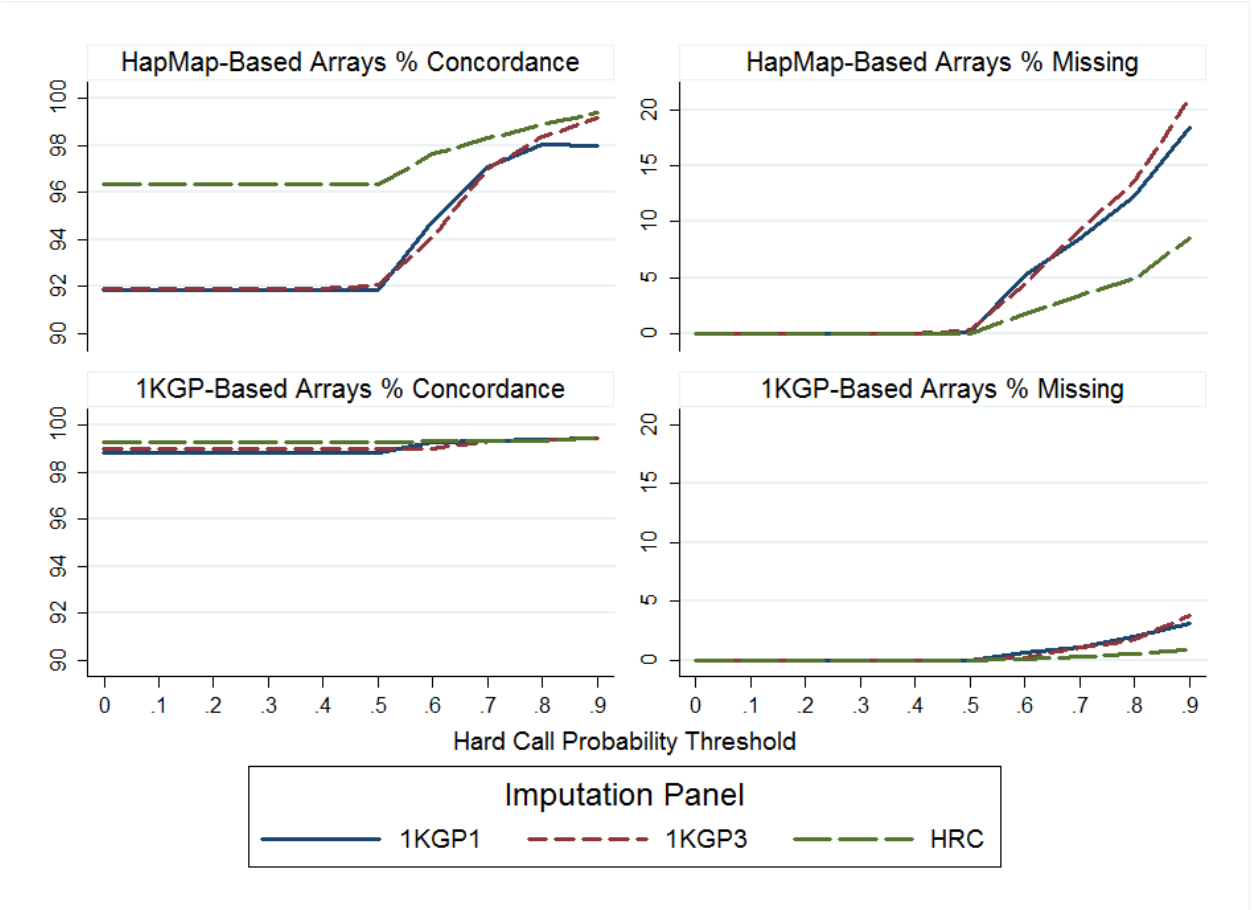

Figure 1. Comparison of measured and imputed *APOE* genotypes showing concordance and missingness.
(See table 1 legend for abbreviations).