

ARTICLE

Low-density lipoprotein receptor mutations generate synthetic genome-wide associations

Daniëlla M Oosterveer^{1,5}, Jorie Versmissen^{1,5}, Joep C Defesche^{2,5}, Suthesh Sivapalaratnam^{2,5}, Mojgan Yazdanpanah³, Monique Mulder¹, Leonie van der Zee¹, André G Uitterlinden^{1,4}, Cornelia M van Duijn⁴, Albert Hofman⁴, John JP Kastelein², Yurii S Aulchenko⁴ and Eric JG Sijbrands^{*,1}

Genome-wide association (GWA) studies have discovered multiple common genetic risk variants related to common diseases. It has been proposed that a number of these signals of common polymorphisms are based on synthetic associations that are generated by rare causative variants. We investigated if mutations in the low-density lipoprotein receptor (*LDLR*) gene causing familial hypercholesterolemia (FH, OMIM #143890) produce such signals. We genotyped 480 254 polymorphisms in 464 FH patients and in 5945 subjects from the general population. A total of 28 polymorphisms located up to 2.4 Mb from the *LDLR* gene were genome-wide significantly associated with FH ($P < 10^{-8}$). We replicated the 10 top signals in 2189 patients with a clinical diagnosis of FH and in 2157 subjects of a second sample of the general population ($P < 0.000087$). Our findings confirm that rare variants are able to cause synthetic genome-wide significant associations, and that they exert this effect at relatively large distances from the causal mutation.

European Journal of Human Genetics (2013) 21, 563–566; doi:10.1038/ejhg.2012.207; published online 12 September 2012

Keywords: familial hypercholesterolemia; synthetic associations; LDLR mutations; genome-wide association studies

INTRODUCTION

Genome-wide association (GWA) studies comparing hundreds of thousands of common polymorphisms between persons with and without disease have successfully identified genetic risk variants with frequencies $> 5\%$. These genetic variants are expected to tag frequent causal variants, and it is considered unlikely that they are coupled to mutations having frequencies far below 1%.^{1,2} However, Dickson *et al*³ performed intriguing simulation studies, suggesting that stochastically occurring coupling between common polymorphisms and rare causal variants can give rise to synthetic association signals in GWA studies. By exploring two autosomal recessive disorders (hearing loss and sickle cell anemia), they showed that GWA signals occur when multiple rare mutations are coupled to a limited number of common polymorphisms. In addition, mutations in the nucleotide-binding oligomerization domain-containing protein 2 (*NOD2*) gene produced synthetic association signals in patients with Crohn's disease, and rare variants of hypertriglyceridemia were identified by a GWA.^{4,5} Remarkably, the functional variants can be located at a large distance (2.5 Mb) from the synthetic association signal. The relevance of synthetic signals has been questioned and the examples may be considered anecdotal.^{6–8} This is supported by extrapolations from preliminary data of the 1000 Genomes Project showing that associations can be synthetic but that this will hardly occur in case of complex disorders.^{9,10} However, these findings are difficult to interpret, because this phase of the project had power to identify variants with a frequency of $> 1\%$ and because undeniable

causal variants were lacking. On the basis of our nationwide molecular screening program for familial hypercholesterolemia (FH, OMIM #143890), the rare causal mutation has been identified in a large population displaying extensive heterogeneity at the low-density lipoprotein receptor (*LDLR*) gene.¹¹ Hence, we had the opportunity to test whether well-known rare variants of the *LDLR* gene causing an autosomal dominant disorder create synthetic associations in real data of independent samples.

In the present study, we investigated whether replicable synthetic associations occur based on mutations in the *LDLR* gene causing FH.

MATERIALS AND METHODS

The 'Association study of coronary heart disease (CHD) Risk factors in the Genome using an Old-versus-young Setting' (ARGOS-NL) population consisted of 500 patients, who were selected from 17 000 FH patients with an identified causal mutation in the *LDLR* gene (segregating in the families and not present in controls) from the nationwide molecular screening program for FH. The scope of the original ARGOS study was to identify previously unknown risk factors for CHD. Phenotypic data were acquired from general practitioners and by reviewing medical records at the lipid and cardiology clinics. We selected the 264 youngest patients with severe premature CHD (defined as myocardial infarction, CBAG and PTCA) and the 236 oldest patients without any sign or symptom of CHD, stratified for sex. At least all first, second and third degree relatives were excluded from the study using the pedigree information. After exclusion of family members and patients, whose DNA was not available, 464 FH subjects were available as cases for the GWA study. The control group consisted of 5945 subjects of the Rotterdam study

¹Department of Internal Medicine–D435, Erasmus University Medical Center, Rotterdam, The Netherlands; ²Department of Internal Medicine, Amsterdam Medical Center, Amsterdam, The Netherlands; ³Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada; ⁴Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands

*Correspondence: Dr E.J.G. Sijbrands, Department of Internal Medicine–D435, Erasmus University Medical Center, PO Box 2040, 3000 CA Rotterdam, The Netherlands. Tel: +31 10 70 33 283; Fax: +31 10 70 33 639; E-mail e.sijbrands@erasmusmc.nl

⁵These authors contributed equally to this work.

Received 7 September 2011; revised 9 August 2012; accepted 9 August 2012; published online 12 September 2012

(RS-I) without excluding FH patients. The Rotterdam Study is a population-based cohort described in detail elsewhere.¹² All cases and controls were of apparent Caucasian descent. All subjects gave informed consent and the ethics institutional review board approved the study protocols.

Illumina Infinium HumanHap550K chips (Illumina, San Diego, CA, USA) were used to genotype the subjects of the ARGOS population and the Rotterdam study. Polymorphisms were excluded if they had (i) unsuccessful genotyping in $\geq 2\%$ of all subjects, (ii) a minor allele frequency $< 5\%$ or (iii) showed deviation from Hardy–Weinberg equilibrium ($P < 0.0001$). For each polymorphism that passed the quality control in both populations, we compared allele frequencies between ARGOS and the Rotterdam study using Plink version 1.06.¹³ Cross-tabulation and stepwise forward regression analysis of genome-wide significant polymorphisms in the *LDLR* gene region were performed using SPSS version 15.0 (IBM, Armonk, NY, USA). After identity-by-descent (IBD) analysis in Plink version 1.06 (Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA, and the Broad Institute of Harvard & MIT, Cambridge, MA, USA), a multidimensional scaling (MDS) plot was made with R package 2.8.1 (University of Auckland, Grafton, New Zealand, Australia). Haplotypes were constructed with haplo.stats in R package 2.8.1.

To replicate our results, we compared the genotype frequencies of the ten most significant polymorphisms between a second cohort of 2189 patients with clinical FH and another population-based control sample consisting of 2157 subjects of the Rotterdam study (RS-II).^{12,14} The above-described Illumina chips were used to genotype the subjects of the RS-II. In the second FH cohort, the polymorphisms were genotyped using TaqMan (Applied Biosystems, Foster City, CA, USA). The replication analyses were performed with SPSS version 15.0. All subjects gave informed consent and the ethics institutional review board approved the study protocols.

RESULTS

We compared the frequencies of 480 254 polymorphisms between 464 FH patients (ARGOS-NL) and 5945 subjects from the general population (RS-I). Overall, 28 polymorphisms were genome-wide significantly associated with FH (Supplementary Figure 1). Of these, 13 polymorphisms were located up to 1 Mb upstream or downstream from the *LDLR* gene (Table 1 and Supplementary Figure 1). In addition, a number of signals were even further away from this gene: 15 polymorphisms were located up to 2.4 Mb from the gene (Table 1). Remarkably, the polymorphisms within the *LDLR* gene were not associated with FH. Entering the polymorphism in a stepwise forward logistic regression model demonstrated that the rs2304240, rs4804149, rs387865 and rs4804636 polymorphisms were independently associated with FH with *P*-values ranging from 2.21×10^{-34} to 1.61×10^{-10} (Table 1). The genomic inflation factor was 1.20. Correction for up to 20 principal components slightly decreased the genomic inflation factor (1.19). The largest cluster found by IBD analysis existed of five persons and between these persons the maximum shared IBD was 4%. Adjustment for age, which might be considered as an adjustment for group, did lower the inflation factor to 1.03 but the significant association between the polymorphisms on chromosome 19 and FH yielded similar results.

Replication of results

We compared the genotype frequencies of the 10 most significant polymorphisms between a second FH cohort of 2189 patients and 2157 subjects of the general population (RS-II): all 10 polymorphisms on chromosome 19 were significantly associated with FH (all $P < 10^{-4}$, Supplementary Table 1). An *LDLR* mutation had been identified in 1258 patients of this second FH cohort. Analyses restricted to these mutation carriers resulted in higher odds ratios with lower *P*-values, whereas analyzing clinical FH patients in whom

Table 1 Genome-wide significant polymorphisms on chromosome 19 associated with familial hypercholesterolemia

Polymorphism	Position	Allele	The Rotterdam study		ARGOS		P-value
			MAF	OR	MAF	OR	
rs7259203	8 672 833 ^a	A	0.16	0.24	1.71	4.32×10^{-11}	
rs2033483	9 115 342 ^a	G	0.11	0.18	1.76	6.51×10^{-10}	
rs10422772	9 584 423 ^a	A	0.14	0.08	0.49	1.41×10^{-8}	
rs10405652	9 602 123 ^a	A	0.14	0.08	0.49	2.03×10^{-8}	
rs10411082	9 602 216 ^a	A	0.21	0.13	0.58	4.51×10^{-8}	
rs10418705	9 629 925 ^a	G	0.20	0.13	0.57	3.16×10^{-8}	
rs10417523	9 681 194 ^a	C	0.14	0.07	0.48	1.87×10^{-8}	
rs10415132	9 724 483 ^a	A	0.21	0.13	0.57	2.08×10^{-8}	
rs7247038	9 739 695 ^a	G	0.21	0.13	0.58	4.16×10^{-8}	
rs2431820	9 996 284 ^a	G	0.06	0.11	2.09	1.05×10^{-10}	
rs10417403	10 033 516 ^a	G	0.26	0.34	1.50	3.33×10^{-8}	
rs3745264	10 290 566	A	0.15	0.25	1.97	6.00×10^{-17}	
rs7507634	10 297 562	A	0.11	0.22	2.24	1.36×10^{-20}	
rs2304240	10 310 392	A	0.16	0.33	2.63	2.21×10^{-35}	
rs280519	10 333 933	G	0.50	0.40	0.67	1.78×10^{-8}	
rs8112449	10 381 064	A	0.30	0.22	0.63	3.63×10^{-8}	
rs7256672	10 440 474	C	0.36	0.45	1.49	1.14×10^{-8}	
rs1529711	10 884 434	A	0.16	0.25	1.73	8.58×10^{-12}	
rs4804149	11 145 028	G	0.29	0.40	1.64	1.83×10^{-12}	
rs387865	11 145 539	A	0.29	0.39	1.57	1.61×10^{-10}	
rs8111456	11 162 147	G	0.34	0.44	1.54	3.30×10^{-10}	
rs313624	11 413 910	A	0.48	0.43	1.48	2.18×10^{-8}	
rs4804636	11 703 323	A	0.34	0.45	1.58	3.15×10^{-11}	
rs286262	11 708 615	A	0.37	0.48	1.55	2.43×10^{-10}	
rs11880217	12 279 795 ^a	A	0.16	0.24	1.66	4.89×10^{-10}	
rs12610507	12 297 410 ^a	C	0.17	0.25	1.62	1.70×10^{-9}	
rs2967890	12 901 647 ^a	G	0.23	0.31	1.55	2.91×10^{-9}	
rs16042	13 202 037 ^a	A	0.11	0.17	1.76	7.05×10^{-10}	

Abbreviations: LDL, low-density lipoprotein; MAF, minor allele frequency; OR, odds ratio.

^aLocated 1–2 Mb from the *LDLR* gene (located 11 061 132–11 105 490 according to HapMap). The other polymorphisms are located < 1 Mb from the *LDLR* gene.

no *LDLR* mutation was detected ($n = 931$) did not result in replication with the exception of rs4804149 (Supplementary Table 2).

Associated polymorphisms do not identify patients with (familial) hypercholesterolemia

As the presence of *LDLR* mutations produced strong signals in our GWA study, we hypothesized that these signals might identify subjects with FH in the general population. The prevalence of FH is $\sim 1:500$.¹¹ This indicates that 12 out of 5945 subjects of the RS-I and 4 out of 2157 subjects of RS-II were expected to carry an *LDLR* mutation. However, a MDS plot based on the DNA similarity at the *LDLR* gene ± 1 Mb between subjects from the ARGOS and all controls (RS-I and RS-II) did not show any pattern that distinguished carriers of *LDLR* mutations from non-carriers.

We constructed four haplotypes of rs2304240, rs4804149, rs387865 and rs4804636 polymorphisms that were independently associated with FH. These haplotypes were associated with having an *LDLR* mutation when comparing ARGOS to RS-I and RS-II ($P < 2.61 \times 10^{-5}$). Within both RS samples the haplotypes were not associated with LDL cholesterol level or possible FH phenotype, defined as total cholesterol level > 7 mmol/l and triglyceride level < 4 mmol/l, or LDL cholesterol

Table 2 Genotype frequencies of two significant polymorphisms close to the *LDLR* gene in the Rotterdam study, in the whole ARGOS and within ARGOS in carriers of the same *LDLR* mutation

	rs3745264				Minor allele	rs2304240				Minor allele
	AA	AC	CC	P-value		AA	AG	GG	P-value	
The Rotterdam study	131 (2.2)	1506 (25.4)	4283 (72.3)		A	143 (2.4)	1619 (27.4)	4154 (70.2)		A
ARGOS	22 (4.7)	193 (41.6)	249 (53.7)	6.00×10^{-17}	A	37 (0.8)	228 (49.5)	196 (42.2)	2.21×10^{-35}	A
<i>Carriers mutation</i>										
191-2	6 (18.8)	26 (81.3)	0 (0.0)	$<1.00 \times 10^{-6}$	C	6 (18.8)	26 (81.3)	0 (0.0)	$<1.00 \times 10^{-6}$	A
W23X	4 (13.8)	10 (34.5)	15 (51.7)	5.30×10^{-5}	A	4 (13.8)	9 (31.0)	16 (55.2)	3.07×10^{-4}	A
313 + ½	5 (9.1)	44 (80.0)	6 (10.9)	$<1.00 \times 10^{-6}$	A	6 (10.9)	43 (78.2)	6 (10.9)	$<1.00 \times 10^{-6}$	A
E207K	0 (0.0)	5 (17.9)	23 (82.1)	0.45	A	0 (0.0)	4 (14.3)	24 (85.7)	0.19	G
S285L	0 (0.0)	19 (95.0)	0 (0.0)	$<1.00 \times 10^{-6}$	C	3 (15.0)	17 (85.0)	0 (0.0)	$<1.00 \times 10^{-6}$	A
G322S	2 (16.7)	9 (75.0)	1 (8.3)	$<1.00 \times 10^{-6}$	C	0 (0.0)	3 (27.3)	8 (72.7)	0.87	G
R329X	0 (0.0)	0 (0.0)	7 (100.0)	0.263	A	0 (0.0)	0 (0.0)	7 (100.0)	0.23	G
Cape Town-2	2 (9.5)	19 (90.5)	0 (0.0)	$<1.00 \times 10^{-6}$	C	4 (19.0)	17 (81.0)	0 (0.0)	$<1.00 \times 10^{-6}$	A
1359-1	0 (0.0)	12 (17.9)	55 (82.1)	0.15	A	11 (16.4)	49 (73.1)	7 (10.4)	$<1.00 \times 10^{-6}$	A
L401P	0 (0.0)	2 (25.0)	6 (75.0)	0.91	A	1 (12.5)	3 (37.5)	4 (50.0)	0.13	A
V408M	0 (0.0)	2 (15.4)	11 (84.6)	0.59	A	0 (0.0)	3 (23.1)	10 (76.9)	0.79	G
L590F	0 (0.0)	0 (0.0)	8 (100.0)	0.22	A	0 (0.0)	8 (100.0)	0 (0.0)	2.50×10^{-5}	A
P664L	0 (0.0)	1 (12.5)	7 (87.5)	0.62	A	0 (0.0)	0 (0.0)	8 (100.0)	0.18	G
P678L	0 (0.0)	2 (40.0)	3 (60.0)	0.73	A	0 (0.0)	2 (40.0)	3 (60.0)	0.78	G

Values are counts (%). P-values are result of χ^2 test with The Rotterdam study as reference group.

level >6 mmol/l and a history of premature CHD (Supplementary Table 3).

As each LDLR mutation might cosegregate with different alleles of common polymorphisms, we compared the genotype frequencies of the polymorphisms found by GWA between the subjects of the ARGOS study and the RS-I, stratified for all mutations that were present in 10 or more patients. Different mutations indeed appeared to cosegregate with different alleles of the polymorphisms, because of the absence of homozygosity for the wild-type allele. The results of rs3745264 and rs2304240, the two most significant hits of our study, are shown as example in Table 2. The cosegregation of a mutation with specific alleles of common polymorphisms was most pronounced in carriers of the *LDLR* mutation 191-2 in intron 2 (association with the minor allele in seven polymorphisms), S285L in exon 6 (association with the minor allele in nine polymorphisms) and a large deletion of 2.5 kb from exon 7–8 called the Cape Town-2 mutation (association with the minor alleles in eight polymorphisms). This cosegregation is confirmed by the disappearance of the associations between FH and the two polymorphisms in Table 2 after removing all carriers of these mutations from the analyses: $P=0.38$ for rs3745264 (191-2, 313 + 1/2, S285L, G322S and Cape Town-2 removed) and $P=0.95$ for rs2304240 (191-2, 313 + 1/2, S285L, Cape Town-2, 1359-1 and L590F removed).

We searched for these particular combinations of minor alleles in persons with a possible FH phenotype in the Rotterdam study and identified seven possible FH patients. Sequencing of the *LDLR* gene including the promoter region and splice sites, as well as the lipid profiles, did not identify FH patients among these subjects.

DISCUSSION

In this study, we demonstrate that *LDLR* mutations give rise to multiple independent synthetic associations and that these signals occur up to 2.4 Mb upstream and downstream of the *LDLR* gene.

Using the associated polymorphisms, we were not able to identify FH patients among the general population.

The synthetic associations in our study could have been the result of cryptic family relations or population stratification. The genomic inflation factor (1.20) was relatively high. This indicates that there might be population stratification. Correction for up to 20 principal components only slightly decreased the genomic inflation factor (1.19) and did not change the association between the polymorphisms and FH.¹⁵ In addition, our selection procedure of ARGOS aimed at genetic heterogeneity at the *LDLR* locus: we took a sample from our nationwide screening program and excluded relatives, resulting in identification of 96 different *LDLR* mutations in 464 FH patients. The most frequent mutation (1359-1 splice defect in intron 9) was found in 14%, which is similar to its frequency in the overall Dutch FH population. To further explore cryptic family relations, we used an IBD analysis. The largest cluster existed of five persons and between these persons the maximum shared IBD was low. Taken together, this confirms that we succeeded in excluding relatives and even distant relationships did not influence our results. Therefore, random inflation by cryptic family relations or population stratification is a very unlikely explanation for the 28 GWA signals at the *LDLR* locus.

Remarkably, the polymorphisms within the *LDLR* gene were not associated with FH. Particular mutations associated with specific polymorphism within the *LDLR* gene, but these numbers were too small to produce a genome-wide significant signal. Alternatively, one could argue that they did not tag the mutations at all because of their location outside the linkage disequilibrium blocks.

We also found a genome-wide significantly associated polymorphism that was not located on chromosome 19 but on chromosome 16, however, genotyping of this polymorphism did not result in distinct genotype clusters, indicating a false-positive result.

As the presence of an *LDLR* mutation produced signals in our GWA study, we hypothesized that these signals might identify subjects

with FH in the general population. An MDS plot based on the DNA similarity at the *LDLR* gene ± 1 Mb between subjects from the ARGOS and the Rotterdam study did not show any pattern that distinguished carriers of *LDLR* mutations from non-carriers. Alternatively to this MDS analysis, we tried to identify subjects with FH in the Rotterdam study by constructing haplotypes of the independently associated polymorphisms. As expected, some of these haplotypes were associated with having an *LDLR* mutation, when comparing ARGOS to the Rotterdam study. However, within the Rotterdam study none of the haplotypes were associated with LDL cholesterol level or possible FH phenotype. As both MDS analysis and haplotype analyses were not able to identify FH subjects in the Rotterdam study, we hypothesized that different mutations might cosegregate with different alleles of the polymorphisms. We therefore compared frequencies of the genotypes of the significantly associated polymorphisms between carriers of the mutations in ARGOS that were present in 10 or more patients. Different mutations indeed appeared to cosegregate with different alleles of the polymorphisms. The significance of these associations was most likely driven by the fact that specific mutations cosegregated with the minor allele of the polymorphism, making it the common allele among carriers of these mutations. For some of the more frequent mutations, cosegregation with the rare allele of a polymorphism at the *LDLR* locus was even more certain, as homozygosity for the common allele of the polymorphism was absent among the carriers of the mutation, however, using this information did not lead to identification of FH patients in the general population. Different mutations cosegregated with different alleles, but of course the power to detect specific *LDLR* mutations in the general population was limited.

Our GWA study picked up genetic heterogeneity at a locus of an exemplary Mendelian trait, supporting that multiple rare variants produce synthetic GWA signals. This suggests that we have opportunities to detect rare variants in a GWA. Our study cannot be used to estimate what proportion of heritabilities will be explained by rare variants.¹⁶ We merely show that if rare undeniable causal variants are present strong signals may appear. A very large population (~ 250 000 persons) is required to identify synthetic signals of *LDLR* mutations with similar power in a population-based setting. Sequencing of large populations is still too expensive. Perhaps, we should enrich the current sequencing initiatives with our cohorts containing monogenic disorders to gain more in depth insight in the behavior of mutated loci. The phenomenon of synthetic associations were first described by Dickson *et al.*³ Recently, Anderson *et al.*⁷ suggested that the consistency of GWA results in different populations is an argument against involvement of synthetic associations. We clearly replicated the signals in an independent sample in line with consistency among populations, but the cases and controls were again Caucasian. It would be of great interest to know whether our results would be consistent in populations of other ethnicities. The distributions of the *LDLR* mutations differ between populations and probably the polymorphisms of the synthetic association signals as well. Unfortunately, cohorts of genetically heterogeneous FH patients from other populations are currently not large enough for this analysis.

Wray *et al.*⁶ argued that the allele frequencies of the signals in GWA's are too high to be explained by rare variants. In the initial and the replication phase, however, we found that a large number of different mutations in the *LDLR* produced a wide range in minor allele frequencies (MAFs) of all found polymorphisms including the independently associated polymorphism (rs2304240, rs4804149, rs387865, rs4804636). Although overall the MAFs varied over a wide range, there might be a trend of higher MAFs of the polymorphisms located <1 Mb from the *LDLR* gene and lower MAFs of the polymorphisms located further away.

We confirm that mutations causing a monogenic disorder give rise to multiple independent synthetic associations and that these signals occur up to 2.4 Mb away from the rare functional variants. However, we could not simplify molecular diagnostics of FH with these synthetic associations.

CONFLICT OF INTEREST

JPK has received research funding from, served as a consultant for, and received honoraria for lectures from AstraZeneca, Genzyme, ISIS, Merck, Novartis, Pfizer, Roche, Schering Plough and Sankyo. EJGS has received not-drug-related research funding from Pfizer and Merck Sharp & Dohme, The Netherlands. The remaining authors declare no conflict of interest.

- McCarthy MI, Abecasis GR, Cardon LR *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet* 2008; **9**: 356–369.
- McCarthy MI, Hirschhorn JN: Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet* 2008; **17**: R156–R165.
- Dickson SP, Wang K, Krantz I *et al.* Rare variants create synthetic genome-wide associations. *PLoS Biol* 2010; **8**: e1000294.
- Wang K, Dickson SP, Stolle CA *et al.* Interpretation of association signals and identification of causal variants from genome-wide association studies. *Am J Hum Genet* 2010; **86**: 730–742.
- Johansen CT, Wang J, Lanktree MB *et al.* Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat Genet* 2010; **8**: 684–687.
- Wray NR, Purcell SM, Visscher PM: Synthetic associations created by rare variants do not explain most GWAS results. *PLoS Biol* 2011; **9**: e1000579.
- Anderson CA, Soranzo N, Zeggini E *et al.* Synthetic associations are unlikely to account for many common disease genome-wide association signals. *PLoS Biol* 2011; **9**: e1000580.
- Dudbridge F, Fletcher O, Walker K *et al.* Estimating causal effects of genetic risk variants for breast cancer using marker data from bilateral and familial cases. *Cancer Epidemiol Biomarkers Prev* 2011; **21**: 262–272.
- 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**: 1061–1073.
- Kent Jr JW: Rare variants, common markers: synthetic association and beyond. *Genet Epidemiol* 2011; **35**(Suppl 1): S80–S84.
- Leigh SE, Foster AH, Whittall RA *et al.* Update and analysis of the University College London low density lipoprotein receptor familial hypercholesterolemia database. *Ann Hum Genet* 2008; **72**: 485–498.
- Hofman A, van Duijn CM, Franco OH *et al.* The Rotterdam study: 2012 objectives and design update. *Eur J Epidemiol* 2011; **26**: 657–686.
- Purcell S, Neale B, Todd-Brown K *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
- Jansen AC, van Aalst-Cohen ES, Tanck MW *et al.* The contribution of classical risk factors to cardiovascular disease in familial hypercholesterolemia: data in 2400 patients. *J Intern Med* 2004; **256**: 482–490.
- Price AL, Patterson NJ, Plenge RM *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–909.
- Goldstein DB: The importance of synthetic associations will only be resolved empirically. *PLoS Biol* 2011; **9**: e1001008.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)