nature genetics

Weighting sequence variants based on their annotation increases power of whole-genome association studies

Gardar Sveinbjornsson[1,2], Anders Albrechtsen[3], Florian Zink[1], Sigurjón A Gudjonsson[1], Asmundur Oddson[1], Gísli Másson[1], Hilma Holm[1,4], Augustine Kong[1,2], Unnur Thorsteinsdottir[1,5], Patrick Sulem[1], Daniel F Gudbjartsson[1,2] & Kari Stefansson[1,5]

The consensus approach to genome-wide association studies (GWAS) has been to assign equal prior probability of association to all sequence variants tested. However, some sequence variants, such as loss-of-function and missense variants, are more likely than others to affect protein function and are therefore more likely to be causative. Using data from whole-genome sequencing of 2,636 Icelanders and the association results for 96 quantitative and 123 binary phenotypes, we estimated the enrichment of association signals by sequence annotation. We propose a weighted Bonferroni adjustment that controls for the family-wise error rate (FWER), using as weights the enrichment of sequence annotations among association signals. We show that this weighted adjustment increases the power to detect association over the standard Bonferroni correction. We use the enrichment of associations by sequence annotation we have estimated in Iceland to derive significance thresholds for other populations with different numbers and combinations of sequence variants.

GWAS have increased understanding of the contribution of common sequence variants to complex traits[1,2]. Recent technological advances have made large-scale whole-genome sequencing feasible, allowing nearly complete coverage of the sequence diversity of a population under study[3,4]. This extended coverage makes it likely that the sequence variant causing the functional perturbation behind an association has itself been identified and tested.

The multiple-testing burden from testing all common variants (minor allele frequency (MAF) greater than 5%) for association in a European sample of 1,000 cases and 1,000 controls has been estimated to correspond to 1 million independent tests[5] and, in turn, a genome-wide significance threshold of $P = 5 \times 10^{-8}$. However, this burden increases with sample size (up to 1.5 million independent tests for 10,000 cases and 10,000 controls[5]) and is higher for African populations[5]. Most sequence variants occur at a frequency of less than 5% (refs. 3,4) and are expected to be young and in weak linkage disequilibrium (LD) with surrounding markers. Thus, if rare sequence variants are included in association analysis, the multiple-testing burden increases.

For some time, GWAS have been based on imputing genotypes for common SNPs from the HapMap Project[6] and using the significance threshold $P = 5 \times 10^{-8}$. The current releases of the 1000 Genomes Project European population and the Genomes of the Netherlands (GoNL) Project reference sets include around 15 and 20 million variants, respectively[4,7], many of which are rare. Hence, a more stringent significance threshold is necessary for GWAS based on imputing from these sets[8]. Permutation of 2,157 cases and 1,150 controls leads to a significance threshold of $P = 1.5 \times 10^{-8}$ based on the 1000 Genomes Project data[8]. Again, a more stringent threshold would have to be applied to larger association studies.

The threshold for significance in future GWAS will depend on the study population and the number of testable variants, and a universal threshold will not be applicable. Because most sequence variants are rare and in weak LD with neighboring markers, we opt for using the number of variants testable in the genome when choosing a threshold for genome-wide significance.

The consensus approach has been to assign equal prior probability of association to all variants[9]. However, some sequence variants, such as coding variants, are predicted to affect protein function and are more likely to be causal than other variants[10–13]. Having nearly complete coverage of the sequence diversity in a population makes weighting sequence variants on the basis of their predicted effects practical, as causal variants are likely to have been discovered and tested. Previously, we used a simple weighted Bonferroni scheme and applied different genome-wide significance thresholds depending on the predicted effects of sequence variants[14]. Here we seek to demonstrate that such a weighting scheme is not only intuitively appealing but also increases power. We do this by analyzing the GWAS results for a diverse set of quantitative and binary phenotypes and estimating the enrichment of association signals by sequence annotation.

We imputed 14.2 million high-quality sequence variants (giving a Bonferroni multiple-testing threshold of $P = 0.05/14.2$ million $= 3.5 \times 10^{-9}$) from whole-genome sequencing of 2,636 Icelanders to a mean depth of at least 10× (median depth of 20×) into 104,220 Icelanders and their first- and second-degree relatives and tested

[1]deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. [2]School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland. [3]Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark. [4]Division of Cardiology, Department of Internal Medicine, Landspitali, National University Hospital of Iceland, Reykjavik, Iceland. [5]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. Correspondence should be addressed to D.F.G. (daniel.gudbjartsson@decode.is) or K.S. (kari.stefansson@decode.is).

Received 11 November 2015; accepted 15 January 2016; published online 8 February 2016; doi:10.1038/ng.3507

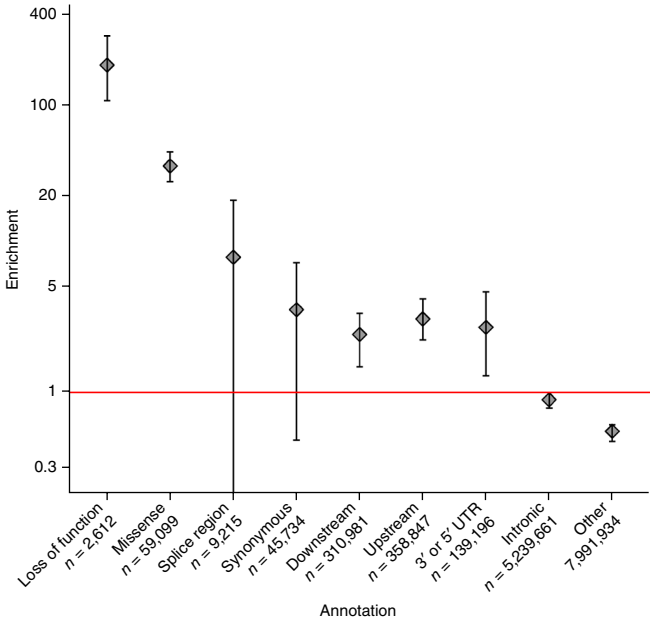
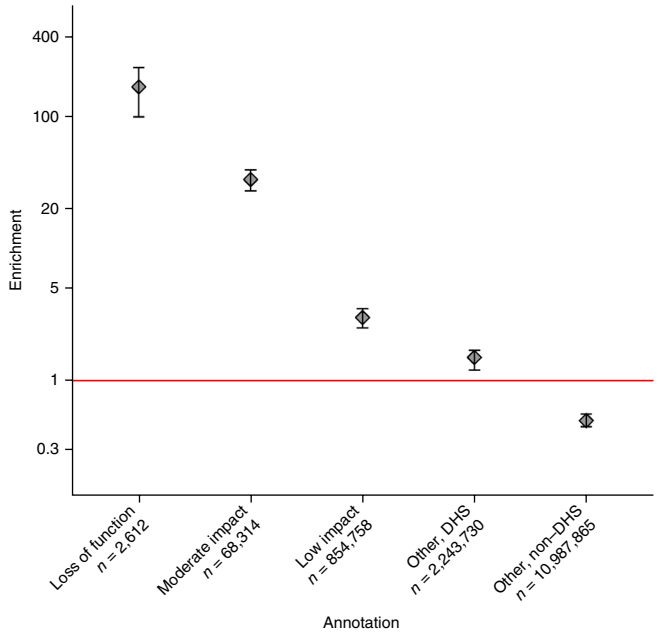


Figure 1 Enrichments of VEP annotations for all quantitative and binary phenotypes. The figure shows enrichment of loss-of-function variants, including splice donor, splice acceptor, stop-gain, frameshift and stop-loss variants. The missense category includes both missense variants and in-frame indels. The error bars indicate 95% confidence intervals derived from bootstrapping.

them for association with 96 quantitative and 123 binary phenotypes (the median number of individuals for quantitative traits was 23,093 and the median number of cases for binary phenotypes was 1,431; **Supplementary Note**)[3]. We used variants associating with a *P* value below 1×10^{-8} to estimate enrichment. We stress that this is a cutoff for estimating enrichment but not a threshold for association discovery. Given the phenotype associations for this set of sequence variants, we sequentially defined a list of association signals by (1) adding the most significantly associating sequence variant to the list, (2) adjusting the association of all variants in the region of the variant from step 1 for its association with the phenotype and (3) repeating steps 1 and 2 while the set of associating sequence variants was not empty. We considered sequence variants with LD (r^2) >0.2 with a significant variant on the list as possibly causal, defining an association signal. We found 609 and 142 association signals for the quantitative and binary phenotypes, respectively.

We used a likelihood-based inference to estimate the enrichment of phenotype-associating sequence variants on the basis of their annotations. This approach accounts for the frequencies of annotations and LD between variants.

We used the Variant Effect Predictor (VEP)[15] to predict the maximal consequence of each sequence variant on all neighboring RefSeq genes[16]. There was substantial variation in the enrichment of VEP sequence annotations (**Fig. 1** and **Supplementary Table 1**). The enrichments corresponded well with negative selection signatures for variants with these annotations, as measured by variant density and the fraction of rare variants[3], with the notable exception that upstream and downstream variants (variants within 5 kb of a gene) showed greater enrichment than intronic and intergenic variants. On the basis of these enrichments, we grouped sequence variants into four categories, in order of decreasing effect on biological function: (i) loss-of-function variants (stop-gain and stop-loss, frameshift indel, donor and acceptor splice-site and initiator codon variants); (ii) moderate-impact variants (missense, in-frame indel and splice region variants); (iii) low-impact variants (synonymous, 3′ and 5′ UTR, and upstream and downstream variants); and (iv) other variants (**Fig. 2** and **Supplementary Table 2**).



Other genomic annotations might also be enriched among association signals. The Encyclopedia of DNA Elements (ENCODE) has annotated regulatory regions across the whole genome[17]. Annotations such as DNase I hypersensitivity sites (DHSs) are enriched in GWAS findings, as well as enhancer-associated marks that were shown to be enriched using tissue-specific epigenomic data[18,19]. Regulatory variants in noncoding regions may also explain more heritability than coding variants[20]. We assessed the enrichment of sequence variants in the DHS peaks observed in any of 218 cell types[20] (18% of the variants), as well as narrowing the DHS peaks to those overlapping ENCODE ChromHMM-predicted enhancers (4% of variants) (**Fig. 2**)[21]. Sequence variants within DHS sites were enriched by 1.8-fold for quantitative traits and 1.7-fold for binary traits in noncoding regions, and predicted enhancers were enriched by 3.0-fold and 3.2-fold, respectively (**Supplementary Table 3**). For intronic and intergenic DHS variants, the enrichment was 1.5-fold as compared to 0.5-fold enrichment for intronic and intergenic variants not in DHSs (**Fig. 2** and **Supplementary Table 2**).

We also annotated sequence variants with genomic evolutionary rate profiling (GERP) conservation scores[22,23] and combined annotation-dependent depletion (CADD) scores[24], both of which were overall enriched but not substantially beyond the VEP annotations (**Supplementary Tables 4** and **5**).

Functional annotations can be used to improve power to detect associations[25–27]. To increase power by incorporating the information

Figure 2 Enrichments for all quantitative and binary phenotypes. The five categories are defined as (i) loss-of-function variants, including stop-gain and stop-loss, frameshift indel, donor and acceptor splice-site, and initiator codon variants; (ii) moderate-impact variants, including missense, in-frame indel and splice region variants; (iii) low-impact variants, including synonymous, 3′ and 5′ UTR, and upstream and downstream variants; and (iv + v) other variants, consisting of deep intronic and intergenic variants. The last category is divided according to broadly defined DHSs. The error bars indicate 95% confidence intervals derived from bootstrapping.

Table 1 Weights derived from different weighting schemes and the resulting number of significant associations among unsettled association signals

		Weighting scheme					
Category	Broad DHS	Standard Bonferroni	Simple	Enrichment QT	Enrichment B	Enrichment QT+B	Range of optimal QT+B weights
Loss of function	–	1	1,355	118	452	182	186–239
Moderate impact	–	1	52	37	36	37	66–70
Low impact	–	1	4.1	3.5	1.8	3.1	1.6–1.9
Other	Yes	1	0.27	1.3	1.2	1.3	2.9–3.1
	No	1	0.27	0.50	0.56	0.52	0.080–0.094
n significant	QT	122	123	136	138	139	151
unsettled	B	24	27	27	27	27	34
association signals	QT+B	146	150	163	165	166	185

The weights for the enrichment-derived weighting schemes were estimated on the basis of settled ($P < 1 \times 10^{-10}$) association signals. Categories are based on VEP annotation, with the "other" category (intronic and intergenic variants) further divided depending on broadly defined DHSs. Enrichment weights were estimated on the basis of 437 settled association signals for quantitate traits (QT) and 103 settled association signals for binary phenotypes (B), or both combined (QT+B). For each weighting scheme, the number of significant unsettled ($P > 1 \times 10^{-10}$) association signals is given. The range of estimated optimal weights and the resulting number of significant unsettled association signals are also given.

that these annotations hold, we suggest a weighting scheme based on the observed enrichment using a weighted Bonferroni correction similar to that suggested by Roeder *et al.* for linkage analysis[28–30]. The weighted Bonferroni correction leads to thresholds for genome-wide significance that depend on sequence annotation instead of being the same for all sequence variants. This is a robust procedure in the sense that informative specification of the weights can increase power substantially, whereas uninformative weighting leads to little loss of power for sparse weights[28–30].

Association signals with very low *P* values will be significant for all reasonable weighting schemes. Enrichment can be estimated from these 'settled' association signals, and the derived significance threshold can then be applied to the remaining 'unsettled' association signals with more moderate *P* values. To evaluate the relative power of different weighting schemes, we counted the number of unsettled association signals that satisfied significance thresholds by evaluating weights using only the settled associations. We chose to consider association signals with a *P* value less than 1×10^{-10} as settled. We used the VEP annotations and split the intronic and intergenic variants according to broadly defined DHSs. We estimated enrichment weights for the quantitative and binary phenotypes separately as well as for the two sets combined (**Table 1**). The enrichment estimates were similar for the quantitative and binary phenotypes. We also calculated the optimal weighting scheme—that is, the weighting scheme that maximized the number of significant association signals—by taking all association signals and optimizing over the weighting schemes that controlled the FWER at 0.05 (**Table 1**). Optimizing over the weighting schemes will overfit the data, but we note that the number of significant signals using the enrichments as weights ($n = 166$) was not far from being optimal ($n = 185$), and this approach had more power than the standard Bonferroni method ($n = 146$) to detect associations. To analyze the robustness of the estimated enrichment, we used the data from odd-numbered chromosomes and evaluated the weighting scheme on the independent set of association signals on even-numbered chromosomes, and vice versa. We found that the enrichment weighting scheme performed well on the independent sets (**Supplementary Table 6**).

The estimated enrichment was similar across Human Phenotype Ontology (HPO)[31] classes (**Supplementary Table 7**). The enrichment was also similar for variants with MAFs greater than and less than 2% (**Supplementary Table 8**).

Several methods have been used to estimate enrichment of functional classes[11,20,25,32]. The method of LD score regression[32] gave similar enrichment estimates and numbers of associations when applied to a weighted Bonferroni procedure (**Supplementary Table 9**).

The genome-wide significance threshold that the enrichment weights gave on the basis of all association signals ($P < 1 \times 10^{-8}$) varied substantially between categories and was $P = 5.8 \times 10^{-7}$ for loss-of-function variants and $P = 1.2 \times 10^{-7}$ for moderate-impact variants, as compared to $P = 1.8 \times 10^{-9}$ for deep intronic and intergenic variants not in DHSs (**Table 2**). The enrichment estimates can be used to derive significance thresholds for different whole-genome sequencing data sets, such as the GoNL and 1000 Genomes projects (**Table 2**), given that the frequency distributions of the categories in these data sets are similar to those in the Icelandic population. The estimated enrichment will in particular depend on the characteristics of the population. The Icelandic, GoNL and 1000 Genomes Project data sets are all of European origin, and the derived thresholds were similar.

Exome sequencing is a common alternative to whole-genome sequencing. Giving all variants in coding regions equal weights and assigning no weight to other variants would yield a genome-wide significance threshold of $P = 4.3 \times 10^{-7}$ and 374 association signals would satisfy this threshold (302 signals for quantitative and 72 signals for binary phenotypes), as compared to 686 significant association signals under the standard Bonferroni correction with a threshold of $P = 3.5 \times 10^{-9}$. The threshold of $P = 4.3 \times 10^{-7}$ differed from the

Table 2 The estimated enrichment of categories among association signals and the resulting significance thresholds

			Iceland		GoNL		1000 Genomes	
Category	Broad DHS	Enrichment	*n* variants	Significance threshold	*n* variants	Significance threshold	*n* variants	Significance threshold
Loss of function	–	165	2,612	5.8×10^{-7}	4,423	4.1×10^{-7}	2,426	5.5×10^{-7}
Moderate impact	–	33	68,314	1.2×10^{-7}	106,328	8.1×10^{-8}	70,468	1.1×10^{-7}
Low impact	–	3.0	854,758	1.1×10^{-8}	1,168,209	7.4×10^{-9}	914,868	1.0×10^{-8}
Other	Yes	1.5	2,243,730	5.3×10^{-9}	3,163,835	3.7×10^{-9}	2,252,786	5.1×10^{-9}
	No	0.50	10,987,865	1.8×10^{-9}	15,585,939	1.2×10^{-9}	11,910,249	1.7×10^{-9}
Total			14,157,279		20,028,734		15,150,797	

The resulting significance thresholds are derived from application of the enrichment weighting scheme on the Icelandic whole-genome sequencing data as well as the GoNL and European 1000 Genomes Project data sets.

enrichment thresholds for loss-of-function and moderate-impact variants by a factor of 1.3 and 0.3, respectively. Thus, the enrichment thresholds for coding sequence variants do not differ substantially from Bonferroni thresholds used in exome sequencing studies, and whole-genome studies are not penalized for including the plethora of intergenic variants with low prior probability of association.

Estimating the enrichment of sequence variants by category is complicated by the LD between causal and non-causal variants. Assessing enrichment on the basis of lists of published association signals[2] will lead to biased results because coverage is incomplete and biased and publication and fine-mapping efforts will depend on category. We avoided several of the major biases by relying on whole-genome sequencing data and by modeling the distribution of categories in a maximum-likelihood framework, taking into account the frequencies of the non-causal sites and LD between sequence variants. Even so, some of our assumptions may be invalid: for example, assuming that only a single variant in an association signal is causal. Also, some causal variants may be missed in our testing, for example, structural variants and other variants not covered by short-read sequencing. In general, misspecification of the model or error in determining the true causal variants will shrink the enrichment estimates for the most enriched categories.

The enrichments derived for quantitative and binary phenotypes were similar, and using the weights derived for the quantitative traits for the binary phenotypes gave similar power as using the binary phenotype weights, and vice versa. Under the enrichment thresholds, ten coding variants became significant that were not significant under standard Bonferroni correction (**Supplementary Table 10**). Six of the ten coding variants were in genes that have been associated with a corresponding phenotype, and two variants in *FUT2*, previously associated with vitamin B_{12} levels[33], associated with ferritin.

The weighting scheme based on quantitative and binary phenotypes can be applied to other annotated sets of sequence variants. This weighting scheme controls the FWER and is more powerful than the standard Bonferroni correction. It has the benefit of being robust and simple to use and implement. Additionally, this scheme can be used both for whole-genome and exome association analyses, and the significance thresholds will not differ substantially. Thus, association studies based on whole-genome sequencing will not have their variants with a high prior probability of association, in particular loss-of-function and missense variants, unduly penalized because of the addition of a large number of sequence variants with low prior probability. At the time the GWAS consensus of equal weight to all variants was reached, the available data were far from capturing all the diversity in the human sequence. The commonly accepted $P = 5 \times 10^{-8}$ threshold is outdated and will not be applicable in future GWAS.

URLs. LD score regression, http://www.github.com/bulik/ldsc.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors thank all the participants in the study. We also thank the staff at the Patient Recruitment Center and the deCODE Genetics core facilities.

AUTHOR CONTRIBUTIONS

The study was designed and results were interpreted by G.S., A.A., G.M., H.H., A.K., U.T., P.S., D.F.G. and K.S. G.S., A.A., F.Z., S.A.G., A.O., G.M., A.K., P.S. and D.F.G. performed the statistical and bioinformatics analyses. The manuscript was drafted by G.S., A.A., P.S., D.F.G. and K.S. All authors contributed to the final version of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106**, 9362–9367 (2009).
2. Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* **42**, D1001–D1006 (2014).
3. Gudbjartsson, D.F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435–444 (2015).
4. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
5. Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* **32**, 381–385 (2008).
6. International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
7. Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* **46**, 818–825 (2014).
8. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
9. Sham, P.C. & Purcell, S.M. Statistical power and significance testing in large-scale genetic studies. *Nat. Rev. Genet.* **15**, 335–346 (2014).
10. Thomas, P.D. & Kejariwal, A. Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc. Natl. Acad. Sci. USA* **101**, 15398–15403 (2004).
11. Schork, A.J. *et al.* All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genet.* **9**, e1003449 (2013).
12. Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43**, 519–525 (2011).
13. Minelli, C. *et al.* Importance of different types of prior knowledge in selecting genome-wide findings for follow-up. *Genet. Epidemiol.* **37**, 205–213 (2013).
14. Sveinbjornsson, G. *et al.* Rare mutations associating with serum creatinine and chronic kidney disease. *Hum. Mol. Genet.* **23**, 6935–6943 (2014).
15. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
16. Eilbeck, K. *et al.* The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol.* **6**, R44 (2005).
17. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
18. Maurano, M.T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
19. Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
20. Gusev, A. *et al.* Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.* **95**, 535–552 (2014).
21. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
22. Cooper, G.M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15**, 901–913 (2005).
23. Goode, D.L. *et al.* Evolutionary constraint facilitates interpretation of genetic variation in resequenced human genomes. *Genome Res.* **20**, 301–310 (2010).
24. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310–315 (2014).
25. Pickrell, J.K. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am. J. Hum. Genet.* **94**, 559–573 (2014).
26. Iversen, E.S., Lipton, G., Clyde, M.A. & Monteiro, A.N. Functional annotation signatures of disease susceptibility loci improve SNP association analysis. *BMC Genomics* **15**, 398 (2014).
27. Kichaev, G. *et al.* Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* **10**, e1004722 (2014).
28. Roeder, K. & Wasserman, L. Genome-wide significance levels and weighted hypothesis testing. *Stat. Sci.* **24**, 398–413 (2009).
29. Roeder, K., Bacanu, S.A., Wasserman, L. & Devlin, B. Using linkage genome scans to improve power of association in genome scans. *Am. J. Hum. Genet.* **78**, 243–252 (2006).
30. Genovese, C.R., Roeder, K. & Wasserman, L. False discovery control with *p*-value weighting. *Biometrika* **93**, 509–524 (2006).
31. Köhler, S. *et al.* The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.* **42**, D966–D974 (2014).
32. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
33. Hazra, A. *et al.* Common variants of *FUT2* are associated with plasma vitamin B_{12} levels. *Nat. Genet.* **40**, 1160–1162 (2008).

ONLINE METHODS

Association signals and conditional association. Genotypes for sequence variants identified in 2,636 whole genome–sequenced Icelanders were imputed into 104,220 Icelanders and their relatives and tested for association[3]. Association with quantitative traits was tested using a generalized linear model, and association with binary phenotype status was tested using logistic regression[3].

All sample identifiers were encrypted in accordance with the regulations of the Icelandic Data Protection Authority. Approval for these studies was provided by the National Bioethics Committee and the Icelandic Data Protection Authority.

If a sequence variant had a P value below 1×10^{-8}, we performed conditional analysis where we tested the variant again conditioning separately on every variant with a lower P value at the locus. If the conditional P value was less than 1×10^{-8} conditional on each of the stronger variants, then a new association signal was assumed to have been detected and the sequence variant was taken as the most significant variant underlying it. An association signal was then defined as all SNPs with an r^2 value greater than 0.2 with the most significant sequence variant and was taken to be the set of all possible causal variants. Sequence variants with MAF <0.1%, imputation information <0.9, multiallelic variants and variants with quality flags were excluded from the analysis.

Assessment of category enrichment. Assuming that there is a single causal variant behind each association signal, the probability of the phenotype y given this causal variant is independent of the other variants at the locus. Thus, for the M_i possibly causal sequence variants at the ith association signal, the probability of the phenotype given the genotype g is

$$p(y|g_1,\ldots,g_{M_i}) = p(y|g_{k_i})$$

where k_i is the index of the causal variant. We assign an annotation c_m to each variant m at the association signal. We denote the probability of a causal variant being of category C as p_C and the probability of a non-causal variant being of category C as q_C. We estimate the non-causal probabilities, q_C, with the genome-wide frequencies of the categories. If we do not know which variant at an association signal is the causal one, then the probability of the phenotype given the genotype data at an association signal is

$$\sum_{m=1}^{M_i} p(y|g_m, k_i = m)\, p(k_i = m)$$

where

$$p(k_i = m) = p(c_m|k_i = m) \prod_{m' \neq m} p(c_{m'}|k_i \neq m') = p_{c_m} \prod_{m' \neq m} q_{c_{m'}}$$

The phenotype and genotype data alone contain almost no information about the probability distribution of the causal variants over categories, p_C, and we assume that $p(y)$ and $p(g)$ do not depend on p_C. Then, the total likelihood for all association signals is

$$p(y|g) \propto p(g|y) = \prod_i^L p(g_i|y) \propto \prod_i^L p(y|g_i) = \prod_i^L \sum_{m=1}^{M_i} p(y|g_m, k_i = m) p_{c_m} \prod_{m' \neq m} q_{c_{m'}}$$

We approximate the term $p(y|g_m, k_i = m)$ with the probability of y given the maximum-likelihood estimate under the linear regression model, which is proportional to the likelihood-ratio test statistic used in the genome-wide association scan. We estimate the probabilities of the categories, p_C, by maximizing the likelihood numerically[34]. We define the enrichment of a category as the probability of a causal variant being from a class divided by the probability of a non-causal variant being from the class (genomic frequencies), or p_C/q_C.

Enrichment confidence intervals. We used percentile bootstrap to obtain 95% confidence intervals for the enrichment estimates[35]. The association signals were sampled 1,000 times to calculate each confidence interval.

Weighted Bonferroni correction. Weighted Bonferroni correction is an extension of Bonferroni correction[28–30]. For T tests, non-negative weights $w = (w_1, \ldots, w_T)$ are specified, and each hypothesis H_j is rejected if

$$\frac{P_j}{wj} \leq \frac{0.05}{T}$$

As long as

$$\frac{1}{T}\sum_j w_j = 1$$

this procedure controls the FWER at 0.05.

We suggest using enrichments as weights

$$w_j = e_{c_j}$$

where c_j is the category to which the jth sequence variant belongs. Let T_C be the number of sequence variants in category C, and recall that we estimate the frequency of category C, q_C, with T_C/T. Then

$$\frac{1}{T}\sum_j w_j = \frac{1}{T}\sum_j e_{c_j} = \frac{1}{T}\sum_C T_C e_C = \frac{1}{T}\sum_C T_C \frac{p_C}{q_C} = \frac{1}{T}\sum_C T p_C = \sum_C p_C = 1$$

showing that using the enrichments as weights in the population from which they were derived controls the FWER.

In other populations, the weights need to be rescaled to control the FWER. If T' sequence variants are being tested and the number of variants in category C is T'_C, then we suggest setting the weight for the jth test as

$$w'_j = \frac{e_{c_j}}{\frac{1}{T'}\sum_C T'_C e_C}$$

It is easy to see that using these weights controls the FWER.

Estimation of partitioned heritability. Stratified LD score regression was used to estimate partitioned heritability and enrichment[32]. Open source ldsc was used in the analysis, and LD score was calculated for all variants using our sequencing data. We estimated enrichment for all quantitative and binary traits. In cases where estimates of enrichment were negative, we set the values to zero. To obtain weights, we averaged the enrichment for each category over all phenotypes and rescaled to fulfill the criteria for weighted Bonferroni correction.

34. Nelder, J.A. & Mead, R. A simplex method for function minimization. *Comput. J.* **7**, 308–313 (1965).
35. Efron, B. & Tibshirani, R.J. *An Introduction to the Bootstrap* (Chapman & Hall/CRC, 1993).

doi:10.1038/ng.3507