Hardy–Weinberg analysis of a large set of published association studies reveals genotyping error and a deficit of heterozygotes across multiple loci

Srijan Sen[1*] *and Margit Burmeister*[1,2]

[1]Molecular & Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI 48109, USA
[2]Departments of Psychiatry and Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA
*Correspondence to: Tel: +1 734 395 8319; E-mail: srijan.sen@yale.edu

Date received (in revised form): 7th April, 2008

Abstract

In genetic association studies, deviation from Hardy–Weinberg equilibrium (HWD) can be due to recent admixture or selection at a locus, but is most commonly due to genotyping errors. In addition to its utility for identifying potential genotyping errors in individual studies, here we report that HWD can be useful in detecting the presence, magnitude and direction of genotyping error across multiple studies. If there is a consistent genotyping error at a given locus, larger studies, in general, will show more evidence for HWD than small studies. As a result, for loci prone to genotyping errors, there will be a correlation between HWD and the study sample size. By contrast, in the absence of consistent genotyping errors, there will be a chance distribution of *p*-values among studies without correlation with sample size. We calculated the evidence for HWD at 17 separate polymorphic loci investigated in 325 published genetic association studies. In the full set of studies, there was a significant correlation between HWD and locus-standardised sample size ($p = 0.001$). For 14/17 of the individual loci, there was a positive correlation between extent of HWD and sample size, with the evidence for two loci (*5-HTTLPR* and *CTSD*) rising to the level of statistical significance. Among single nucleotide polymorphisms (SNPs), 15/23 studies that deviated significantly from Hardy–Weinberg equilibrium (HWE) did so because of a deficit of heterozygotes. The inbreeding coefficient (F(is)) is a measure of the degree and direction of deviation from HWE. Among studies investigating SNPs, there was a significant correlation between F(is) and HWD ($R = 0.191$; $p = 0.002$), indicating that the greater the deviation from HWE, the greater the deficit of heterozygotes. By contrast, for repeat variants, only one in five studies that deviated significantly from HWE showed a deficit of heterozygotes and there was no significant correlation between F(is) and HWD. These results indicate the presence of HWD across multiple loci, with the magnitude of the deviation varying substantially from locus to locus. For SNPs, HWD tends to be due to a deficit of heterozygotes, indicating that allelic dropout may be the most prevalent genotyping error.

Keywords: *meta-analysis*, *polymorphism*, *variant*, *deviation*

Introduction

Genotyping errors are an important and increasingly recognised problem in modern genetics.[1] Traditional family-based genetic studies allow for straightforward identification of genotyping errors through a familial Mendelian inheritance check. Over the past decade, however, there has been increasing interest in case-control association

studies, a type of study in which investigators generally compare a group of subjects having a particular disease with another group not having the disease, to identify a genotypic difference between the groups. Unfortunately, these association studies do not allow for simple inheritance checks to identify errors and, as a result, we have limited insight into the prevalence and nature of genotyping errors in published association studies.

Hardy–Weinberg law states that if conditions of population equilibrium are met (random mating and negligible mutation, migration, stratification, genetic drift and selection), then genotype frequencies should fit a predictable binomial distribution calculable from the allele frequencies. Significant deviation from the predicted distribution has been used as a marker for genotyping error.[2] Previous work has estimated that the control sample genotype distribution violates Hardy–Weinberg equilibrium (HWE) in approximately 10 per cent of published association studies.[3–5] Furthermore, exclusion of studies that violate HWE alters the results of a substantial fraction of gene association meta-analyses.[6]

The inbreeding coefficient (F(is)) can be used as a measure of the degree and direction of deviation from HWE (HWD). Positive F(is) values indicate an excess of homozygotes and negative F(is) values indicate a deficit of homozygotes. Salanti and colleagues[4] found that with a moderate level of HWD (F(is) = 0.10), only 7 per cent of association studies had at least 80 per cent power to find significant evidence for violation of HWE. Because of this low level of power, focusing on statistically significant violation of HWE in individual association studies substantially limits the insight that we can gain into potential genotyping errors from HWE analysis.[7] A complementary approach that bypasses the problem of limited power in individual studies is the analysis of HWD patterns across a set of studies. As originally demonstrated by Weir,[8] if a locus is prone to genotyping error, the evidence for HWD will increase with increasing sample size. By contrast, if there is no substantial genotyping error, or if the error is random, there will be no relationship between HWD and sample size. By examining a set of studies at a given locus, we can learn about the level of genotyping error present at that locus. Furthermore, by looking at the evidence across multiple loci, we can gain insight into the level and nature of genotyping error in association studies in general.

Here, we investigate: (1) the relationship between sample size and HWD across well-studied loci, and (2) the direction of deviation in a set of association studies compiled from previous meta-analyses.

Materials and methods

Studies

Genetic loci for analysis were identified through published meta-analyses. Meta-analyses were identified through PubMed at the National Library of Medicine, limiting the search to meta-analyses published between 2001 and 2005 and using the search terms: (1) association genetic; (2) association polymorphism; (3) association variant. These results were supplemented by a database of meta-analyses compiled by Ioannidis and colleagues.[9,10] Loci were subsequently chosen using the criteria: (1) biallelic markers; (2) at least ten independent studies; and (3) sample size data for all three genotype groups included in the publication. For each included study, we recorded the control group sample size for the three genotype groups (Supplementary Table 1).

Analyses

The most straightforward way to assess HWD in a set of studies investigating a given locus is to pool the genotype cell counts from each of the relevant studies and assess HWD among the three pooled genotype groups. All of these studies investigated population samples with different ethnicities, however, and consequently different allele frequencies. As a result, simply combining data from different studies would find substantial HWD due to lack of heterozygotes, even in the absence of genotyping error.

We took an alternative approach to assessing HWD among a set of studies investigating a given

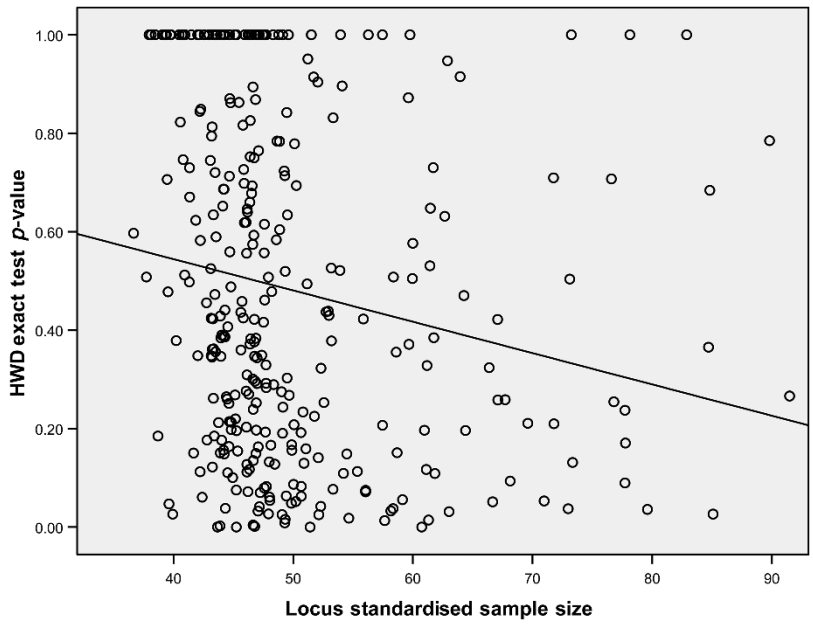

Figure 1. Hardy–Weinberg disequilibrium (HWD) *p*-value vs sample size across 325 studies.

Table 1. Relationship between sample size and Hardy–Weinberg exact test *p*-values for individual loci

Variant	Variant	No. of studies	Correlation (*p*-value)
PON192	SNP	39	−0.120 (0.467)
GPIIIa	SNP	33	−0.050 (0.783)
5-HTTLPR	Repeat	31	−0.444 (0.014)
L-myc-ECOR1	Repeat	28	−0.296 (0.126)
MTHFR677	SNP	23	−0.201 (0.358)
VDR	SNP	17	−0.140 (0.593)
CTSD	SNP	16	−0.582 (0.018)
DRD2	SNP	21	−0.191 (0.407)
Neurod1	SNP	14	−0.433 (0.122)
TPH	SNP	13	0.297 (0.324)
COLIA1	SNP	13	0.188 (0.538)
ADD1	SNP	12	−0.275 (0.387)
SRD5A2	SNP	12	−0.326 (0.301)
BSM1	SNP	11	−0.188 (0.503)
IL-1	SNP	11	−0.520 (0.101)
CYP17	SNP	10	−0.175 (0.628)

locus. For each locus, we determined the correlation between the HWD exact test *p*-value of each study and study sample size. The stronger the correlation, the stronger the evidence for HWD at that locus. Given that many included studies had small homozygote minor allele cell counts (fewer than five subjects), and that the chi square test is an unreliable test of HWD in the presence of small cell counts, an exact test was used to determine the strength of evidence for HWD.[11]

In addition to investigating the correlation between HWD and sample size among studies investigating each individual locus, we also wanted to explore the strength and significance of this correlation across all studies, regardless of locus. A straightforward assessment of correlation between sample size and HWD, however, would be confounded by statistical artefact. Specifically, the mean sample size varies substantially across loci. Because the level of HWD varies substantially across loci (as demonstrated by our initial analyses), a correlation between sample size and HWD *p*-value among the set of all studies could merely represent that loci with larger mean sample sizes have greater HWD. In order to control for this potential confound, we calculated a standardised sample size for each study,

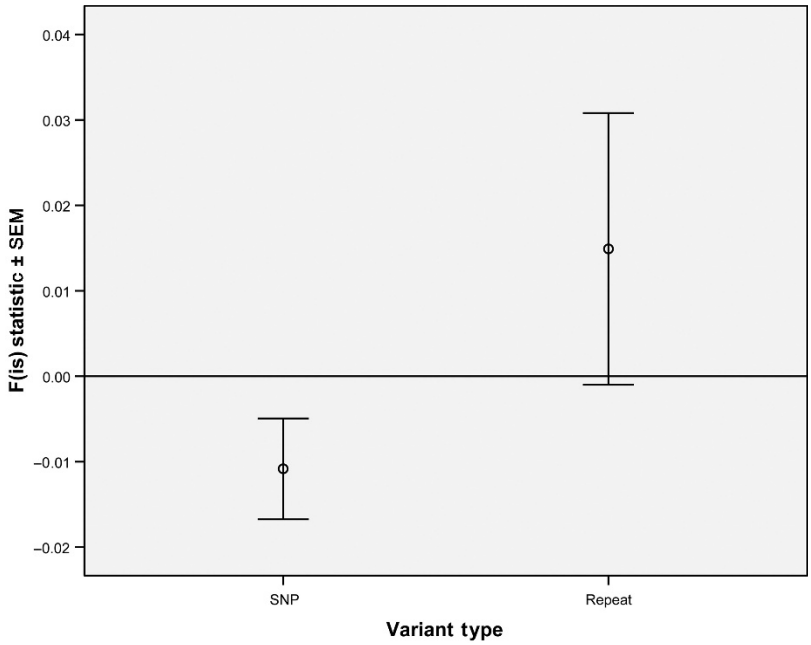

Figure 2. Mean F(is) statistic stratified by variant type.

such that each locus had a mean sample size = 50 and sample size standard deviation = 10. Subsequently, we calculated the strength and significance of the correlation between this locus-standardised sample size and HWD *p*-value for the set of all studies. The raw sample size for each study was converted to a T-score so that each locus had an overall mean standardised sample size of 50 ± 10. Subsequently, the correlation between standardised sample size and exact test *p*-value was calculated for the set of all studies.

Inbreeding coefficient was calculated using the following formula:

$$\mathrm{F(is)} = \mathrm{P(AA)/P(A)} + \mathrm{P(aa)/P(a)} - 1$$

where P = frequency; A = major allele; a = minor allele; AA = homozygous major allele; aa = homozygous minor allele. All analyses were carried out in SPSS 12.0 (SPSS Inc., Chicago, IL, USA).

Results

In total, 325 studies, investigating 17 loci, fit the criteria for analysis. Twenty-eight studies (9 per cent) showed significant HWD. This proportion is in line with the results of previous studies.[3–5] The number of studies per locus ranged from ten (*CYP1*) to 39 (*PON1 Q192R*). The average sample size per locus ranged from 71 (*DRD2*) to 1,020 (*ADD1*) (Figure 1).

Among individual loci, 14/17 variants showed a negative correlation between sample size and HWD *p*-value, indicating that the majority of studied variants show evidence of consistent genotyping error. Overall, the correlations ranged from $R = 0.29$ (*TPH*) to $R = -0.59$ (*CTSD*) and was significant for two loci (*CTSD* and *5-HTTLPR*) (Table 1). Among the set of all 325 studies, 23 studies had a homozygote minor allele cell count = 0. The strength and significance of correlations were not substantially changed with the exclusion of these studies (data not shown).

The 325 studies investigated 15 single nucleotide polymorphism (SNP) loci (267 studies) and two repeat polymorphism loci (58 studies). The percentage of individual studies that significantly deviated from HWE was the same (9 per cent) for both the SNP and repeat polymorphism categories. Similarly, the standardised sample size–HWD correlation was statistically significant for both SNP

($p = 0.018$) and repeat polymorphism ($p = 0.004$) groups. Of the 28 studies that showed significant deviation from HWE, 23 studies were SNP studies and five were repeat polymorphism studies. Fifteen out of 23 HWE-violating SNP studies showed a deficit of heterozygotes, while only one in five HWE-violating repeat polymorphism studies showed a deficit of heterozygotes. In addition, for SNP studies, there was a significant correlation between F(is) and HWD *p*-value ($R = 0.190$; $p = 0.002$), while repeat polymorphisms showed no evidence of correlation ($R = 0.03$). In the set of all 325 studies, there was a significant correlation between standardised sample size and HWD ($R = 0.18$; $p = 0.001$) (Figure 2).

To gain insight into the reliability of the results found among controls, and to help to differentiate between selection and genotyping error as the primary cause of HWD, we investigated the correlation between F(is) among cases (F(cases)) and controls (F(controls)) for each individual study. If the HWD among control subjects is due to selection, then we would expect the genotype that is deficient among controls to be over-represented among cases, and thus F(is) among control and case studies would show a *negative* correlation. By contrast, if the HWD among control subjects is due to genotyping error, then we would expect the genotype that is deficient among controls also to be deficient among cases, and thus the inbreeding coefficients would show a *positive* correlation. Lastly, if the HWD among controls were due purely to chance, then we would expect no correlation whatsoever between F(is) statistics.

Looking across 12 loci and 221 studies for which we had data for both cases and controls, we found a significant *positive* correlation between F (controls) and F (cases) ($r = 0.174$; $p = 0.01$). Further, the correlation was in the positive direction for 11/12 loci. These findings indicate that for any given study, the direction and magnitude of HWD among cases is similar to the direction of magnitude of HWD among controls. This result is consistent with genotyping error rather than selection as the primary source of HWD, and provides further evidence that these findings are not due purely to chance.

Discussion

The primary finding of this analysis was the identification of HWD across a large subset of published association studies investigating both SNP and repeat variants. Although deviation was present at most loci, the degree of deviation varied substantially across loci. At least among SNP studies, the predominant cause of this deviation was a deficit of heterozygotes.

In addition to genotyping error, other factors can contribute to HWD. For example, strong selection against a specific genotype can skew the genotypic distribution of a population. In fact, HWD *among cases* has been used as a test for genotype–phenotype association,[12,13] and Wittke-Thompson and colleagues[14] have demonstrated a pattern of expected deviation among cases and, under some conditions, controls for various disease models. Our finding that the HWD among cases has a strong tendency to be in the same direction as the deviation found among controls is contrary to the expected result under the selection model, however.

Population stratification is another factor that can contribute to HWD. To eliminate the possibility of ethnic differences *between* studies causing stratification and HWD in our study, we did not pool the three genotype counts for all studies investigating a given locus and calculate a HWD *p*-value from this pooled sample. Instead, for each locus, we determined the correlation between the HWD exact test *p*-value and study sample size. Thus, any effect of stratification in our study is not due to allele frequency differences *between* studies investigating the same locus. Although population stratification *within* individual studies may contribute to HWD in our study, there are multiple considerations that are likely to mitigate its effect. First, most studies included in our analysis utilise samples that are ethnically homogeneous. Secondly, a significant proportion of the studies formally tested and rejected the presence of population stratification in

their sample. Thirdly, the consistent direction of deviation across studies and the different patterns of deviation found between SNP and repeat variants are more consistent with genotyping error than stratification as a primary cause of HWD. We cannot however, definitively exclude stratification as a contributing cause of HWD among these studies.

Previous studies investigating the nature and consequences of genotyping error based on simulations or experimental samples specifically designed to assess genotyping error have proposed allelic dropout as one of the most frequent causes of genotypic error.[2,15,16] Intuitively, it is clear that heterozygotes, which get half a dose of each allele compared with homozygotes, may be more often missed or misclassified. In fact, even in the most sophisticated high-throughput algorithms, heterozygotes have a lower call rate than homozygotes.[17] Our investigation of a large set of published studies is consistent with this prediction. Further, our findings are consistent with the hypothesis that genotyping error is not stochastic, but more common at certain loci.[18–21] These findings raise concerns about the level and widespread nature of genotyping errors in genetic association studies and the conclusions drawn from those studies. In light of this finding, the approach employed here could be useful to identify loci most prone to error. For example, Yonan and colleagues[22] recently used HWD to identify genotyping errors at the 5-hydroxytryptamine transporter 5-HTTLPR variant and developed an alternate assay less prone to error.

We propose that future genetic association meta-analyses examine the correlation between sample size and HWE to determine the level of genotyping error among included studies. Further, we believe that the method and points that this analysis highlight can be of utility to investigators performing individual association studies. First, this result should caution investigators against dismissing the possibility of genotyping error merely because their sample does not show significant deviation from HWE. Instead, investigators should further examine the magnitude and direction of deviation. For instance, a large F(is) statistic in the same direction among cases and controls raises the concern for genotyping error, and should prompt investigators to perform genotyping quality checks.

Acknowledgments

The authors are very grateful to Pratima Naik for her contribution to this study and to Scott Stoltenberg and Laura Scott for advice and helpful discussion. They also thank the reviewers for their thorough and helpful comments, which helped them significantly to improve the manuscript.

References

1. Pompanon, F., Bonin, A., Bellemain, E. and Taberlet, P. (2005), 'Genotyping errors: Causes, consequences and solutions', *Nat. Rev. Genet*. Vol. 6, pp. 847–859.
2. Hosking, L., Lumsden, S., Lewis, K. *et al*. (2004), 'Detection of genotyping errors by Hardy–Weinberg equilibrium testing', *Eur. J. Hum. Genet*. Vol. 12, pp. 395–399.
3. Xu, J., Turner, A., Little, J. *et al*. (2002), 'Positive results in association studies are associated with departure from Hardy–Weinberg equilibrium: Hint for genotyping error?', *Hum. Genet*. Vol. 111, pp. 573–574.
4. Salanti, G., Amountza, G., Ntzani, E.E. and Ioannidis, J.P. (2005), 'Hardy–Weinberg equilibrium in genetic association studies: An empirical evaluation of reporting, deviations, and power', *Eur. J. Hum. Genet*. Vol. 13, pp. 840–848.
5. Bardoczy, Z., Gyorffy, B., Kocsis, I. and Vasarhelyi, B. (2004), 'Re-calculated Hardy–Weinberg values in papers published in *Atherosclerosis* between 1995 and 2003', *Atherosclerosis* Vol. 173, pp. 141–143.
6. Trikalinos, T.A., Salanti, G., Khoury, M.J. and Ioannidis, J.P. (2006), 'Impact of violations and deviations in Hardy–Weinberg equilibrium on postulated gene-disease associations', *Am. J. Epidemiol*. Vol. 163, pp. 300–309.
7. Cox, D.G. and Kraft, P. (2006), 'Quantification of the power of Hardy–Weinberg equilibrium testing to detect genotyping error', *Hum. Hered*. Vol. 61, pp. 10–14.
8. Weir, B.S. (1996), *Genetic Data Analysis II*. Sinauer, Sunderland.
9. Ioannidis, J.P., Trikalinos, T.A., Ntzani, E.E. and Contopoulos-Ioannidis, D.G. (2003), 'Genetic associations in large versus small studies: An empirical assessment', *Lancet* Vol. 361, pp. 567–571.
10. http://image.thelancet.com/extras/02art8007webappendix.pdf.
11. Wiggintonn, J.E., Cutler, D.J. and Abecasis, G.R. (2005), 'A note on exact tests of Hardy–Weinberg equilibrium', *Am. J. Hum. Genet*. Vol. 76, pp. 887–893.
12. Luo, X., Kranzler, H.R., Zuo, L. *et al*. (2006), 'ADH4 gene variation is associated with alcohol dependence and drug dependence in European Americans: Results from HWD tests and case-control association studies', *Neuropsychopharmacology* Vol. 31, pp. 1085–1095.
13. Feder, J.N., Gnirke, A., Thomas, W. *et al*. (1996), 'A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis', *Nat. Genet*. Vol. 13, pp. 399–408.
14. Wittke-Thompson, J.K., Pluzhnikov, A. and Cox, N.J. (2005), 'Rational inferences about departures from Hardy–Weinberg equilibrium', *Am. J. Hum. Genet*. Vol. 76, pp. 967–986.
15. Taberlet, P., Griffin, S., Goossens, B. *et al*. (1996), 'Reliable genotyping of samples with very low DNA quantities using PCR', *Nucleic Acids Res*. Vol. 24, pp. 3189–3194.
16. Gagneux, P., Boesch, C. and Woodruff, D.S. (1997), 'Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair', *Mol. Ecol*. Vol. 6, pp. 861–868.
17. Carvalho, B., Bengtsson, H., Speed, T.P. and Irizarry, R.A. (2007), 'Exploration, normalization, and genotype calls of high density oligonucleotide SNP array data', *Biostatistics* Vol. 8, pp. 485–499.

18. Mitchell, A.A., Cutler, D.J. and Chakravarti, A. (2003), 'Undetected genotyping errors cause apparent overtransmission of common alleles in the transmission/disequilibrium test', *Am. J. Hum. Genet*. Vol. 72, pp. 598–610.
19. Jeffrey, G.P. and Adams, P.C. (2000), 'Pitfalls in the genetic diagnosis of hereditary hemochromatosis', *Genet. Test*. Vol. 4, pp. 143–146.
20. Creel, S., Spong, G., Sands, J.L. *et al.* (2003), 'Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes', *Mol. Ecol*. Vol. 12, pp. 2003–2009.
21. Constable, J.L., Ashley, M.V., Goodall, J. and Pusey, A.E. (2001), 'Noninvasive paternity assignment in Gombe chimpanzees', *Mol. Ecol.* Vol. 10, pp. 1279–1300.
22. Yonan, A.L., Palmer, A.A. and Gilliam, T.C. (2006), 'Hardy–Weinberg disequilibrium identified genotyping error of the serotonin transporter (SLC6A4) promoter polymorphism', *Psychiatr. Genet*. Vol. 16, pp. 31–34.

Supplementary Table. Included association studies stratified by locus

Study	locus	std N	a1/a1	a1/a2	a2/a2	N	*p*-value
Brummett	5-HTTLPR	47.62162	33	91	78	202	0.4612
Comings	5-HTTLPR	47.72973	58	95	51	204	0.3294
Du	5-HTTLPR	46.75676	40	86	60	186	0.3763
Ebstein	5-HTTLPR	43.24324	32	66	23	121	0.3611
Flory	5-HTTLPR	48.86486	37	112	76	225	0.7835
Greenberg	5-HTTLPR	58.16216	66	217	114	397	0.0328
Gusatavsson	5-HTTLPR	46.16216	35	83	57	175	0.6461
Gusatavsson	5-HTTLPR	43.45946	22	66	37	125	0.4725
Hamer	5-HTTLPR	70.97297	108	336	190	634	0.053
Herbst	5-HTTLPR	59.67568	79	198	148	425	0.3712
Hu	5-HTTLPR	77.72973	135	390	234	759	0.2373
Jorm	5-HTTLPR	77.72973	155	350	254	759	0.0896
Katsuragi	5-HTTLPR	42.16216	66	31	4	101	1
Kumakiri-TCI	5-HTTLPR	44.48649	85	48	11	144	0.26
Lang	5-HTTLPR	49.02703	41	102	85	228	0.2748
Lesch	5-HTTLPR	52.05405	52	141	91	284	0.9039
Lesch	5-HTTLPR	48.64865	43	106	72	221	0.7841
Mazzanti	5-HTTLPR	48.32432	41	106	68	215	1
Melke	5-HTTLPR	46.97297	35	84	71	190	0.2915
Murakami	5-HTTLPR	46.91892	124	55	10	189	0.2523
Nakamura	5-HTTLPR	46.75676	128	55	3	186	0.4221
Osher-TPQ	5-HTTLPR	44.7027	39	73	36	148	0.8703
Ricketts	5-HTTLPR	38.7027	10	14	13	37	0.185
Samachowiec	5-HTTLPR	43.51351	18	67	41	126	0.356
Schmidt	5-HTTLPR	39.78378	12	29	16	57	1
Sen	5-HTTLPR	59.13514	83	183	149	415	0.0557

Continued

Supplementary Table. Continued

Study	locus	std N	a1/a1	a1/a2	a2/a2	N	*p*-value
Stoltenberg	5-HTTLPR	41.35135	17	45	24	86	0.6704
Strobel	5-HTTLPR	43.35135	22	67	34	123	0.3619
Tsai	5-HTTLPR	47.08108	100	71	21	192	0.1629
Umekage	5-HTTLPR	49.89189	161	70	13	244	0.156
O'Donnell	ACE DI	54.48314	492	845	313	1650	0.1486
O'Donnell	ACE DI	53.34439	437	719	288	1444	0.8315
Agerholm-Larsen	ACE DI	89.81205	2113	4006	1922	8041	0.7849
Barley	ACE DI	46.52294	55	109	46	210	0.678
Benetos	ACE DI	46.06965	47	56	25	128	0.2764
Berge	ACE DI	46.13599	34	77	29	140	0.3092
Busjahn	ACE DI	46.13046	33	79	27	139	0.1272
Cambien	ACE DI	49.41404	200	390	143	733	0.0632
Castellano	ACE DI	46.40685	76	90	23	189	0.7523
Celermajer	ACE DI	46.37922	49	89	46	184	0.6599
Friedl	ACE DI	45.72692	16	37	13	66	0.4583
Kauma	ACE DI	48.20896	148	264	103	515	0.4783
Kiema	ACE DI	46.64456	75	115	42	232	0.8941
Kiema	ACE DI	46.65561	54	127	53	234	0.239
Ludwig	ACE DI	47.58983	117	206	80	403	0.6152
Mattu	ACE DI	52.1393	442	556	228	1226	0.0251
Puija	ACE DI	46.09176	46	70	16	132	0.203
Rigat	ACE DI	45.80431	29	37	14	80	0.8164
Tiret	ACE DI	46.44555	60	103	33	196	0.3825
Busch	ADD1	48.02101	405	76	0	481	0.0608
Clark	ADD1	46.77722	162	80	14	256	0.347
Ju	ADD1	49.49696	166	357	225	748	0.3028
Manunta	ADD1	45.95909	80	26	2	108	1
Morrison	ADD1	56.05307	1227	643	64	1934	0.0747
Mulatero	ADD1	46.28524	117	43	7	167	0.2699
Narita	ADD1	46.88778	56	150	70	276	0.1494
Nicod	ADD1	46.79934	167	83	10	260	1

Continued

Supplementary Table. Continued

Study	locus	std N	a1/a1	a1/a2	a2/a2	N	*p*-value
Persu	ADD1	46.41791	121	63	7	191	0.8258
Ranade	ADD1	51.2272	296	530	235	1061	0.951
Shioji	ADD1	67.08184	241	560	305	1106	0.4218
Yamagishi	ADD1	60.96739	599	1365	859	2823	0.1967
berg	bsm1	41.67598	12	19	18	49	0.1504
boschitsch	bsm1	48.04469	36	67	60	163	0.0539
garnero	bsm1	53.91061	38	134	96	268	0.5213
gennari	bsm1	61.84358	71	219	120	410	0.1087
gomez	bsm1	47.93296	27	72	62	161	0.5075
hansen	bsm1	50.11173	46	98	56	200	0.7787
jorgensen	bsm1	69.60894	77	276	196	549	0.2109
kiel	bsm1	45.2514	22	17	74	113	2.2E-10
kroger	bsm1	40.22346	2	14	7	23	0.3787
langdahl	bsm1	43.40782	25	34	21	80	0.1848
marc	bsm1	44.63687	19	59	24	102	0.1634
mcclure	bsm1	44.69274	8	43	52	103	1
melhus	bsm1	43.18436	7	35	34	76	0.7943
riggs	bsm1	44.02235	15	36	40	91	0.1765
vandevyer	bsm1	71.78771	107	306	175	588	0.2098
aerssens	COLIA1	50.90116	151	73	15	239	0.1295
alvarez	COLIA1	44.65116	21	3	0	24	1
de vernejoul	COLIA1	47.93605	85	51	1	137	0.0267
efstathiodou	COLIA1	47.18023	73	29	9	111	0.0413
heegaard	COLIA1	47.18023	82	27	2	111	1
hustmyer	COLIA1	46.22093	58	16	4	78	0.0719
keen	COLIA1	47.73256	85	40	5	130	1
langdahl	COLIA1	48.13953	94	48	2	144	0.1664
liden	COLIA1	45.90116	44	20	3	67	0.6981
mcguigan	COLIA1	46.51163	70	17	1	88	1
roux	COLIA1	47.06395	81	24	2	107	1
uitterlinden	COLIA1	82.87791	905	392	42	1339	1

Continued

Supplementary Table. Continued

Study	locus	std N	a1/a1	a1/a2	a2/a2	N	*p*-value
weichetova	COLIA1	47.61628	94	30	2	126	1
bagnoli	CTSD	42.01754	1	26	99	126	1
bertram	CTSD	46.92982	1	29	152	182	1
bhojak	CTSD	58.68421	0	56	260	316	0.151
crawford	CTSD	41.49123	0	20	100	120	1
crawford	CTSD	40.78947	2	28	82	112	1
emahazion	CTSD	44.03509	3	27	119	149	0.3899
ingegni	CTSD	41.49123	1	21	98	120	1
mateo	CTSD	61.31579	8	54	284	346	0.0143
matsui	CTSD	72.98246	1	7	471	479	0.0372
mcilroy	CTSD	47.36842	1	16	170	187	0.3491
menzer	CTSD	57.45614	1	33	268	302	1
papassotiropoulos	CTSD	61.75439	0	47	304	351	0.3847
papassotiropoulos	CTSD	47.10526	0	18	166	184	1
prince	CTSD	46.22807	0	22	152	174	1
styczynska	CTSD	39.73684	0	9	91	100	1
chang	CYP17	45.82569	26	79	77	182	0.4248
gsur	CYP17	43.25688	12	67	47	126	0.1219
habuchi	CYP17	52.75229	69	157	107	333	0.4371
haiman	CYP17	73.34862	127	350	305	782	0.1312
kittles	CYP17	42.56881	10	46	55	111	1
latil	CYP17	44.63303	24	84	48	156	0.2511
lunn	CYP17	44.77064	18	73	68	159	0.8621
stanford	CYP17	61.46789	79	256	188	523	0.6477
wadelius	CYP17	44.81651	26	88	46	160	0.1979
yamada	CYP17	46.65138	29	120	51	200	0.004
amadeo	drd2	43.48837	0	7	36	43	1
Anghelescu	drd2	56.27907	3	32	63	98	1
Bau	drd2	60	6	36	72	114	0.5764
blum	drd2	39.06977	0	4	20	24	1
blum	drd2	40.69767	0	6	25	31	1

Continued

Supplementary Table. Continued

Study	locus	std N	a1/a1	a1/a2	a2/a2	N	*p*-value
bolos	drd2	63.02326	8	30	89	127	0.0314
comings	drd2	58.60465	0	24	84	108	0.3553
cook	drd2	38.13953	0	6	14	20	1
geijer	drd2	52.32558	5	24	52	81	0.3226
gelernter	drd2	49.30233	3	21	44	68	0.7138
goldman	drd2	41.86047	2	11	23	36	0.6232
heinz	drd2	59.76744	4	35	74	113	1
Hietala	drd2	45.11628	0	11	39	50	1
lawford	drd2	44.18605	3	11	32	46	0.1562
neiswanger	drd2	40.46512	0	4	26	30	1
noble	drd2	46.97674	3	14	41	58	0.3437
Ovchiunikov	drd2	51.16279	4	23	49	76	0.494
parsian	drd2	39.30233	0	3	22	25	1
Pastorelli	drd2	48.37209	2	13	49	64	0.2895
Samochoweic	drd2	78.13953	5	51	136	192	1
suarez	drd2	53.95349	2	23	63	88	1
abbate	gpIIIa	43.2963	3	19	51	73	0.4229
aleksic	gpIIIa	60.74074	0	141	403	544	0.000039
anderson	gpIIIa	50.81481	9	65	202	276	0.2337
anderson	gpIIIa	46.88889	6	42	122	170	0.3835
ardissino	gpIIIa	48	4	33	163	200	0.1324
boncler	gpIIIa	43.55556	0	19	61	80	0.5896
bottiger	gpIIIa	53.18519	9	84	247	340	0.5261
carter	gpIIIa	44.81481	0	28	86	114	0.2131
carter	gpIIIa	48.59259	3	57	156	216	0.5836
carter	gpIIIa	43.92593	2	24	64	90	1
corral	gpIIIa	44.33333	0	35	66	101	0.038
durante-mangoni	gpIIIa	43.22222	0	19	52	71	0.3451
garcia	gpIIIa	44.2963	1	12	87	100	0.3864
gardemann	gpIIIa	84.7037	31	297	863	1191	0.3654
grand maison	gpIIIa	44.2963	1	23	76	100	1

Continued

Supplementary Table. Continued

Study	locus	std N	a1/a1	a1/a2	a2/a2	N	*p*-value
hermann	gpIIIa	47.11111	4	43	129	176	0.7646
hermann	gpIIIa	59.96296	10	143	370	523	0.5047
hooper	gpIIIa	47.44444	2	39	144	185	1
joven	gpIIIa	49.85185	3	81	166	250	0.0483
kekomaki	gpIIIa	42.22222	2	7	35	44	0.1123
kekomaki	gpIIIa	43.62963	1	17	64	82	1
laule	gpIIIa	76.59259	20	254	698	972	0.7073
mamotte	gpIIIa	61.7037	12	136	422	570	0.7302
marian	gpIIIa	46.66667	7	38	119	164	0.135
moshfegh	gpIIIa	43.88889	6	14	69	89	0.0023
osborn	gpIIIa	46.77778	8	27	132	167	0.0015
pastinen	gpIIIa	46.18519	2	26	123	151	0.6399
ridker	gpIIIa	66.66667	22	164	518	704	0.0513
samani	gpIIIa	49.2963	5	97	133	235	0.0086
scaglione	gpIIIa	44.22222	1	27	70	98	0.6863
senti	gpIIIa	45.62963	3	28	105	136	0.4363
weiss	gpIIIa	43.11111	1	12	55	68	0.525
zotz	gpIIIa	43.96296	0	23	68	91	0.3467
Combarros	IL-1	52.10145	195	104	7	306	0.1408
Du	IL-1	43.76812	126	62	3	191	0.2122
Green	IL-1	66.37681	221	217	65	503	0.3238
Grimaldi	IL-1	54.2029	142	163	30	335	0.109
Hedley	IL-1	55.36232	153	168	30	351	0.113
Ki	IL-1	36.66667	72	21	0	93	0.5969
Minster	IL-1	46.73913	115	99	18	232	0.75
Nicoll	IL-1	42.02899	82	74	11	167	0.3481
Pirskanen	IL-1	67.10145	248	209	56	513	0.2582
Rebeck	IL-1	43.47826	97	74	16	187	0.7202
Tsai	IL-1	42.24638	147	22	1	170	0.5822
chenevix-Trench	LmycECOR1	57.46667	46	72	43	161	0.2068
chernitsa	LmycECOR1	46.26667	18	38	21	77	1

Continued

Supplementary Table. Continued

Study	locus	std N	a1/a1	a1/a2	a2/a2	N	*p*-value
crossen	LmycECOR1	49.33333	43	43	14	100	0.5194
dlugosz	LmycECOR1	44.66667	11	38	16	65	0.2145
dolcetti	LmycECOR1	46.4	24	35	19	78	0.3718
ejarque	LmycECOR1	50.66667	40	45	25	110	0.0825
fernandez	LmycECOR1	49.46667	30	49	22	101	0.842
ge	LmycECOR1	39.46667	6	12	8	26	0.7061
hseih	LmycECOR1	47.73333	22	39	27	88	0.2921
isbir	LmycECOR1	47.06667	39	29	15	83	0.0323
isbir	LmycECOR1	42.8	23	26	2	51	0.1768
ishizaki	LmycECOR1	49.33333	17	63	20	100	0.0157
kato	LmycECOR1	49.06667	17	61	20	98	0.0254
kondratieva	LmycECOR1	49.6	28	52	22	102	1
kuminoto	LmycECOR1	68.13333	59	134	48	241	0.0934
murakami	LmycECOR1	79.6	69	183	75	327	0.0358
saranath	LmycECOR1	49.46667	30	49	22	101	0.842
shibuta	LmycECOR1	50.26667	34	55	18	107	0.6938
shibuta	LmycECOR1	50.26667	34	55	18	107	0.6938
shih	LmycECOR1	53.33333	43	54	33	130	0.0767
taylor	LmycECOR1	46.13333	22	31	23	76	0.1118
tefre	LmycECOR1	53.2	35	59	35	129	0.3782
togo	LmycECOR1	76.8	85	143	78	306	0.2544
weston	LmycECOR1	43.33333	10	22	23	55	0.2616
weston	LmycECOR1	40.8	11	17	8	36	0.7464
weston	LmycECOR1	37.73333	2	4	7	13	0.5079
yaylim	LmycECOR1	40.93333	14	16	7	37	0.5121
young	LmycECOR1	42.4	16	29	3	48	0.0606
Adams	MTHFR C677T	47.57246	29	97	96	222	0.557
brugada	MTHFR C677T	45.14493	12	73	70	155	0.2683
Brulhart	MTHFR C677T	56.05072	73	195	188	456	0.0715
Christensen	MTHFR C677T	43.91304	13	61	47	121	0.4287
de Franchis	MTHFR C677T	48.87681	39	129	90	258	0.6041

Continued

Supplementary Table. Continued

Study	locus	std N	a1/a1	a1/a2	a2/a2	N	*p*-value
Deloughery	MTHFR C677T	61.12319	94	262	240	596	0.117
Gallagher	MTHFR C677T	43.33333	7	45	53	105	0.6343
Izumi	MTHFR C677T	46.81159	25	102	74	201	0.2965
Kluijtmans	MTHFR C677T	43.55072	6	42	63	111	1
Kluijtmans	MTHFR C677T	84.81884	106	527	617	1250	0.6841
Ma	MTHFR C677T	50.03623	39	116	135	290	0.0868
malinow	MTHFR C677T	43.22464	8	45	49	102	0.8129
markus	MTHFR C677T	45.36232	22	63	76	161	0.1545
morita	MTHFR C677T	67.71739	79	361	338	778	0.2587
Narang	MTHFR C677T	41.34058	5	19	26	50	0.7298
salden	MTHFR C677T	45.47101	18	75	71	164	0.8626
Schmitz	MTHFR C677T	46.34058	27	90	71	188	1
Schwartz	MTHFR C677T	51.77536	43	141	154	338	0.2251
tosetto	MTHFR C677T	44.23913	17	71	42	130	0.1486
van bockxmeer	MTHFR C677T	44.71014	15	58	70	143	0.5591
Verhoef	MTHFR C677T	43.15217	7	48	45	100	0.3479
verhoef	MTHFR C677T	57.64493	72	200	228	500	0.013
Wilcken	MTHFR C677T	47.68116	24	113	88	225	0.1929
Awata	Neurod1	71.75824	1	55	327	383	0.7094
Cinek	Neurod1	61.42857	42	130	117	289	0.5308
Dupont	Neurod1	42.1978	18	53	43	114	0.8444
Dupont	Neurod1	42.1978	18	53	43	114	0.8444
Hansen	Neurod1	58.35165	48	108	105	261	0.0374
Iwata	Neurod1	48.79121	0	17	157	174	1
Jackson	Neurod1	64.3956	2	73	241	316	0.1963
Kanatsuka	Neurod1	49.12088	0	22	155	177	1
Malecki	Neurod1	44.94505	14	75	50	139	0.1004
Malecki	Neurod1	48.46154	25	68	78	171	0.1277
Mockizuki	Neurod1	42.96703	0	12	109	121	1
Owerback	Neurod1	38.46154	10	36	34	80	1
Yamada	Neurod1	43.07692	4	33	85	122	0.7447

Continued

Supplementary Table. Continued

Study	locus	std N	a1/a1	a1/a2	a2/a2	N	*p*-value
Ye	Neurod1	43.2967	0	13	111	124	1
antikainen	PON1 Q192R	45.24735	87	75	7	169	0.0753
aubo	PON1 Q192R	47.73852	154	123	33	310	0.2833
aynacioglu	PON1 Q192R	44.11661	11	43	51	105	0.652
ayub	PON1 Q192R	43.14488	32	15	3	50	0.4242
cascorbi	PON1 Q192R	59.62898	521	391	71	983	0.8721
chen	PON1 Q192R	49.52297	208	166	37	411	0.6341
ferre	PON1 Q192R	46.06007	106	93	16	215	0.6192
gardemann	PON1 Q192R	51.71378	279	216	40	535	0.9141
hasselwander	PON1 Q192R	49.11661	179	178	31	388	0.1905
heijman	PON1 Q192R	52.93286	291	263	50	604	0.4386
hermann	PON1 Q192R	54.64664	362	265	74	701	0.018
hong	PON1 Q192R	45.63604	75	84	32	191	0.3597
imai	PON1 Q192R	49.87633	59	182	190	431	0.1672
ko	PON1 Q192R	46.11307	30	96	92	218	0.5562
lawlor	PON1 Q192R	91.4841	1430	1115	241	2786	0.2662
letellier	PON1 Q192R	43.9576	55	38	3	96	0.3843
leus	PON1 Q192R	44.27562	56	48	10	114	1
liu	PON1 Q192R	44.52297	25	74	29	128	0.1104
mackness	PON1 Q192R	47.24382	156	99	27	282	0.0698
odawara	PON1 Q192R	44.41696	25	53	44	122	0.2648
ombres	PON1 Q192R	45.86572	106	84	14	204	0.7264
osei-hyiaman	PON1 Q192R	46.34276	181	44	6	231	0.1172
pati	PON1 Q192R	43.67491	60	12	8	80	0.0001
pfohl	PON1 Q192R	45.26502	73	77	20	170	1
rice	PON1 Q192R	52.98587	312	241	54	607	0.4298
robertson	PON1 Q192R	85.08834	1317	910	197	2424	0.0263
ruiz	PON1 Q192R	46.90813	140	110	13	263	0.1968
salonen	PON1 Q192R	44.18728	59	43	7	109	1
sangera	PON1 Q192R	46.57244	41	123	80	244	0.6933
sangera	PON1 Q192R	45.17668	77	66	22	165	0.2199

Continued

Supplementary Table. Continued

Study	locus	std N	a1/a1	a1/a2	a2/a2	N	*p*-value
sen-banerjee	PON1 Q192R	51.41343	279	226	13	518	0.000013
senti	PON1 Q192R	49.25795	193	165	38	396	0.7234
serrato	PON1 Q192R	46.62544	120	99	28	247	0.3007
suehiro	PON1 Q192R	46.71378	34	124	94	252	0.5929
tuban	PON1 Q192R	47.57951	136	143	22	301	0.0794
wang	PON1 Q192R	50.65371	193	230	52	475	0.1919
watzinger	PON1 Q192R	46.85512	147	96	17	260	0.8684
yamada	PON1 Q192R	62.89753	523	516	129	1168	0.9473
zama	PON1 Q192R	44.29329	17	61	37	115	0.4408
Febbo	SRD5A2	73.11111	78	330	391	799	0.5038
Hsing	SRD5A2	51.06667	105	136	62	303	0.1591
Latil	SRD5A2	44.53333	8	64	84	156	0.4069
Lunn	SRD5A2	44.17778	13	58	77	148	0.6865
Lunn	SRD5A2	37.95556	1	5	2	8	1
Margiotti	SRD5A2	42.75556	9	40	67	116	0.4555
Nam	SRD5A2	44.8	21	69	72	162	0.488
Pearce	SRD5A2	64.26667	76	263	261	600	0.4703
Pearce	SRD5A2	50.22222	43	156	85	284	0.0518
Pearce	SRD5A2	55.86667	21	159	231	411	0.4226
Soderstrom	SRD5A2	44.66667	16	66	77	159	0.7128
Yamada	SRD5A2	46.62222	50	97	56	203	0.5742
abbar	TPH	58.38095	30	133	118	281	0.5079
bellivier	TPH	40.57143	11	45	38	94	0.8226
du	TPH	39.61905	13	52	19	84	0.047
furlong	TPH	73.2381	67	208	162	437	1
geijer	TPH	40.95238	13	47	38	98	1
kunugi	TPH	51.52381	55	105	49	209	1
ono	TPH	44.19048	26	71	35	132	0.3875
paik	TPH	54.09524	66	116	54	236	0.8961
rujescu	TPH	62.66667	40	155	131	326	0.6315
souery	TPH	47.52381	27	74	66	167	0.4161

Continued

Supplementary Table. Continued

Study	locus	std N	a1/a1	a1/a2	a2/a2	N	*p*-value
tsai	TPH	50.66667	33	113	54	200	0.0624
turecki	TPH	43.90476	18	71	40	129	0.1507
zaisman	TPH	42.28571	34	54	24	112	0.8488
Blazer	VDR TaqI	50.06579	35	74	59	168	0.2079
Blazer	VDR TaqI	39.93421	3	2	9	14	0.0261
Correa-Cerro	VDR TaqI	45.26316	11	52	32	95	0.1957
Furuya	VDR TaqI	42.96053	1	18	41	60	1
Gsur	VDR TaqI	51.51316	22	87	81	190	1
Habuchi	VDR TaqI	61.18421	3	81	253	337	0.3282
Hamasaki	VDR TaqI	47.76316	8	34	91	133	0.0823
Kibel	VDR TaqI	41.31579	7	15	13	35	0.4978
Kibel	VDR TaqI	39.40789	1	3	2	6	1
Luscombe	VDR TaqI	49.14474	30	67	57	154	0.2436
Ma	VDR TaqI	77.76316	86	299	204	589	0.1706
Medeiros	VDR TaqI	52.56579	41	92	73	206	0.2529
Suzuki	VDR TaqI	45.92105	2	20	83	105	0.6184
Tayeb	VDR TaqI	63.94737	62	181	136	379	0.915
Taylor	VDR TaqI	49.67105	36	73	53	162	0.2677
Taylor	VDR TaqI	39.53947	1	6	1	8	0.4779
Watanabe	VDR TaqI	52.30263	6	36	160	202	0.042