American Journal of Epidemiology
Vol. 172, No. 8
DOI: 10.1093/aje/kwq234
Advance Access publication:
September 28, 2010

Human Genome Epidemiology (HuGE) Review

Genome-wide Significant Associations for Variants With Minor Allele Frequency of 5% or Less—An Overview: A HuGE Review

Orestis A. Panagiotou, Evangelos Evangelou, and John P. A. Ioannidis*

* Correspondence to Dr. John P. A. Ioannidis, Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, P.O. Box 1186, Ioannina 45110, Greece (e-mail: jioannid@cc.uoi.gr).

Initially submitted March 8, 2010; accepted for publication June 24, 2010.

The authors survey uncommon variants (minor allele frequency, ≤5%) that have reached genome-wide significance ($P \leq 10^{-7}$) in genome-wide association study(ies) (GWAS). They examine the typical effect sizes of these associations; whether they have arisen in multiple GWAS on the same phenotype; and whether they pertain to genetic loci that have other variants discovered through GWAS, perceived biologic plausibility from the candidate gene era, or known mutations associated with related phenotypes. Forty-three associations with minor allele frequency of 5% or less and $P \leq 10^{-7}$ were studied, 12 of which involved nonsynonymous variants. Per-allele odds ratios ranged from 1.03 to 22.11. Thirty-two associations had $P \leq 10^{-8}$. Eight uncommon variants were identified in multiple GWAS. For 14 associations, also other common polymorphisms with genome-wide significance were identified in the same loci. Thirteen associations pertained to genetic loci considered to have biologic plausibility for association in the candidate gene era, and mutations with related phenotypic effects were identified for 11 associations. Twenty-five uncommon variants are common in at least 1 of the 4 different ancestry samples of the International HapMap Project. Although the number of uncommon variants with genome-wide significance is still limited, these data suggest a possible confluence of rare/uncommon and common genetic variation on the same genetic loci.

epidemiology; gene frequency; genes; genetics; genome-wide association study; genomic structural variation; Human Genome Project; polymorphism, single nucleotide

Abbreviations: AIDS, acquired immunodeficiency syndrome; GWAS, genome-wide association study(ies); HDL, high density lipoprotein; HuGE, Human Genome Epidemiology; LDL, low density lipoprotein; MAF, minor allele frequency; SNP, single-nucleotide polymorphism.

Editor's note: *This article also appears on the website of the Human Genome Epidemiology Network (http://www.cdc.gov/genomics/hugenet/default.htm).*

The large majority of discoveries in human genome epidemiology in the last 5 years pertain to associations of common genetic variants with diverse phenotypes (1, 2). In particular, genome-wide association study(ies) (GWAS) have dramatically increased the yield of associations with very high levels of statistical significance (3–6). GWAS conducted to date have used common genetic markers and have found mostly low penetrance variants with small effects (7, 8). Their genotyping platforms offer very good coverage across the genome for variants with minor allele frequency (MAF) of greater than 5% (8, 9). However, variants with lower MAF are either excluded routinely from commercial platforms or inadequately covered (8, 10). For most diseases, the associations identified to date through GWAS account for only a small portion of the estimated total heritability (11–13). There are many speculations about the reasons underlying the residual unknown component of the genetic architecture—also described as the "genetic dark matter" (13, 14). One explanation is the presence of associations involving uncommon (MAF, ≤5%) and rare (MAF, <0.5%) variants (8, 13). Associations with uncommon/rare

Table 1. Eligible Associations With a Minor Allele Frequency of 5% or Less and $P \leq 10^{-7}$

Disease/Trait	Region	Reported Gene(s)	Variant-Risk Allele	Position/Function	Risk Allele Frequency	Population Descent	*P* Value	Odds Ratio	95% Confidence Interval	Reference
ALL	12q24.22	*KRTHB5*	rs2089222-A	Intronic	0.03	European	8.0×10^{-8}	2.26	1.60, 3.00	39
AIDS progression	6p21.33	*HCP5, MICB, MCCD1, BAT1, LTB, TNF*	rs2395029-G	Nonsynonymous coding (missense)	0.03	European	3.0×10^{-19}	3.47	2.39, 5.04	40
	6p21.3	*C6orf48*	rs9368699-C	5′-UTR	0.03	European	2.0×10^{-11}	NR	NR	
Blue vs. green eyes	15q13.1	*OCA2*	rs1667394-A	Intronic	0.97	European	2.0×10^{-53}	3.67	2.67, 5.05	41
Freckles	16q24.3	*MC1R*	rs1805007-T	Nonsynonymous coding (missense)	0.05	European	1.0×10^{-96}	3.33	2.92, 3.80	
BMD (lumbar spine)	13q14	*AKAP11*	rs180851-C (I)	Intergenic[a]	0.95	European	2.0×10^{-12}	1.46[b]	1.31, 1.63	42
	13q14	*AKAP11*	rs7326472-A	Intergenic[a]	0.95	European	1.0×10^{-10}	1.39[b]	1.24, 1.54	
	13q14	*AKAP11*	rs12854504-T (I)	Intergenic[a]	0.95	European	1.0×10^{-10}	1.39[b]	1.24, 1.54	
	13q14	*AKAP11*	rs7998154-T (I)	Intergenic[a]	0.02	European	2.0×10^{-8}	1.75[b]	1.46, 2.10	
	13q14	*TNFSF11*	rs6561055-G (I)	Intergenic[a]	0.95	European	3.0×10^{-10}	1.39[b]	1.24, 1.54	
	13q14	*TNFSF11*	rs17639156-T (I)	Intergenic[a]	0.95	European	5.0×10^{-10}	1.39[b]	1.24, 1.54	
Cognitive performance	Xp22.2	*HCCS*	rs5934953-C	Intronic	0.02	European	1.0×10^{-7}	NR	NR	43
Crohn's disease	16q12.1	*NOD2*	rs2066844-T	Nonsynonymous coding (missense)	0.05	European	1.0×10^{-18}	2.48	1.98, 3.10	37
	16q12.1	*NOD2*	rs2066845-C	Nonsynonymous coding (missense)	0.01	European	8.0×10^{-10}	3.04	2.09, 4.42	
	16q12.1	*NOD2*	rs2066847-C	Frameshift coding	0.04	European	3.0×10^{-49}	4.30	3.42, 5.42	
	12q12	*LRRK2, MUC19*	rs11175593-T (I)	Intronic	0.02	European	3.0×10^{-10}	1.54	1.34, 1.76	38
HDL cholesterol	20q13.12	*HNF4A*	rs1800961-C (I)	Nonsynonymous coding (missense)	0.97	European	8.0×10^{-10}	1.41[b]	1.27, 1.57	44
	9q31.1	*ABCA1*	rs9282541-T	Nonsynonymous coding (missense)	0.03	European, Mexicans, Asian Indians	5.0×10^{-8}	1.33[b,c]	1.21, 1.45	45
Hematocrit	6p22.1	*HFE*	rs1800562-A (I)	Nonsynonymous coding (missense)	0.04[d]	European	2.0×10^{-9}	1.74	1.45, 2.09	36
Hemoglobin	6p22.1	*HFE*	rs1800562-A (I)	Nonsynonymous coding (missense)	0.04[d]	European	6.0×10^{-19}	1.33	1.25, 1.42	
LDL cholesterol	1p32.3	*PCSK9*	rs11591147-G	Nonsynonymous coding (missense)	0.99	European	2.0×10^{-44}	2.34[b]	2.07, 2.64	46
MCH	6p22.2	*SLC17A3*	rs1408272-G (I)	Unknown	0.03[d]	European	4.0×10^{-39}	1.03	1.02, 1.04	36
MCV	6p22.1	*HFE*	rs1800562-A (I)	Nonsynonymous coding (missense)	0.04[d]	European	1.0×10^{-23}	12.83	7.73, 20.92	35
NCP	3p22.2	*ITGA9*	rs189897-A	Intronic	0.03	Asian	7.0×10^{-8}	3.18	1.94, 5.21	48
	3p22.2	*ITGA9*	rs197757-T	Intronic	0.03	Asian	1.0×10^{-7}	3.09	1.89, 5.05	
NSCL	18q22.3	Intergenic	rs17085106-T	Intergenic	0.02	European	4.0×10^{-8}	4.07	2.37, 7.00	47

Panic disorder	12p13.31	*TMEM16B*	rs12579350-A	Intronic	0.01	Asian	4.0×10^{-9}	22.11	5.30, 92.14	49
	1q32.1	*PKP1*	rs860554-T	Intronic	0.05	Asian	5.0×10^{-8}	4.03	2.40, 6.76	
Prostate cancer	8q24.21	Intergenic	rs16901979-A	Intergenic	0.03	European	1.1×10^{-12}	1.79	1.53, 2.11	33
Primary biliary cirrhosis	6p21.3	*C6orf10*	rs2395148-A	Intronic	0.02	European	4.0×10^{-14}	2.87	2.16, 3.82	50
ALP	12q12	*PDZRN4, CNTN1*	rs1880887-C	Intronic	0.03	European	1.0×10^{-10}	NR	NR	51
fT3	17p12	*HS3ST3B1*	rs3848445-C	Unknown	0.05	European	8.4×10^{-9}	NR	NR	
Psoriasis	6p21.33	*HLA-C*	rs2395029-C	Nonsynonymous coding (missense)	0.03	European	2.1×10^{-26}	4.10	3.10, 5.30	52
Response to treatment for ALL	10p12.33	*ST8SIA6*	rs359312-T	Intronic	0.04	European, African, other	9.0×10^{-8}	3.91	1.52, 10.10	53
Response to antipsychotic therapy	2p12	Intergenic	rs17022444-G	Intergenic	0.03	European, African, other	1.0×10^{-10}	NR	NR	54
	4q24	Intergenic	rs7669317-C	Intergenic	0.04	European, African, other	8.0×10^{-8}	NR	NR	
SLE	6q23.3	*TNFAIP3*	rs5029939-G	Intronic	0.03	European	3.0×10^{-12}	2.28	1.80, 2.88	55
	6q23.3	*TNFAIP3*	rs2230926-C	Nonsynonymous coding (missense)	0.04	Asian	1.0×10^{-17}	1.72	1.52, 1.94	56
Tanning	5p13.3	*MATP*	rs35391-C	Intronic	0.97	European	3.0×10^{-10}	2.22	1.72, 2.86	57
Triglycerides	11q23.3	*APOA1, APOC3, APOA4, APOA5*	rs662799-G (I)	Upstream	0.05	European	2.0×10^{-15}	1.31[b,c]	1.22, 1.40	58
	11q23.3	*APOA1, APOC3, APOA4, APOA5, DSCAML1*	rs10892151-A	Intronic	0.03	European	3.0×10^{-29}	NR	NR	59
Type 1 diabetes	7p12.1	*COBL*	rs4948088-C (I)	Unknown	0.95	European	4.0×10^{-8}	1.30	1.11, 1.49	60
Type 2 diabetes	10q25.2	*TCF7L2*	rs7903146-T	Intronic	0.04	Asian	8.0×10^{-12}	1.54	1.36, 1.74	61

Abbreviations: AIDS, acquired immunodeficiency syndrome; ALL, acute lymphoblastic leukemia; ALP, alkaline phosphatase; BMD, bone mineral density; fT3, free triiodothyronine; GWAS, genome-wide association study(ies); HDL, high density lipoprotein; I, risk variants imputed rather than directly genotyped; LDL, low density lipoprotein; MAF, minor allele frequency; MCH, mean corpuscular hemoglobin; MCV, mean corpuscular volume; NCP, nasopharyngeal carcinoma; NHANES, National Health and Nutrition Examination Survey; NR, not reported and data not adequate for computing the missing values; NSCL, nonsyndromic cleft lip with or without cleft palate; SLE, systemic lupus erythematosus; SNP, single-nucleotide polymorphism; 5′-UTR, 5′-untranslated region.

[a] For these SNPs, *AKAP11* and *TNFSF11* were reported as the closest genes in the GWAS, but WGA Viewer and the Ensembl characterized them as "intergenic."

[b] Odds ratio equivalent was calculated from the standardized mean difference.

[c] Odds ratio equivalent was computed from the mean difference using also the population standard deviation from NHANES data on HDL cholesterol and triglyceride levels, because the population standard deviation was not given in the GWAS.

[d] MAFs reported in the original GWAS were based on the International HapMap Project frequencies.

Table 2. Number of Discovered[a] and Expected Associations Implicating Uncommon Variants Split According to Odds Ratio Quartiles and According to Minor Allele Frequency Categories

Minor Allele Frequency, %	Odds Ratio Range in Quartiles							
	<1.40 (Median, 1.33)		1.40–<2.24 (Median, 1.72)		2.24–3.40 (Median, 2.87)		>3.40 (Median, 4.07)	
	Discovered Associations, no.	Expected Associations, no.	Discovered Associations, no.	Expected Associations, no.	Discovered Associations, no.	Expected Associations, no.	Discovered Associations, no.	Expected Associations, no.
1–2	0	—[b]	2	4	3	3	2	2
3–4	3	28	6	6	4	4	6	6
5	6	26	1	1	2	2	1	1

[a] The total number of observed associations in Table 2 is 36 and not 43 as expected from Table 1, because for 7 associations the effect estimates were not retrievable.
[b] Not possible to calculate.

variants may even have substantial genetic effects, but they have been difficult to discover to date, presumably because of inadequate coverage in most GWAS, very large sample size requirements, or inefficient analytical methods (8, 15).

Newer genotyping platforms (including exome and full-genome sequencing) (16–18) and analysis methods (15, 19, 20) are already being explored in the pursuit of associations involving uncommon and rare variants. Nevertheless, even traditional GWAS occasionally have discovered associations that pertain to such single- nucleotide polymorphisms (SNPs). Given that over 400 GWAS have been published to date (21, 22), an overview of this literature can already assemble a substantial corpus of associations with uncommon variants. Such an overview could yield some preliminary insights about these associations and their respective genetic loci. The following questions may be asked: What are the typical effect sizes of these associations, and how robustly are they replicated? Do they arise in single or multiple GWAS on the same phenotype? Are common variants also identified in the same loci? Have these genetic loci been considered to have biologic plausibility for association in the candidate gene era? Are any mutations with related phenotypes already known for these same loci? Are uncommon variants common in populations of different ancestry?

Here, we systematically evaluated these questions by perusing all associations for single-nucleotide variants with MAF of 5% or less that have been discovered in GWAS with strong statistical support.

MATERIALS AND METHODS

Search strategy and eligibility criteria

We screened *A Catalog of Published Genome-Wide Association Studies* (22) hosted by the National Human Genome Research Institute, Office of Population Genetics. The catalog is an online, regularly updated database of SNP–trait associations extracted from published GWAS, which attempt to assay at least 100,000 SNPs. It lists associations with $P<10^{-5}$ (21, 22). We identified all GWAS reporting at least 1 genome-wide significant association ($P \leq 10^{-7}$), regardless of the minor allele frequency of the involved SNP. Because the catalog reports only 1 SNP per gene locus for each association, we also searched all genome-wide significant studies (main articles and supplements) to identify additional associations involving rare/uncommon polymorphisms, regardless of whether they were mentioned in the GWAS *Catalog* or not. The last search was conducted on December 8, 2009.

Eligible associations for this overview were those involving variants with a risk allele frequency of 5% or less or 95% or greater (i.e., MAF, $\leq$5%) and that had attained genome-wide significance by using a threshold of $P \leq 10^{-7}$ in at least 1 GWAS when both the discovery and replication data were combined (23). The risk allele frequency criterion pertained to the control group for case-control designs and to the whole population for other designs. We focused on single-nucleotide variants and excluded genetic associations based on haplotypes or structural variants. If the same variant was found in more than 1 GWAS on the same

phenotype, we counted this as 1 association but recorded all pertinent GWAS.

Data extraction

For each association, we extracted the following data: first author; publication date; journal; title; disease/phenotype; gene; variant (rs number); chromosome region; race/ethnicity of study populations; discovery and replication sample sizes; effect estimates (odds ratios per copy of risk allele for binary outcomes, standardized mean differences for continuous outcomes); and *P* value of the effect estimates including all data (discovery and replication).

Data extraction was conducted independently by 2 of the authors, and disagreements were discussed and resolved with a third investigator. Data extraction was performed directly from the respective GWAS articles and their supplements, because we have noted some discrepancies in the information already extracted in the GWAS *Catalog* and we required increased accuracy and additional information besides what was listed in the *Catalog*.

Evaluation of the eligible associations

We summarized descriptively the phenotypes involved in the eligible associations, the distribution of the risk allele frequencies, *P* values, and effect estimates. Whenever the effect estimates were not given and could not be calculated from the published information, we contacted the authors. To express all effect estimates on the same scale, we converted standardized mean differences to odds ratio equivalents multiplying the respective standardized mean difference by 1.81 to obtain the natural logarithm of the odds ratio (24). This method transforms a standardized mean difference of a quantitative trait into an odds ratio for the dichotomized version of that trait and uses a normality assumption for the effects.

Additionally, we estimated the average sample size of the eligible GWAS. For this typical sample size, we performed calculations to estimate the power to detect associations with various MAFs and odds ratio values at $\alpha = 1 \times 10^{-7}$ under a multiplicative (log-additive) genetic model and under the optimal scenario where there is no loss of power due to multistage process in SNP selection. We used the QUANTO software (25). We categorized associations according to quartiles of odds ratio and according to MAF 1–2%, 3%–4%, and 5%. For each of the resulting 12 categories, we estimated the power *G* of a typical GWAS (average sample size of the analyzed GWAS) to detect an association of that odds ratio and MAF at the GWAS level. We used the median value of odds ratio and the midvalue of MAF in each category for these calculations. For each category of odds ratio and MAF values, one can calculate the total number of variants (those that have been discovered plus those that have not been discovered because of limited power), by multiplying the number of discovered variants by $1/G$.

Using WGAViewer (26, 27), the University of California, Santa Cruz, Human Genome Browser (http://genome.ucsc.edu/cgi-bin/hgGateway) (28), the Single-Nucleotide Polymorphism Database (dbSNP) Build 130 (http://www.ncbi.nlm.nih.gov/projects/SNP/), and the Ensembl (www.ensembl.org) Database, we identified the functional position of the eligible uncommon/rare variants within the respective genes, that is, whether they are located in exons, introns, or promoter regions and whether they cause nonsynonymous changes or frameshift changes.

For each eligible genotype–phenotype association, we identified also all other GWAS listed in the GWAS catalog (22) that had evaluated the same phenotype. We examined if the eligible uncommon/rare variant had been reported by any other GWAS on the same phenotype, regardless of whether it had reached genome-wide significance or not. Moreover, we evaluated whether any other GWAS on the same phenotype reported on associations with any other variants in the same gene locus as the eligible uncommon variant. We use the term ''locus'' here to denote either a single gene or several genes, if the authors of the GWAS could not pinpoint which gene among the several listed was most likely to harbor the functional causative variants (e.g., when genes overlapped or when associated SNPs were located in an area lying between 2 genes). Whenever such other variants were reported, we recorded their effect estimates and *P* values. Then, we examined whether the uncommon variants were in high linkage disequilibrium ($r^2 \geq 0.8$) with the other variants in the same gene locus, using the Web-based tool, SNP Annotation and Proxy Search (SNAP), version 2.1 (29), selecting data based on the International HapMap Project, Phase 3, Release 2, for the HapMap panel with similar ancestry as the population where the uncommon variant was discovered. Upon unavailability of results, we used HapMap, Release 22.

Furthermore, we searched on Human Genome Epidemiology (HuGE) Navigator, a continuously updated database in human genome epidemiology (2), whether the gene loci containing the eligible uncommon/rare variants had been investigated by candidate-gene association studies conducted prior to the discovery of these loci in a GWAS. We recorded the number of studies on gene–phenotype associations involving the same gene locus and phenotype published until the end of the year before the first GWAS proposing the association with the uncommon variant gene locus, as well as the total number of studies published to date. We also recorded any comments made in the eligible GWAS that had identified the uncommon variant regarding prior evidence on the proposed gene locus, for example, if it had been proposed by previous linkage or candidate-gene studies or GWAS. Additionally, for each gene locus, we recorded whether any Mendelian mutations have been previously reported in association with the same, similar/related, or unrelated phenotype(s), using the Online Mendelian Inheritance in Man (OMIM) Database (http://www.ncbi.nlm.nih.gov/omim/).

Finally, we recorded the minor allele frequencies of the eligible uncommon variants in the populations genotyped in the International HapMap Project (30, 31), using data from HapMap, Phases 1 and 2 (31, 32), on people of European, African, and Asian (Chinese and Japanese) ancestry. We then examined whether the eligible SNPs had estimated MAFs of 5% or less in all of these populations or only in some of them.

Table 3. Variants in the Same Gene Loci as the Uncommon Variants, Described in Other Genome-wide Association Study(ies) on the Same Phenotype

Disease/Trait	Uncommon Variant(s)	Gene Locus	Other GWAS That Found Variants in Same Locus, reference	Timing of Other GWAS	Variant	Risk Allele Frequency	*P* Value	Per-Allele Odds Ratio
Blue vs. green eyes	rs1667394-A	*OCA2*	62	Subsequent	(Same)	0.13[a]	3×10^{-87}	1.82
					rs7495174	0.05[a]	0.018	1.60
			63	Subsequent	(Same)[b]	0.15	8.50×10^{-31}	NR[c]
					rs11855019	0.19	8.60×10^{-25}	NR
					rs6497268	0.18	3.70×10^{-19}	NR
					rs7495174	0.10	2.00×10^{-22}	NR
BMD (lumbar spine)	rs6561055-A, rs17639156-G	*TNFSF11*	72	Previous	rs9533093	0.80	5.40×10^{-11}	1.22
					rs9594738	0.42[a]	4.00×10^{-23}	1.34
					rs9594759	0.62	1.50×10^{-17}	1.24
			73	Previous	rs9594759	0.49[a]	NR	NR
					rs9594738	0.42[a]	NR	NR
			74	Previous	rs9594759	0.63	1.10×10^{-16}	1.27
					rs10507507	0.82	1.60×10^{-5}	1.26
					rs7992970	0.78	8.50×10^{-7}	1.27
					rs9594738	0.56	2.00×10^{-21}	1.36
Crohn's disease	rs2066844-T, rs2066845-C, rs2066847-C	*NOD2*	75	Previous	rs2076756	0.35[a]	5.10×10^{-10}	NR
					rs2066843	0.36[a]	2.90×10^{-9}	NR
			76	Previous	rs2076756	0.24	7.00×10^{-14}	NR
			77	Previous	rs5743289	0.17	3.80×10^{-10}	1.45
					rs17221417	0.29	9.40×10^{-12}	1.29
			38	Subsequent	(Same: rs2066847)	0.02	3.00×10^{-24}	3.99
			78	Subsequent	rs2076756	0.26	9.70×10^{-8}	1.33
			79	Subsequent	rs2076756	0.35[a]	1.00×10^{-9}	NR
HDL cholesterol	rs9282541-T	*ABCA1*	46	Concurrent	rs3890182	0.87	3.00×10^{-10}	1.19[d]
			58	Concurrent	rs4149268	0.35	1.20×10^{-10}	1.10[d,e]
					rs4149274	0.69	7.40×10^{-8}	1.20[d,e]
			80	Concurrent	rs3890182	0.12	2.00×10^{-6}	5.58[d,e]
			81	Subsequent	rs3905000	0.86	8.60×10^{-13}	1.24[d]
					rs3847303	0.88	3.40×10^{-12}	1.25[d]
			44	Subsequent	rs1883025	0.26	1.00×10^{-9}	1.16[d]
			64	Subsequent	rs4149268	0.27[a]	0.69	1.20[d]
			82	Subsequent	rs2740491	0.36	3.10×10^{-4}	1.06[d]
					rs3847303	0.13	3.20×10^{-3}	1.06[d]

Hemoglobin	rs1800562-A	*HFE*	35	Concurrent	(Same)	0.04	1.60×10^{-4}	1.25[d]
			83	Concurrent	rs198833	0.08[a]	1.40×10^{-8}	NR
					rs129128	0.08[a]	3.30×10^{-8}	NR
					rs198851	0.08[a]	3.40×10^{-8}	NR
					rs1799945	0.14[a]	4.30×10^{-8}	NR
					rs198846	0.11[a]	8.70×10^{-6}	1.23[d]
LDL cholesterol	rs11591147-G	*PCSK9*	44	Subsequent	(Same)	0.02	9.00×10^{-6}	2.65[d]
					rs11206510	0.19	4.00×10^{-8}	1.17[d]
			64	Subsequent	(Same)	0.02	1.60×10^{-7}	1.88[d]
			58	Concurrent	rs11206510	0.81	3.50×10^{-11}	1.15[d,e]
			82	Subsequent	rs11206510	0.01	2.00×10^{-12}	1.33[d]
MCV	rs1800562-A	*HFE*	36	Concurrent	(Same)	0.04	1.00×10^{-46}	1.02[d]
			83	Concurrent	rs198846	0.11	8.60×10^{-13}	4.91[d]
Prostate cancer	rs16901979-A	Intergenic	34	Subsequent	(Same)	0.04	2.50×10^{-14}	1.80
Psoriasis	rs2395029-C	*HLA-C*	84	Previous	rs3134792	0.15[a]	1.00×10^{-9}	NR
			85	Subsequent	rs12191877	0.15	$<1.00 \times 10^{-100}$	2.64
Triglycerides	rs662799-G	*APOA1, APOC3, APOA4, APOA5*	86	Previous	rs481843	0.11	3.30×10^{-5}	NR
			46	Concurrent	rs28927680	0.07	2.00×10^{-17}	1.60[d]
			45	Concurrent	rs2075292	0.16	5.30×10^{-8}	1.10[d,e]
					rs7124741	0.17	8.60×10^{-7}	1.10[d,e]
					rs17120139	0.17	2.30×10^{-6}	1.09[d,e]
			80	Concurrent	rs6589566	0.06	3.00×10^{-11}	10.90[d,e]
			64	Subsequent	(Same)	0.06	2.90×10^{-15}	1.60[d]
					rs3135506	0.06	5.50×10^{-12}	1.57[d]
			81	Subsequent	rs12272004	0.93	5.40×10^{-13}	1.39[d]
					rs480878	0.86	8.00×10^{-9}	1.19[d]
					rs28927680	0.93	3.90×10^{-9}	1.64[d]
					rs12292921	0.07	9.06×10^{-13}	1.39[d]
					rs35120633	0.93	2.30×10^{-10}	1.75[d]
					rs3135506	0.06	7.40×10^{-10}	1.74[d]
					rs2075292	0.13	5.70×10^{-12}	1.23[d]
					rs588918	0.87	4.90×10^{-8}	1.19[d]
					rs1351452	0.86	7.40×10^{-10}	1.23[d]
			44	Subsequent	rs964184	0.14	4.00×10^{-62}	1.72[d]
			82	Subsequent	rs12292921	0.06	1.40×10^{-3}	1.20[d]

Table continues

Table 3. Continued

Disease/Trait	Uncommon Variant(s)	Gene Locus	Other GWAS That Found Variants in Same Locus, reference	Timing of Other GWAS	Variant	Risk Allele Frequency	*P* Value	Per-Allele Odds Ratio
Triglycerides	rs10892151-A	*APOA1, APOC3, APOA4, APOA5, DSCAML1*	44	Previous	rs964184	0.14	4.00×10^{-62}	1.72[d]
			46	Previous	rs28927680	0.07	2.00×10^{-17}	1.60[d]
			64	Previous	rs3135506	0.06	5.50×10^{-12}	1.57[d]
					rs662799	0.06	3.00×10^{-15}	1.60[d]
			58	Previous	rs12286037	0.94	1.00×10^{-26}	1.51[d,e]
			45	Previous	rs2075292	0.16	5.30×10^{-8}	1.10[d,e]
					rs7124741	0.17	8.60×10^{-7}	1.10[d,e]
					rs17120139	0.17	2.30×10^{-6}	1.09[d,e]
			86	Previous	rs481843	0.11	3.30×10^{-5}	NR
			81	Previous	rs12272004	0.93	5.40×10^{-13}	1.39[d]
					rs480878	0.86	8.00×10^{-9}	1.19[d]
					rs28927680	0.93	3.90×10^{-9}	1.64[d]
					rs12292921	0.07	9.00×10^{-13}	1.39[d]
					rs35120633	0.93	2.30×10^{-10}	1.75[d]
					rs3135506	0.06	7.40×10^{-10}	1.74[d]
					rs2075292	0.13	5.70×10^{-12}	1.23[d]
					rs588918	0.87	4.90×10^{-8}	1.19[d]
					rs1351452	0.86	7.40×10^{-10}	1.23[d]
			80	Previous	rs6589566	0.06	3.00×10^{-11}	10.90[d,e]
			82	Concurrent	rs12292921	0.06	1.40×10^{-3}	1.20[d]
Type 2 diabetes	rs7903146-T	*TCF7L2*	65	Previous	(Same)	0.25[a]	4.20×10^{-15}	1.43
					rs7901695	0.28[a]	8.30×10^{-13}	1.37
			66	Previous	(Same)	0.25[a]	3.00×10^{-23}	1.37
			67	Previous	(Same)	0.25[a]	5.50×10^{-8}	1.71
					rs7901695	0.28[a]	3.40×10^{-7}	1.66
					rs12255372	0.22[a]	5.30×10^{-7}	1.64
			69	Previous	(Same)	0.18	1.00×10^{-48}	1.37
			71	Previous	(Same)	0.29	1.50×10^{-34}	1.65
			68	Previous	(Same)	0.49	0.005[f]	1.28[f]
					rs7100927	0.49	0.007[f]	1.56[f]
			70	Subsequent	(Same)	0.27	1.20×10^{-30}	1.48
			77	Previous	rs4506565	0.32	5.70×10^{-13}	NR
			87	Previous	rs7901695	0.28[a]	1.00×10^{-48}	1.37

88	Previous	rs7100927	0.40[a]	0.007[f]	1.56[f]	
		rs10509966	0.25[a]	0.64[f]	1.09[f]	
		rs10509969	0.20[a]	0.60[f]	1.12[f]	
		rs290483	0.42[a]	0.93[f]	1.01[f]	
		rs7917983	0.39[a]	0.17[f]	1.22[f]	
		rs10509970	0.23[a]	0.51[f]	1.14[f]	
		rs10509967	0.26[a]	0.82[f]	1.04[f]	

Abbreviations: BMD, bone mineral density; GWAS, genome-wide association study(ies); HDL, high density lipoprotein; LDL, low density lipoprotein; MCV, mean corpuscular volume; NHANES, National Health and Nutrition Examination Survey; NR, not reported and data not adequate for computing the missing values.

[a] Risk allele frequency was retrieved from the International HapMap Project phases 1 + 2 data on the same ancestry populations as the eligible variants, because it was not reported in the GWAS.

[b] rs1667394 was reported to be located in the *HERC2* gene locus.

[c] The odds ratio was not retrievable based on the data given, but based on the *P* value and sample size of the eligible GWAS and of the GWAS reporting the same uncommon variant, we compared the 2 effect estimates and determined that the missing odds ratio is probably smaller than that of the eligible uncommon variant.

[d] The odds ratio equivalent was computed from the standardized mean difference.

[e] The odds ratio equivalent was computed from the mean difference by using also the population standard deviation from NHANES data on LDL cholesterol, HDL cholesterol, and triglyceride levels, because the population standard deviation was not given in the GWAS.

[f] *P* values and hazard ratios from Cox survival analysis.

RESULTS

Description of the eligible associations

We screened 440 GWAS with a total of 2,497 entries in the GWAS catalog. Of those, 74 GWAS were excluded because they reported no genome-wide significant ($P \leq 10^{-7}$) SNP–disease association. Of the remaining 366 studies listed in the catalog, we identified 91 entries with associations that had MAFs of 5% or less. We excluded 61 entries because they were not significant at the $P \leq 10^{-7}$ level, 4 because they had MAFs of greater than 5% upon scrutinizing the respective article, and another 3 because the respective associations were based on haplotypes. Of the remaining 23 associations, 1 (rs16901979 and prostate cancer) had been identified by 2 different GWAS and, thus, we regarded as eligible the one published earlier (33) and the subsequent study (34) as a replication. Thus, 22 different associations discovered in 18 different GWAS were eligible through the catalog search.

The main articles and the supplements of those 366 GWAS reporting at least 1 association significant at the $P \leq 10^{-7}$ level were further scrutinized for uncommon/rare variants with genome-wide significance. Hence, we identified 23 additional SNP–disease associations with genome-wide significance ($P \leq 10^{-7}$) implicating uncommon/rare SNPs (MAF, $\leq$5%), of which 1 (rs1800562 and mean corpuscular volume) had been reported by 2 GWAS published at the same time; thus, we regarded as eligible one of them (35), and the other study (36) was recorded as concurrent. Hence, a total of 44 associations were identified by combing the catalog-based and the full text-based searches. Of those associations, 1 (rs2066847 and Crohn's disease) had been discovered by 2 different GWAS, of which the 1 published earlier (37) was included in our analysis and the subsequent was recorded as a replication (38). Finally, 43 different genome-wide significant associations implicating 40 uncommon/rare SNPs discovered in 28 GWAS (33, 35–61) were eligible (Table 1). One uncommon SNP was implicated in 2 different phenotypes and another in 3 different phenotypes. Among these 40 SNPs, the authors of the respective GWAS implicated a single gene for 31 cases; for 4 SNPs, they implicated more than 1 gene; for 1 SNP, they implicated a single gene in 1 GWAS and more genes in another; and 4 SNPs were not allocated to any specific gene. Overall, 30 different locus–phenotype pairs were implicated (some had been implicated for $\geq$1 SNP).

The phenotypes for these 43 associations were acquired immunodeficiency syndrome (AIDS) progression ($n = 2$ associations), bone mineral density ($n = 6$ associations in 2 loci), Crohn's disease ($n = 4$), high density lipoprotein (HDL) cholesterol ($n = 2$), nasopharyngeal carcinoma ($n = 2$ associations in the same locus), panic disorder ($n = 2$), response to antipsychotic therapy ($n = 2$), systemic lupus erythematosus ($n = 2$ associations in the same locus), triglyceride levels ($n = 2$ associations in the same locus), acute lymphoblastic leukemia in children, eye color, cognitive performance, freckles, low density lipoprotein (LDL) cholesterol, hematocrit levels, hemoglobin levels, mean corpuscular hemoglobin, mean corpuscular volume, nonsyndromic

Table 4. Number of Candidate–Gene Association Studies on Each Gene Locus–Disease Association (per HuGE Navigator) and Comments Regarding Previous Knowledge on the Loci Containing the Uncommon Variants as They Appear in the Eligible GWAS

Disease/Trait	Reported Gene(s)	Uncommon Variant(s)	No. of Studies Until December of the Year Before the First GWAS Proposal	Total No. of Studies on Gene–Phenotype to Date	Comments on Gene Locus in Text
ALL	*KRTHB5*	rs2089222-A	0	0	No comment
AIDS progression	*HCP5, MICB, MCCD1, BAT1, LTB, TNF*	rs2395029-G	1 for *HCP*, 1 for *TNF*, 0 for the rest	1	*HCP5* was previously identified by the GWAS-based Euro-CHAVI cohort and also proposed by candidate–gene association studies
	C6orf48	rs9368699-C	0	0	No comment
Blue vs. green eyes	*OCA2*	rs1667394-A	0	1	Previously reported to be associated with albinism, eye color, hair color, and skin pigmentation; *OCA2* mutations are known to be a major cause for albinism; *OCA2* has been discovered in linkage studies.
Freckles	*MC1R*	rs1805007-T	0	3	It was known by previous reports; previously documented mutations in *MC1R*
BMD (lumbar spine)	*AKAP11*	rs180851-G, rs7326472-G, rs12854504-G, rs7998154-T	0	0	No comment
	TNFSF11	rs6561055-A, rs17639156-G	3	11	No comment
Cognitive performance	*HCCS*	rs5934953-C	0	0	No comment
Crohn's disease	*NOD2*	rs2066844-T, rs2066845-C, rs2066847-C	176	301	*NOD2* is a previously known Crohn's disease locus.
	LRRK2, MUC19	rs11175593-T	1 for *MUC19*, 0 for *LRRK2*	1 for *MUC19*, 0 for *LRRK2*	*LRRK2*: evidence from a previous cell study; *MUC19*: evidence from a previous animal study
HDL cholesterol	*HNF4A*	rs1800961-C	3	5	Function in humans has previously been studied; although mice lacking either *Hnf4a* or *Hnf1a* have altered plasma cholesterol levels, there has been only modest evidence to date connecting these genes to either HDL or LDL cholesterol concentrations in humans.
	ABCA1	rs9282541-T	34	53	It is a well-recognized association.

Table continues

Table 4. Continued

Disease/Trait	Reported Gene(s)	Uncommon Variant(s)	No. of Studies Until December of the Year Before the First GWAS Proposal	Total No. of Studies on Gene–Phenotype to Date	Comments on Gene Locus in Text
Hematocrit	*HFE*	rs1800562-A	3	4	Mutations in the *HFE* gene are already known to underlie hereditary hemochromatosis. The *HFE* gene induces expression of the iron-regulatory hormone hepcidin.
Hemoglobin	*HFE*	rs1800562-A	17	21	Mutations in the *HFE* gene are already known to underlie hereditary hemochromatosis. The *HFE* gene induces expression of the iron-regulatory hormone hepcidin.
LDL cholesterol	*PCSK9*	rs11591147-G	10	26	Prior evidence for association with LDL cholesterol concentrations; has also been shown to cause Mendelian syndromes or to harbor multiple rare alleles that contribute to trait variation
MCH	*SLC17A3*	rs1408272-G	0	0	No comment
MCV	*HFE*	rs1800562-A	3	5	*HFE* is known to be associated with iron homeostasis.
NSCL	Intergenic	rs17085106-T	N/A	N/A	
Nasopharyngeal carcinoma	*ITGA9*	rs189897-A	0	1	The gene is located at the chromosomal 3p22-21.3 segment, which is known to be commonly deleted in various types of carcinoma including NPC. A linkage study also mapped an NPC susceptibility locus to chromosome 3p21.31-21.2, indicating that the genes in this region are crucial for the formation of NPC.
		rs197757-T	0	1	
Panic disorder	*TMEM16B*[a]	rs12579350-A	0	1	No comment
	PKP1	rs860554-T	0	1	The gene has an important role in the cytoskeleton–cell membrane interaction. The protein of *PKP1*, plackoglobin, acts as linker molecules at adherence junctions and desmosome at the plasma membrane.
Prostate cancer	Intergenic	rs16901979-A	N/A	N/A	

Table continues

Table 4. Continued

Disease/Trait	Reported Gene(s)	Uncommon Variant(s)	No. of Studies Until December of the Year Before the First GWAS Proposal	Total No. of Studies on Gene–Phenotype to Date	Comments on Gene Locus in Text
Primary biliary cirrhosis	*C6orf10*	rs2395148-A	0	0	No comment
ALP	*PDZRN4, CNTN1*	rs1880887-C	0	1	No comment (locus found only in supplement)
fT3	*HS3ST3B1*	rs3848445-C	0	1	No comment (locus found only in supplement)
Psoriasis	*HLA-C*	rs2395029-C	57	65	Strongest association with this region is consistent with previous results from our group and others.
Response to treatment for ALL	*ST8SIA6*	rs359312-T	0	0	No comment
Response to antipsychotic therapy	Intergenic	rs17022444-G	N/A	N/A	
	Intergenic	rs7669317-C	N/A	N/A	
SLE	*TNFAIP3*	rs5029939-G	0	10	Previously unreported for SLE susceptibility; recent reports for influencing rheumatoid arthritis risk. This GWAS identifies *TNFAIP3* as a new susceptibility locus in SLE.
SLE	*TNFAIP3*	rs2230926-C	0	10	Reported by previous GWAS
Tanning	*MATP*	rs35391-T	0	0	SNPs in *MATP* were previously evaluated in the GWAS of natural hair color by our group. Three SNPs in the *MATP* gene have been associated with human pigmentation.
Triglycerides	*APOA1, APOC3, APOA4, APOA5*	rs662799-G	118 for *APOA1, APOC3, APOA4, APOA5*	165 combined for *APOA1, APOC3, APOA4, APOA5*	These loci have been previously implicated in lipid metabolism.
Triglycerides	*APOA1, APOC3, APOA4, APOA5, DSCAML1*	rs10892151-A	118 for *APOA1, APOC3, APOA4, APOA5*	165 combined for *APOA1, APOC3, APOA4, APOA5*	*APOA1, APOC3, APOA4, APOA5* is a cluster of more likely candidate genes, given the established key roles of their products in lipid metabolism.
Type 1 diabetes	*COBL*	rs4948088-C	0	0	No comment
Type 2 diabetes	*TCF7L2*	rs7903146-T	14	140	It was reported by previous studies.

Abbreviations: AIDS, acquired immunodeficiency syndrome; ALL, acute lymphoblastic leukemia; ALP, alkaline phosphatase; BMD, bone mineral density; fT3, free triiodothyronine; GWAS, genome-wide association study(ies); HDL, high density lipoprotein; HuGE, Human Genome Epidemiology; LDL, low density lipoprotein; MCH, mean corpuscular hemoglobin; MCV, mean corpuscular volume; N/A, nonapplicable because the variants are in intergenic regions; NPC, nasopharyngeal carcinoma; NSCL, nonsyndromic cleft lip with or without cleft palate; SLE, systemic lupus erythematosus; SNP, single-nucleotide polymorphism.

[a] *TMEM16B* was found as *ANO2* in HuGE Navigator.

cleft lip with or without cleft palate, prostate cancer, alkaline phosphatase, free triiodothyronine, primary billiary cirrhosis, psoriasis, response to treatment for childhood acute lymphoblastic leukemia, tanning, type 1 diabetes, and type 2 diabetes.

Location and function of gene variants

Nine of the 40 uncommon variants (22.5%) constituted nonsynonymous coding SNPs, whereas 15 (37.5%) were intronic, 10 (25%) were intergenic (although 6 of them were related to specific genes by the authors of the GWAS), 1 was located in the 5′-untranslated region (5′-UTR), 1 was found upstream of the respective gene, 1 interfered with the function of the frameshift, and for 3 SNPs the function/location was unknown.

Frequency and effect sizes

All 40 variants had MAFs that would characterize them as uncommon rather than rare. Of the 22 associations pertaining to diseases rather than quantitative traits or nondisease-related phenotypes, 21 had risk variants with a risk allele frequency of 5% or less, and only 1 association had a risk allele frequency of 95%. The latter was actually the association with the smallest odds ratio estimate. Thirty-three associations had been discovered and replicated exclusively in populations of European ancestry, whereas 10 were discovered and/or replicated in non-European or mixed populations.

Eleven of the 43 associations had P values between 10^{-7} and 10^{-8}, and 32 had greater statistical significance. Odds ratios were extracted, obtained from the authors, or calculated in 36 associations (no data were retrievable for 7 associations). Per-allele odds ratios ranged from 1.03 (for rs1408272 contributing to mean corpuscular hemoglobin levels) to 22.11 (for rs12579350 in panic disorder). The median was 2.24 (interquartile range, 1.40–3.40).

Power calculations and observed and expected distributions of uncommon variants

The average sample size utilized in the 28 identified GWAS was 7,637 individuals for case-control studies and 10,647 individuals for all studies (case-control and cohort). Table 2 shows the number of discovered associations implicating uncommon variants split according to odds ratio quartiles and according to MAFs = 1%–2%, 3%–4%, and 5% categories. As shown, no variant with an odds ratio of less than 1.40 and a MAF = 1%–2% is included, because the power to detect such variants with the typical sample size used in these GWAS in minimal (0.37%). Power calculations suggest that only 11% and 23% of the variants with similar odds ratio and a MAF = 3%–4% or 5%, respectively, would have been discovered with the average sample size of the GWAS that we considered. Variants with an odds ratio = 1.40–2.24 and a MAF = 1%–2% had a 56% chance to be discovered. In all other categories of odds ratio and MAF combinations, the power is greater than 99%. This means that, with a sample size of 10,647, it should be possible to discover almost all variants with an odds ratio greater than 1.40 and MAF = 3%–5% and those with an odds ratio greater than 3 and a MAF greater than 1%. Consideration of the power calculations suggests that the number of variants with an odds ratio less than 1.40 and a MAF = 3%–5% may be 3-fold larger than that with an odds ratio greater than 1.40 and a similar MAF, but the latter variants are far easier to discover with the typical sample size used in these GWAS.

Variants in the same loci in other GWAS

For 37 of the 43 associations, we identified at least 1 other GWAS on the same phenotype (Web Table 1). (This information is described in a supplementary table posted on the *Journal*'s website (http://aje.oxfordjournals.org/).) No other GWAS was found for 6 associations (freckles, panic disorder ($n = 2$ associations), primary biliary cirrhosis, free triiodothyronine, and response to treatment for acute lymphoblastic leukemia).

For 15 associations, additional GWAS had presented data on the same uncommon SNP ($n = 16$) (34–36, 38, 44, 62–71) and/or other SNPs in the same locus ($n = 74$ associations) (44–46, 58, 62–65, 67, 68, 72–88) (Table 3). For 1 association (prostate cancer and rs16901979), no other polymorphisms except the same uncommon variant were identified; hence, for 14 uncommon variant–phenotype associations (corresponding to 10 gene locus–phenotype associations), other GWAS discovered 1 or more common SNPs at the same locus with the uncommon variant. For 4 associations (eye color, LDL cholesterol, triglycerides, type 2 diabetes), the same additional GWAS had presented data on both the same uncommon/rare SNP and 1 or more other SNPs (44, 62–65, 67, 68).

Whenever the same uncommon SNPs were identified by additional GWAS (8 uncommon SNPs in 16 additional GWAS), the odds ratio estimates were larger than those proposed by the first study with genome-wide significance in 6 cases and smaller in 10 cases. Twelve of the 16 estimates were genome-wide significant. All 16 were nominally significant ($P < 0.05$).

When other GWAS had presented other SNPs in the same locus, almost all (72/74) of the additional SNPs were common (MAF, >5%). The odds ratio per risk allele was smaller than the effect size of the index uncommon variant with 14 exceptions. Fifty of these 74 additional associations had reached levels of genome-wide significance, and 65 were nominally significant ($P < 0.05$), whereas for 2 associations the exact P value was not reported.

Evaluation of these variants in SNAP showed that the 2 uncommon variants in *TNFAIP3* that were associated with systemic lupus erythematosus in 2 different GWAS were in high linkage disequilibrium ($r^2 = 1$ and $D' = 1$ in both Europeans and Asians). Furthermore, the bone mineral density-associated uncommon SNP rs180851 was in high linkage disequilibrium with the uncommon SNPs rs7326472 and rs12854504 ($r^2 = 0.82$ and $D' = 1$ for pairwise comparison), which were discovered in the same GWAS. Also in the same GWAS, the uncommon SNP-pair rs7326472 and rs12854504, as well as the SNP-pair

Table 5. Mutations in the Same Gene Loci With the Uncommon Variants Causing Related Phenotypic Effects, as Found in the Online Mendelian Inheritance in Man Database

Reported Gene(s)	Region	Uncommon Variant(s)	Disease/Trait	Mutations With the Same or Related Phenotypic Effects	Phenotypic Effects of Mutations
OCA2	15q13.1	rs1667394-A	Blue vs. green eyes	2.7-kb del, ex7del	Oculocutaneous albinism type 2
				IVS17DS, G-T, +1	
				Pro743Leu	
				1-bp del	
				Ala334Val	
				122.5-kb del	
				Trp679Cys	
				Asn489Asp	
				Met394Ile	
TNFSF11	13q14	rs6561055-A, rs17639156-G	BMD	5-bp del, IVS7+4	Osteopetrosis, autosomal recessive type 2
				Met199Lys	
				2-bp del, 828CG	
NOD2	16q12.1	rs2066844-T, rs2066845-C, rs2066847-C	Crohn's disease	3020insC	Crohn's disease
				Gly881Arg	
				Arg675Trp	
				IVS8+158	
MC1R	16q24.3	rs1805007-T	Freckles	3-bp del 439TTC	UV-induced skin damage
				Thr157Ile	
				Pro159Thr	
HNF4A	20q13.12	rs1800961-C	HDL cholesterol	Gln268Ter	Maturity-onset diabetes of the young, type 1
				Arg154Ter	
				Arg127Trp	
				1-bp del Phe75T	
				IVS5, Del A, -2	
				Met364Arg	
				Val393Ile	Type 2 Diabetes
ABCA1	9q31.1	rs9282541-T	HDL cholesterol	Cys1417Arg	Tangier disease (HDL deficiency type 1)
				IVS24DS, G-C	
				Gln537Arg	
				110-bp in/14-bp del	
				2-bp del, 3283TC	
				1-bp del, 2665C	
				Ser1446Leu	
				int12-14 del, int16-31 del	
				Arg1680Trp	
				Asp1229Asn	
				Arg2021Trp	
				1-bp del, 1764G	
				Asn875Ser	
				Ala877Val	
				Trp530Ser	
				1-bp del, 1764G	
				Tyr573Ter	
				3-bp del	HDL deficiency type 2

Table continues

Table 5. Continued

Reported Gene(s)	Region	Uncommon Variant(s)	Disease/Trait	Mutations With the Same or Related Phenotypic Effects	Phenotypic Effects of Mutations
HFE	6p22	rs1800562-A	Hematocrit, Hemoglobin, MCV	Cys282Tyr	Hemochromatosis
				His63Asp	
				Arg330Met	
				Gln283Pro	
PCSK9	1p32.3	rs11591147-G	LDL cholesterol	Asp374Tyr	Familial hypercholesterolemia, type 3
				Tyr142Ter	LDL cholesterol level quantitative trait locus 1
				Cys679Ter	
				3-bp del 290_292delGCC	
HLA-C	6p21.33	rs2395029-C	Psoriasis	HLA-C, HLA-Cw6 allele	Psoriasis
MATP	5p13.3	rs35391-T	Tanning	IVS2, G-A, -1	Oculocutaneous albinism type 4
				1-bp del, 986C	
				3-bp del	
				Ala486Val	
				Asp157Asn	
				1-bp del, 1121T	
APOA1, APOC3, APOA4, APOA5	11q23.3	rs662799-G, rs10892151-A	Triglycerides	Gln84Ter (*APOA1/APOC3*)	Apolipoprotein A-I deficiency
				Val156Glu (*APOA1/APOC3*)	
				Gln-2Ter (*APOA1/APOC3*)	Analphalipoproteinemia
				1-bp ins (*APOA1/APOC3*)	Primary hypoalphalipoproteinemia
				Gln32Ter (*APOA1/APOC3*)	Periorbital xanthelasma
				Gln139Ter (*APOA5*)	Hyperlipoproteinemia type 4

Abbreviations: BMD, bone mineral density; HDL, high density lipoprotein; kb, kilobase(s); LDL, low density lipoprotein; MCV, mean corpuscular volume; SNP, single-nucleotide polymorphism; UV, ultraviolet.

rs6561055 and rs17639156, were in high linkage disequilibrium ($r^2 = 1$ and $D' = 1$). Moreover, the type 2 diabetes susceptibility uncommon variant rs7903146 located in *TCF7L2* was in linkage disequilibrium with rs7901695 ($r^2 = 1$ and $D' = 1$) and rs4506565 ($r^2 = 1$ and $D' = 1$), which have been highlighted by 3 and 1 previous GWAS, respectively. Both rs7901695 and rs4506565 are common in the European populations used in these GWAS but not in Japanese populations where rs7903146 reached genome-wide significance. None of the other SNPs in the same genetic loci as the uncommon variants had high linkage disequilibrium with them based on the r^2. Besides these associations that had both $D' = 1$ and $r^2 = 1$, another 41 pairs of uncommon-other SNPs had $D' = 1$ but not $r^2 = 1$.

Prior literature

In HuGE Navigator, we identified 2 prior studies for the association between AIDS progression and the *HCP-TNF* gene locus; 3 studies for the association between *TNFSF11* and bone mineral density; 176 studies for the association between Crohn's disease and *NOD2*; 1 study for the association between Crohn's disease and *MUC19*; 3 studies for HDL cholesterol levels and *HNF4A*; 34 studies for the association between *ABCA1* and HDL cholesterol; 3 studies for the association between *HFE* and hematocrit; 17 studies for the association between *HFE* and hemoglobin; 10 studies for LDL cholesterol levels and *PCSK9*; 3 studies for the association between *HFE* and mean corpuscular volume; 57 studies for psoriasis and *HLA-C*; 118 studies for triglyceride levels and any gene in the *APOA1-APOC3-APOA4-APOA5* complex, and 14 studies for type 2 diabetes and *TCF7L2*. Results are summarized in Table 4 along with the total number of studies on each locus published to date.

On the basis of the comments of the GWAS authors (Table 4), several of the loci of discovered uncommon variants had some evidence support from prior studies, although not necessarily gene–disease association studies on human populations.

Known mutations in the same gene loci

According to the Online Mendelian Inheritance in Man Database, for 11 gene loci (implicated in a total of 13 gene locus–phenotype associations) where uncommon variants

Table 6. Minor Allele Frequencies of the Eligible Uncommon Variants in the 4 HapMap Phases 1 + 2 Populations

Strongest SNP-Risk Allele	Reported Gene(s)	Region	MAF in GWAS	GWAS Population	MAF in CEU HapMap 1 + 2	MAF in CHB HapMap 1 + 2	MAF in JPT HapMap 1 + 2	MAF in YRI HapMap 1 + 2
rs2089222-A	*KRTHB5*	12q24.22	0.03	European	0.04	0.26	0.33	0.19
rs2395029-G	*HCP5, MICB, MCCD1, BAT1, LTB, TNF*	6p21.33	0.03	European	0.05	0.01	0	0
	HLA-C		0.03	European				
rs9368699-C	*C6orf48*	6p21.3	0.03	European	0.06	0.18	0.10	0
rs1667394-A	*OCA2*	15q13.1	0.02	European	0.13	0.20	0.14	0.05
rs1805007-T	*MC1R*	16q24.3	0.05	European	0.15	0	0	0
rs5934953-C	*HCCS*	Xp22.2	0.02	European	0.04	0	0	0
rs180851-G	*AKAP11*	13q14	0.05	European	0.05	0.08	0.03	0.14
rs7326472-G	*AKAP11*	13q14	0.05	European	0.04	0.11	0.09	0.16
rs12854504-G	*AKAP11*	13q14	0.05	European	0.04	0.10	0.07	0.03
rs7998154-T	*AKAP11*	13q14	0.02	European	0.02	0	0	0
rs6561055-A	*TNFSF11*	13q14	0.05	European	0.04	0	0	0
rs17639156-G	*TNFSF11*	13q14	0.05	European	0.04	0.07	0.09	0
rs2066844-T	*NOD2*	16q12.1	0.05	European	0.11	0	0	0
rs2066845-C	*NOD2*	16q12.1	0.01	European	0.02	0	0	0
rs2066847-C	*NOD2*	16q12.1	0.04	European	0	0	0	0
rs11175593-T	*LRRK2, MUC19*	12q12	0.02	European	0.02	0.03	0.01	0
rs1800961-C	*HNF4A*	20q13.12	0.03	European	0.05	0.01	0	0
rs9282541-T	*ABCA1*	9q31.1	0.03	Mixed	0	0	0	0
rs1800562-A	*HFE*	6p22.1	0.04	European	0.04	0	0	0
			0.04	European				
			0.04	European				
rs11591147-G	*PCSK9*	1p32.3	0.01	European	0	0	0	0
rs1408272-G	*SLC17A3*	6p22.1	0.03	European	0.03	0	0	0
rs17085106-T	Intergenic	18q22.3	0.02	European	0	0	0	0.16
rs189897-A	*ITGA9*	3p22.2	0.03	Asian	0.27	0.08	0.12	0.01
rs197757-T	*ITGA9*	3p22.2	0.03	Asian	0	0.07	0.17	0.02
rs12579350-A	*TMEM16B*	12p13.31	0.01	Asian	0	0.01	0	0
rs860554-T	*PKP1*	1q32.1	0.05	Asian	0.20	0.12	0.06	0.01
rs16901979-A	Intergenic	8q24.21	0.03	European	0.02	0.29	0.16	0.46
rs2395148-A	*C6orf10*	6p21.3	0.02	European	0.05	0.21	0.07	0.09
rs1880887-C	*PDZRN4, CNTN1*	12q12	0.03	European	0.03	0.14	0.08	0.38
rs3848445-C	*HS3ST3B1*	17p12	0.05	European	0.05	0.32	0.26	0.17
rs359312-T	*ST8SIA6*	10p12.33	0.04	European, African, other	0	0.47	0.42	0
rs17022444-G	Intergenic	2p12	0.03	European, African, other	0	0	0	0.07

rs7669317-C	Intergenic	4q24	0.04	European, African, other	0.05	0	0	0	
rs5029939-G	*TNFAIP3*	6q23.3	0.03	European	0.04	0.09	0.20	0.50	
rs2230926-C	*TNFAIP3*	6q23.3	0.04	Asian	0.01	0.09	0.18	0.48	
rs35391-T	*MATP*	5p13.3	0.03	European	0	0.38	0.35	0.47	
rs662799-G	*APOA1, APOC3, APOA4, APOA5*	11q23.3	0.05	European	0.02	0.27	0.29	0.13	
rs10892151-A	*APOA1, APOC3, APOA4, APOA5, DSCAML1*	11q23.3	0.03	European	0.02	0.06	0	0.41	
rs4948088-C	*COBL*	7p12.1	0.05	European	0.02	0	0	0.04	
rs7903146-T	*TCF7L2*	10q25.2	0.04	Asian	0.25	0.02	0.02	0.29	

Abbreviations: CEU, Utah residents with Northern and Western European ancestry from the CEPH (Centre de'Etude du Polymorphism Humain) collection; CHB, Han Chinese in Beijing, China; GWAS, genome-wide association study(ies); HapMap, International HapMap Project; JPT, Japanese in Tokyo, Japan; MAF, minor allele frequency; SNP, single-nucleotide polymorphism; YRI, Yoruba in Ibadan, Nigeria.

had been identified by GWAS (*OCA2* and eye color; *TNFSF11* and bone mineral density; *NOD2* and Crohn's disease; *MC1R* and freckles; *HNF4A* and HDL cholesterol levels; *ABCA1* and HDL cholesterol; *HFE* and hematocrit, hemoglobin, and mean corpuscular volume; *PCSK9* and LDL cholesterol levels; *MATP* and tanning; *HLA-C* and psoriasis; and *APOA1/C3/A4/A5* and triglycerides), there were known mutations conferring the same or related phenotypic effects (Table 5).

In 5 loci (*HCCS*, *LRRK2*, *PKP1*, *CNTN1*, *HLA-C*), mutations had been described with phenotypic effects (syndromic micropthalmia, Parkinson's disease, ectodermal dysplasia/skin fragility syndrome, Compton-North myopathy, human immunodeficiency virus, type 1 (HIV-1), viremia, respectively) that were not similar to those implicated in the GWAS-identified uncommon variants.

Confluence of common SNPs, prior candidate variants, or mutations in loci with uncommon variants discovered in GWAS

Overall, GWAS have discovered 30 different gene locus–phenotype associations involving uncommon variants where a single or multiple genes have been implicated. Of those, for 16 associations other common SNPs have been described by GWAS ($n = 10$), variants have been proposed by candidate gene studies prior to the first GWAS proposing the respective locus ($n = 13$), or mutations conferring similar or related phenotypes have been described ($n = 13$). For 4 of the 16 locus–phenotype associations, 2 of the 3 statements hold true, and for another 8 all 3 statements hold true.

For the remaining 14 gene locus–phenotype associations (*KRTHB5* and acute lymphoblastic leukemia, *C6orf48* and AIDS progression, *AKAP11* and bone mineral density, *HCCS* and cognitive performance, *SLC17A3* and mean corpuscular hemoglobin, *ITGA9* and nasopharyngeal carcinoma, *TMEM16B* and panic disorder, *PKP1* and panic disorder, *C6orf10* and primary biliary cirrhosis, *PDZRN4/CNTN1* and alkaline phosphatase, *HS3ST3B1* and free triiodothyronine, *ST8SIA6* and response to treatment for acute lymphoblastic leukemia, *TNFAIP3* and systemic lupus erythematosus, *COBL* and type 1 diabetes), we did not identify common SNPs in the same locus with the uncommon variants, prior candidate-gene association studies, or mutations with a similar/related phenotypic effect.

Allele frequencies in populations of different ancestry

Three variants (rs11591147-T, rs9282541-T, and rs2066847-C) were not found in any of the 4 HapMap samples. Another 12 variants were uncommon in all 4 HapMap samples (Table 6). Therefore, 25 of the 40 variants were common in at least 1 HapMap sample.

DISCUSSION

Here, we systematically evaluated the characteristics of variants with a MAF of 5% or less that have reached levels of genome-wide significance ($P \leq 10^{-7}$) in GWAS. We identified 43 eligible SNP–disease associations, in 12 of

which the implicated SNPs (9 in total) were exonic. Most were discovered and replicated in populations of European descent. The effect sizes were typically large. Some of these variants were identified in more than 1 GWAS on the same phenotype and, for 14 uncommon variant–phenotype associations (corresponding to 10 gene locus–phenotype associations), GWAS had also identified common variants for the same phenotype. Eleven loci implicated in 13 different locus–phenotype associations also had some evidence support from prior studies. Additionally, for 11 loci implicated in a total of 13 locus–phenotype associations, there was evidence for mutations conferring the same or related phenotypic effects. Most of the eligible uncommon SNPs would be common in at least 1 HapMap sample.

There are considerable debate and some preliminary evidence regarding the ''rare variant–common disease'' model of susceptibility to many complex diseases such as cancer, diabetes, and lupus (7, 89–95). According to this hypothesis, the multiplicative action of uncommon (13, 90) and rare (13) variants with modest and high odds ratios may explain a significant fraction of genetic variance in many common traits (89–91). In almost all the eligible associations that we overviewed that pertained to diseases, the risk allele had a frequency of 5% or less rather than 95% or greater. The only exception was a *COBL* variant apparently conferring susceptibility to type 1 diabetes, where the effect size was atypically small and the statistical support was among the weakest. Uncommon risk alleles may have an evolutionary disadvantage, and this does not allow them to become more prevalent in the population. They may also tend to be more recent, even if their effects are evolutionary neutral. Additionally, most of the associations in our study had odds ratios above 2, which is the usual odds ratio expected for associations involving uncommon variants (89–92, 94). However, odds ratios exceeding by far the small effect sizes typical of most GWAS-identified common variants (7) do not necessarily prove that uncommon variants routinely should always have such large effects. Because of power considerations, current studies are expected to identify predominantly those uncommon variants that have the largest effects (4, 13, 90). This is also supported by our analysis, which showed that the average sample sizes of most GWAS conducted to date are insufficient to detect the majority of uncommon SNPs with an odds ratio of less than 1.40. There are likely to be far more associations of uncommon variants with modest effects rather than large effects in the genetic architecture of complex traits. The majority of associations in the latter group have probably already been discovered, especially when large sample sizes have been amassed in GWAS.

Although uncommon and rare variants may constitute about 60% of variation in the human genome (90, 96), they are poorly covered in GWAS (8, 91, 97) and are often excluded from GWAS analyses by default, since a MAF threshold of 1% or greater or even 5% is often adopted as a quality control criterion by GWAS conducted to date. This may also explain the fact that all the SNPs that we identified were uncommon rather than rare; that is, they have a MAF = 0.5%–5%. Indeed, in our study, only a small minority of the variants indexed in the *Catalog of Published Genome-Wide Association Studies* had a MAF of 5% or less, and an even smaller minority were genome-wide significant. Detection of uncommon variants requires sample sizes (4, 98) much larger than those of most GWAS conducted to date (13). The situation may improve with much larger studies (99) or meta-analysis of multiple GWAS (100).

The finding that most of the uncommon variants in this overview were detected in populations of European ancestry simply reflects the fact that most GWAS have been conducted to date in these ethnic groups (101). As we have shown, relatively few of the identified uncommon variants are uncommon across all different ancestry groups. Conversely, several of the discovered common variants in GWAS are uncommon in other ancestry groups (102). Hence, investigating loci in other ethnicities that are statistically significantly associated with traits in 1 ethnicity may be a mechanism for discovering further associated rare variants.

Finally, we have identified several gene loci that contain both uncommon and common variants with genome-wide significance. The effect estimates of the uncommon variants were generally larger than the effects of the common variants. This supports the hypothesis that genes containing common variants with modest effects on common traits may also contain uncommon variants with much larger effects (13). Alternatively, uncommon and rare variants may create ''synthetic associations'' by occurring, stochastically, more often in association with one of the alleles at a common SNP site (103). However, we found few examples where common and uncommon variants had high linkage disequilibrium. Furthermore, some of these same loci carry known mutations causing related traits. Overall, this picture is more consistent with a confluence of rare, uncommon and common genetic variation on the same genetic loci, perhaps conferring independent effects in shaping complex traits (14).

Our study has some limitations. First, the number of the eligible associations is still limited. Second, the MAF of a specific allele may differ significantly between different studies, depending on the populations studied; thus, the same allele may be characterized as uncommon in 1 population and as common in another (104). The emergence of mature data from the 1,000 Genomes Project should give better accuracy in allele frequencies and a better characterization of rare/uncommon variants than is currently possible (105, 106). Third, we did not have data on the examined variants from all agnostic GWAS done on the same phenotype, since for some of them their effect estimate, *P* values, and MAFs were not retrievable. Effect sizes may be smaller than what we observed based on published data that may suffer to some extent from winner's curse (107–109).

The number of associations with uncommon/rare variants discovered in agnostic genotyping methods is expected to rise with new technologies for whole genome or exome sequencing (16–18). A current debate is whether focusing on exons rather than sequencing the whole genome may suffice for identifying a large share of the missing genetic dark matter. On the basis of our series, exons may include only a minority of these uncommon variants and, thus, full genome sequencing may be unavoidable for successful identification of most variants of interest. Moreover, given technical and power considerations, GWAS to date have not been able to tell us

anything about the rare variants with a MAF less than 0.5%. Even with newer technologies, these will be captured only if they confer extremely large causal effects.

ACKNOWLEDGMENTS

Author affiliations: Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece (Orestis A. Panagiotou, Evangelos Evangelou, John P. A. Ioannidis); Tufts Medical Center and Tufts University School of Medicine, Boston, Massachusetts (John P. A. Ioannidis); Stanford Prevention Research Center, Stanford, California (John P. A. Ioannidis); and Harvard School of Public Health, Boston, Massachusetts (John P. A. Ioannidis).

Conflict of interest: none declared.

REFERENCES

1. Lin BK, Clyne M, Walsh M, et al. Tracking the epidemiology of human genes in the literature: the HuGE Published Literature database. *Am J Epidemiol*. 2006;164(1):1–4.
2. Yu W, Gwinn M, Clyne M, et al. A navigator for human genome epidemiology. *Nat Genet.* 2008;40(2):124–125.
3. Vineis P, Brennan P, Canzian F, et al. Expectations and challenges stemming from genome-wide association studies. *Mutagenesis*. 2008;23(6):439–444.
4. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet.* 2008;9(5):356–369.
5. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science*. 2008;322(5903):881–888.
6. Ioannidis JPA, Thomas G, Daly MJ. Validating, augmenting and refining genome-wide association signals. *Nat Rev Genet.* 2009;10(5):318–329.
7. Campbell H, Manolio T. Commentary: rare alleles, modest genetic effects and the need for collaboration. *Int J Epidemiol*. 2007;36(2):445–448.
8. Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. *Nat Genet.* 2006;38(6):659–662.
9. Zondervan KT, Cardon LR. The complex interplay among factors that influence allelic association. *Nat Rev Genet.* 2004;5(2):89–100.
10. Anderson CA, Pettersson FH, Barrett JC, et al. Evaluating the effects of imputation on the power, coverage, and cost efficiency of genome-wide SNP platforms. *Am J Hum Genet.* 2008;83(1):112–119.
11. Goldstein DB. Common genetic variation and human traits. *N Engl J Med*. 2009;360(17):1696–1698.
12. Ioannidis JPA. Prediction of cardiovascular disease outcomes and established cardiovascular risk factors by genome-wide association markers. *Circ Cardiovasc Genet.* 2009;2(1):7–15.
13. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461(7265):747–753.
14. Galvan A, Ioannidis JPA, Dragani TA. Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer. *Trends Genet.* 2010;26(3):132–141.
15. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet.* 2008;83(3):311–321.
16. Ng SB, Turner EH, Robertson PD, et al. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 2009;461(7261):272–276.
17. Choi M, Scholl UI, Ji W, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A*. 2009;106(45):19096–19101.
18. Hodges E, Xuan Z, Balija V, et al. Genome-wide in situ exon capture for selective resequencing. *Nat Genet.* 2007;39(12):1522–1527.
19. Morris AP, Zeggini E, Lindgren CM. Identification of novel putative rheumatoid arthritis susceptibility genes via analysis of rare variants [electronic article]. *BMC Proc.* 2009;3(suppl 7):S131. (DOI: 10.1186/1753-6561-3-S7-S131).
20. Zhu X, Feng T, Li Y, et al. Detecting rare variants for complex traits using family and unrelated data. *Genet Epidemiol*. 2010;34(2):171–187.
21. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009;106(23):9362–9367.
22. Hindorff LA, Junkins HA, Hall PN, et al. A catalog of published genome-wide association studies. Bethesda, MD: National Human Genome Research Institute; 2008. (www.genome.gov/gwastudies). (Accessed December 8, 2009).
23. Hoggart CJ, Clark TG, De Iorio M, et al. Genome-wide significance for dense SNP and resequencing data. *Genet Epidemiol*. 2008;32(2):179–185.
24. Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med*. 2000;19(22):3127–3131.
25. Gauderman J, Morrison JQUANTO. 1.1: a computer program for power and sample size calculations for genetic-epidemiology studies. Los Angeles, CA: University of Southern California, 2006. (http://hydra.usc.edu/gxe/).
26. Ge D, Zhang K, Need AC, et al. WGAViewer: software for genomic annotation of whole genome association studies. *Genome Res.* 2008;18(4):640–643.
27. Ge D, Goldstein DB. WGAViewer. Durham, NC: Duke University School of Medicine; 2010. (http://people.genome.duke.edu/~dg48/WGAViewer/).
28. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res.* 2002;12(6):996–1006.
29. Johnson AD, Handsaker RE, Pulit SL, et al. SNAP: a Web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*. 2008;24(24):2938–2939.
30. International HapMap Consortium. The International HapMap Project. *Nature*. 2003;426(6968):789–796.
31. International HapMap Consortium. A haplotype map of the human genome. *Nature*. 2005;437(7063):1299–1320.
32. Thorisson GA, Smith AV, Krishnan L, et al. The International HapMap Project Web site. *Genome Res.* 2005;15(11):1592–1593.
33. Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet.* 2007;39(5):631–637.
34. Gudmundsson J, Sulem P, Gudbjartsson DF, et al. Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat Genet.* 2009;41(10):1122–1126.
35. Soranzo N, Spector TD, Mangino M, et al. A genome-wide meta-analysis identifies 22 loci associated with eight

hematological parameters in the HaemGen Consortium. *Nat Genet*. 2009;41(11):1182–1190.
36. Ganesh SK, Zakai NA, van Rooij FJ, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet*. 2009;41(11):1191–1198.
37. Franke A, Hampe J, Rosenstiel P, et al. Systematic association mapping identifies *NELL1* as a novel IBD disease gene [electronic article]. *PLoS One*. 2007;2(1e691.
38. Barrett JC, Hansoul S, Nicolae DL, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet.* 2008;40(8):955–962.
39. Treviño LR, Yang W, French D, et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat Genet.* 2009;41(9):1001–1005.
40. Limou S, Le Clerc S, Coulonges C, et al. Genomewide association study of an AIDS-nonprogression cohort emphasizes the role played by HLA genes (ANRS Genomewide Association Study 02). *J Infect Dis*. 2009;199(3):419–426.
41. Sulem P, Gudbjartsson DF, Stacey SN, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat Genet.* 2007;39(12):1443–1452.
42. Rivadeneira F, Styrkársdottir U, Estrada K, et al. Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. *Nat Genet.* 2009;41(11):1199–1206.
43. Need AC, Attix DK, McEvoy JM, et al. A genome-wide study of common SNPs and CNVs in cognitive performance in the CANTAB. *Hum Mol Genet.* 2009;18(23):4650–4661.
44. Kathiresan S, Willer CJ, Peloso GM, et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet.* 2009;41(1):56–65.
45. Kooner JS, Chambers JC, Aguilar-Salinas CA, et al. Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. *Nat Genet.* 2008;40(2):149–151.
46. Kathiresan S, Melander O, Guiducci C, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet.* 2008;40(2):189–197.
47. Grant SF, Wang K, Zhang H, et al. A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. *J Pediatr.* 2009;155(6):909–913.
48. Ng CC, Yew PY, Puah SM, et al. A genome-wide association study identifies ITGA9 conferring risk of nasopharyngeal carcinoma. *J Hum Genet.* 2009;54(7):392–397.
49. Otowa T, Yoshida E, Sugaya N, et al. Genome-wide association study of panic disorder in the Japanese population. *J Hum Genet.* 2009;54(2):122–126.
50. Hirschfield GM, Liu X, Xu C, et al. Primary biliary cirrhosis associated with HLA, IL12A, and IL12RB2 variants. *N Engl J Med*. 2009;360(24):2544–2555.
51. Melzer D, Perry JR, Hernandez D, et al. A genome-wide association study identifies protein quantitative trait loci (pQTLs) [electronic article]. *PLoS Genet.* 2008;4(5):e1000072.
52. Liu Y, Helms C, Liao W, et al. A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci [electronic article]. *PLoS Genet.* 2008;4(3):e1000041.
53. Yang JJ, Cheng C, Yang W, et al. Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia. *JAMA*. 2009;301(4):393–403.
54. Aberg K, Adkins DE, Bukszár J, et al. Genomewide association study of movement-related adverse antipsychotic effects. *Biol Psychiatry.* 2010;67(3):279–282.
55. Graham RR, Cotsapas C, Davies L, et al. Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat Genet.* 2008;40(9):1059–1061.
56. Han JW, Zheng HF, Cui Y, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat Genet*. 2009;41(11):1234–1237.
57. Nan H, Kraft P, Qureshi AA, et al. Genome-wide association study of tanning phenotype in a population of European ancestry. *J Invest Dermatol*. 2009;129(9):2250–2257.
58. Willer CJ, Sanna S, Jackson AU, et al. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet.* 2008;40(2):161–169.
59. Pollin TI, Damcott CM, Shen H, et al. A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. *Science*. 2008;322(5908):1702–1705.
60. Barrett JC, Clayton DG, Concannon P, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet.* 2009;41(6):703–707.
61. Takeuchi F, Serizawa M, Yamamoto K, et al. Confirmation of multiple risk loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population. *Diabetes*. 2009;58(7):1690–1699.
62. Sulem P, Gudbjartsson DF, Stacey SN, et al. Two newly identified genetic determinants of pigmentation in Europeans. *Nat Genet.* 2008;40(7):835–837.
63. Kayser M, Liu F, Janssens AC, et al. Three genome-wide association studies and a linkage analysis identify *HERC2* as a human iris color gene. *Am J Hum Genet.* 2008;82(2):411–423.
64. Chasman DI, Paré G, Zee RY, et al. Genetic loci associated with plasma concentration of low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, apolipoprotein A1, and apolipoprotein B among 6382 white women in genome-wide analysis with replication. *Circ Cardiovasc Genet.* 2008;1(1):21–30.
65. Timpson NJ, Lindgren CM, Weedon MN, et al. Adiposity-related heterogeneity in patterns of type 2 diabetes susceptibility observed in genome-wide association data. *Diabetes*. 2009;58(2):505–510.
66. Zeggini E, Scott LJ, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet.* 2008;40(5):638–645.
67. Salonen JT, Uimari P, Aalto JM, et al. Type 2 diabetes whole-genome association study in four populations: the DiaGen Consortium. *Am J Hum Genet.* 2007;81(2):338–345.
68. Florez JC, Manning AK, Dupuis J, et al. A 100K genome-wide association scan for diabetes and related traits in the Framingham Heart Study: replication and integration with other genome-wide datasets. *Diabetes*. 2007;56(12):3063–3074.
69. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*. 2007;316(5829):1341–1345.
70. Rung J, Cauchi S, Albrechtsen A, et al. Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat Genet.* 2009;41(10):1110–1115.
71. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*. 2007;445(7130):881–885.

72. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, et al. New sequence variants associated with bone mineral density. *Nat Genet.* 2009;41(1):15–17.
73. Timpson NJ, Tobias JH, Richards JB, et al. Common variants in the region around Osterix are associated with bone mineral density and growth in childhood. *Hum Mol Genet.* 2009; 18(8):1510–1517.
74. Styrkarsdottir U, Halldorsson BV, Gretarsdottir S, et al. Multiple genetic loci for bone mineral density and fractures. *N Engl J Med*. 2008;358(22):2355–2365.
75. Duerr RH, Taylor KD, Brant SR, et al. A genome-wide association study identifies *IL23R* as an inflammatory bowel disease gene. *Science*. 2006;314(5804):1461–1463.
76. Rioux JD, Xavier RJ, Taylor KD, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet.* 2007;39(5):596–604.
77. Burton PR, Clayton DG, Cardon LR, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661–678.
78. Kugathasan S, Baldassano RN, Bradfield JP, et al. Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. *Nat Genet.* 2008;40(10): 1211–1215.
79. Raelson JV, Little RD, Ruether A, et al. Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. *Proc Natl Acad Sci U S A*. 2007;104(37):14747–14752.
80. Wallace C, Newhouse SJ, Braund P, et al. Genome-wide association study identifies genes for biomarkers of cardiovascular disease: serum urate and dyslipidemia. *Am J Hum Genet.* 2008;82(1):139–149.
81. Aulchenko YS, Ripatti S, Lindqvist I, et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet.* 2009;41(1):47–55.
82. Sabatti C, Service SK, Hartikainen AL, et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat Genet.* 2009;41(1):35–46.
83. Chambers JC, Zhang W, Li Y, et al. Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. *Nat Genet.* 2009;41(11): 1170–1172.
84. Capon F, Bijlmakers MJ, Wolf N, et al. Identification of *ZNF313/RNF114* as a novel psoriasis susceptibility gene. *Hum Mol Genet.* 2008;17(13):1938–1945.
85. Nair RP, Duffin KC, Helms C, et al. Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. *Nat Genet.* 2009;41(2):199–204.
86. Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research. *Science*. 2007;316(5829): 1331–1336.
87. Zeggini E, Weedon MN, Lindgren CM, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*. 2007;316(5829): 1336–1341.
88. Meigs JB, Manning AK, Fox CS, et al. Genome-wide association with diabetes-related traits in the Framingham Heart Study [electronic article]. *BMC Med Genet.* 2007;8(suppl 1): S16. (doi:10.1186/1471-2350-8-S1-S16).
89. Schork NJ, Murray SS, Frazer KA, et al. Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev.* 2009;19(3):212–219.
90. Gorlov IP, Gorlova OY, Sunyaev SR, et al. Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet.* 2008;82(1):100–112.
91. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet.* 2008;40(6):695–701.
92. Cohen JC, Pertsemlidis A, Fahmi S, et al. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. *Proc Natl Acad Sci U S A*. 2006;103(6):1810–1815.
93. Khoury MJ, Little J, Gwinn M, et al. On the synthesis and interpretation of consistent but weak gene-disease associations in the era of genome-wide association studies. *Int J Epidemiol*. 2007;36(2):439–445.
94. Benn M, Stene MC, Nordestgaard BG, et al. Common and rare alleles in apolipoprotein B contribute to plasma levels of low-density lipoprotein cholesterol in the general population. *J Clin Endocrinol Metab.* 2008;93(3):1038–1045.
95. McClellan JM, Susser E, King MC. Schizophrenia: a common disease caused by multiple rare alleles. *Br J Psychiatry.* 2007;190(3):194–199.
96. Wong GK, Yang Z, Passey DA, et al. A population threshold for functional polymorphisms. *Genome Res.* 2003;13(8): 1873–1879.
97. Ku CS, Loy EY, Pawitan Y, et al. The pursuit of genome-wide association studies: where are we now? *J Hum Genet.* 2010;55(4):195–206.
98. Zeggini E, Rayner W, Morris AP, et al. An evaluation of HapMap sample size and tagging SNP performance in large-scale empirical and simulated data sets. *Nat Genet.* 2005; 37(12):1320–1322.
99. Collins FS, Manolio TA. Merging and emerging cohorts: necessary but not sufficient [commentary]. *Nature*. 2007; 445(7125):259.
100. Zeggini E, Ioannidis JPA. Meta-analysis in genome-wide association studies. *Pharmacogenomics*. 2009;10(2): 191–201.
101. Need AC, Goldstein DB. Next generation disparities in human genomics: concerns and remedies. *Trends Genet.* 2009;25(11):489–494.
102. Adeyemo A, Rotimi C. Genetic variants associated with complex human diseases show wide variation across multiple populations. *Public Health Genomics*. 2010;13(2):72–79.
103. Dickson SP, Wang K, Krantz I, et al. Rare variants create synthetic genome-wide associations [electronic article]. *PLoS Biol.* 2010;8(1):e1000294.
104. Ioannidis JPA. Population-wide generalizability of genome-wide discovered associations. *J Natl Cancer Inst*. 2009; 101(19):1297–1299.
105. 1000 Genomes. A deep catalog of human genetic variation. Bethesda, MD: National Human Genome Research Institute, National Institutes of Health, 2010. (http://www.1000genomes.org/). (Accessed February 25, 2010).
106. Via M, Gignoux C, Burchard EG. The 1000 Genomes Project: new opportunities for research and social challenges [electronic article]. *Genome Med.* 2010;2(1):3.
107. Ioannidis JPA. Why most discovered true associations are inflated. *Epidemiology.* 2008;19(5):640–648.
108. Ioannidis JPA, Ntzani EE, Trikalinos TA, et al. Replication validity of genetic association studies. *Nat Genet.* 2001; 29(3):306–309.
109. Lohmueller KE, Pearce CL, Pike M, et al. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet.* 2003;33(2):177–182.