nature genetics

LD Score regression distinguishes confounding from polygenicity in genome-wide association studies

Brendan K Bulik-Sullivan[1–3], Po-Ru Loh[1,4], Hilary K Finucane[4,5], Stephan Ripke[2,3], Jian Yang[6], Schizophrenia Working Group of the Psychiatric Genomics Consortium[7], Nick Patterson[1], Mark J Daly[1–3], Alkes L Price[1,4,8] & Benjamin M Neale[1–3]

Both polygenicity (many small genetic effects) and confounding biases, such as cryptic relatedness and population stratification, can yield an inflated distribution of test statistics in genome-wide association studies (GWAS). However, current methods cannot distinguish between inflation from a true polygenic signal and bias. We have developed an approach, LD Score regression, that quantifies the contribution of each by examining the relationship between test statistics and linkage disequilibrium (LD). The LD Score regression intercept can be used to estimate a more powerful and accurate correction factor than genomic control. We find strong evidence that polygenicity accounts for the majority of the inflation in test statistics in many GWAS of large sample size.

Variants in LD with a causal variant show an elevation in test statistics in association analysis proportional to their LD (measured by r^2) with the causal variant[1–3]. The more genetic variation an index variant tags, the higher the probability that this index variant will tag a causal variant. In contrast, inflation from cryptic relatedness within or between cohorts[4–6] or population stratification purely from genetic drift will not correlate with LD.

Under a polygenic model, in which effect sizes for variants are drawn independently from distributions with variance proportional to $1/(p(1-p))$, where p is the minor allele frequency (MAF), the expected χ^2 statistic of variant j is:

$$E[\chi^2 \mid \ell_j] = Nh^2\ell_j/M + Na + 1 \quad (1)$$

where N is the sample size; M is the number of SNPs, such that h^2/M is the average heritability explained per SNP; a measures the contribution of confounding biases, such as cryptic relatedness and population stratification; and $\ell_j = \Sigma_k r_{jk}^2$ is the LD Score of variant j, which measures the amount of genetic variation tagged by j (a full derivation of this equation is provided in the **Supplementary Note**).This relationship holds for meta-analyses and also for ascertained studies of binary phenotypes, in which case h^2 is on the observed scale. Consequently, if we regress the χ^2 statistics from GWAS against LD Score (LD Score regression), the intercept minus one is an estimator of the mean contribution of confounding bias to the inflation in the test statistics.

RESULTS

Overview of methods

We estimated LD Scores from the European-ancestry samples in the 1000 Genomes Project[7] (EUR) using an unbiased estimator[8] of r^2 with 1-cM windows, singletons excluded (MAF > 0.13%) and no r^2 cutoff. Standard errors were estimated by jackknifing over blocks of individuals, and we used these standard errors to correct for attenuation bias in LD Score regression (that is, the downward bias in the magnitude of the regression slope that occurs when the regressor is measured noisily; Online Methods).

For LD Score regression, we excluded variants with EUR MAF < 1% because the LD Score standard errors for these variants were very high (note that the variants included in LD Score regression are a subset of the variants included in LD Score estimation). In addition, we excluded loci with extremely large effect sizes or extensive long-range LD from all regressions because these loci can be considered outliers in such an analysis and would have disproportionate influence on the regression (Online Methods).

An important consideration in the estimation of LD Score is the extent to which the sample from which LD Score is estimated matches the sample for the association study. If there is a mismatch between the LD Scores from the reference population and the target population used for GWAS, then LD Score regression can be biased in two ways. First, if LD Scores in the reference population are equal to LD Scores in the target population plus mean-zero noise, then the intercept will

[1]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [2]Analytical and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. [3]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [4]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. [5]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [6]Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. [7]A full list of members and affiliations appears in the **Supplementary Note**. [8]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA. Correspondence should be addressed to B.M.N. (bneale@broadinstitute.org).

Received 7 March 2014; accepted 7 January 2015; published online 2 February 2015; doi:10.1038/ng.3211

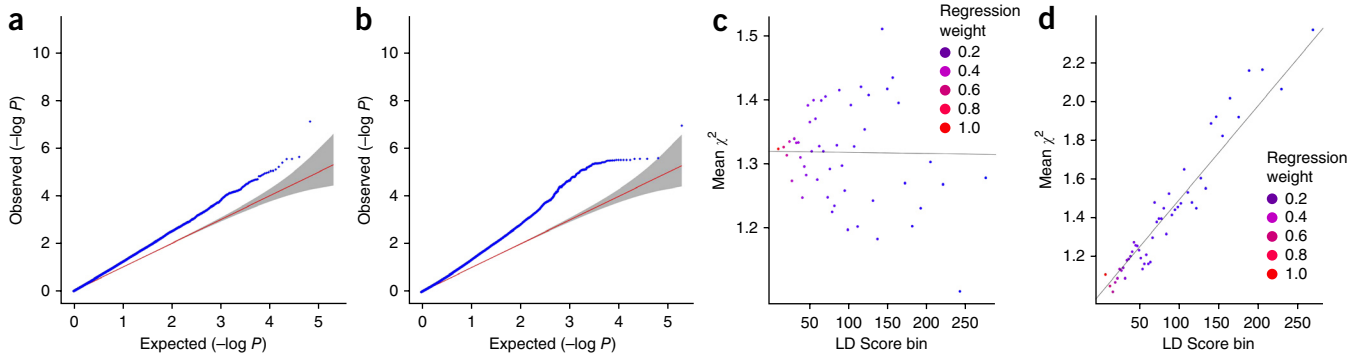

Figure 1 Results from selected simulations. (**a**) Quantile-quantile plot with population stratification (λ_{GC} = 1.32, LD Score regression intercept = 1.30). (**b**) Quantile-quantile plot with a polygenic genetic architecture where 0.1% of SNPs are causal (λ_{GC} = 1.32, LD Score regression intercept = 1.006). (**c**) LD Score plot with population stratification. Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ^2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line. (**d**) LD Score plot as in **c** but with polygenic genetic architecture.

be biased upward and the slope will be biased downward. This is conceptually equivalent to increasing the measurement error for LD Score. Second and perhaps more importantly, consider the scenario where there is a directional bias in the average LD Score such that the LD Scores in the reference population are systematically higher or lower than those in the target population. Under such a scenario, the LD Score regression intercept would be biased downward or upward, respectively (Online Methods).

To explore the stability of LD Score across European-ancestry populations, we estimated LD Scores using each of the 1000 Genomes Project EUR subpopulations separately (Utah residents with Northern and Western European ancestry (CEU), British in England and Scotland (GBR), Toscani in Italia (TSI) and Finnish in Finland (FIN)). The LD Scores from all four subpopulations were highly correlated, but mean LD Score increased with latitude (Online Methods), consistent with the observation that southern European populations have gone through less severe bottlenecks than northern European populations[9]. For example, in comparison to the combined EUR LD Score, the mean LD Score for the FIN population was 7% larger and the mean LD Score for the TSI population was 8% smaller. We evaluated the impact of these differences on the behavior of the LD Score regression analysis and found that the EUR reference panel was adequate for studies in outbred populations of predominantly northern European ancestry, such as European-American or UK populations (Online Methods). For other populations, a different reference panel should be used.

Under strong assumptions about the effect sizes of rare variants, the slope of the LD Score regression can be rescaled to be an estimate of the heritability explained by all SNPs used in the estimation of the LD Scores (**Supplementary Table 1**). Relaxing these assumptions to obtain a robust estimate of the heritability explained by all 1000 Genomes Project SNPs is a direction for further research; however, we note that the LD Score regression intercept is robust to these assumptions.

Simulations with polygenic genetic architectures

To verify the relationship between LD and χ^2 statistics, we performed a variety of simulations to model scenarios with population stratification, cryptic relatedness or polygenic architecture.

To model a polygenic quantitative trait, we assigned per-allele effect sizes drawn from the distribution $N(0, h^2/(2Mp(1-p)))$ to varying numbers of causal variants and for varying heritabilities in an approximately unstructured cohort of 1,000 Swedes. In all simulation settings, the average LD Score regression intercept was close to one (**Supplementary Figs. 1** and **2**). When there were few causal variants, the LD Score regression estimates were still unbiased but the standard errors became very large, meaning that this approach is best suited to polygenic traits (**Supplementary Figs. 3**–**5**).

Simulations with confounding

The model assumes that there is no systematic correlation between F_{ST} (a measure of the between-population variance in allele frequency) and LD Score (**Supplementary Note**). This assumption may be violated in practice as a result of linked selection (positive selection[10] and background selection[11]). If there were a positive correlation between LD Score and F_{ST}, the LD Score regression intercept would underestimate the contribution of population stratification to the inflation in χ^2 statistics. To quantify the bias that this might introduce into the LD Score regression intercept, we performed a series of simulations with real population stratification.

We obtained unimputed genotypes for Psychiatric Genomics Consortium (PGC) controls from seven European cohorts genotyped on the same array (**Supplementary Table 2**). To simulate population stratification on a continental scale, we assigned case or control status on the basis of cohort membership and then computed association statistics for each pair of cohorts (note that in this simulation setup the expected mean χ^2 statistic is $1 + bNF_{ST}$, where b is the correlation between phenotype and ancestry and N is sample size; ref. 12). To simulate population stratification on a national scale, we computed the top three principal components within each cohort and then computed association statistics using each of these principal components as a phenotype. Quantile-quantile plots from simulations with population stratification and polygenicity showed indistinguishable patterns of inflation (**Fig. 1a**,**b**), but the average LD Score regression intercept was approximately equal to the genomic control inflation factor (λ_{GC}) in simulations with population stratification (see **Supplementary Table 3a** for simulations with continental-scale stratification and **Supplementary Table 4a** for simulations with national-scale stratification) and near one in simulations with polygenicity (**Supplementary Figs. 1**–**5**).Furthermore, the qualitative appearance of the pattern of inflation as a function of LD Score was completely different in each set of simulations (**Fig. 1c**,**d**). The observed correlations between F_{ST} and LD Score in all simulations were negligible (generally 1×10^{-5} to 1×10^{-4}; **Supplementary Tables 3b** and **4b**). We note that, in simulations with population stratification, the slope

of the LD Score regression was slightly greater than zero on average (**Supplementary Tables 3c** and **4c**), likely as a result of linked selection. Nevertheless, the performance of the LD Score regression intercept was comparable to λ_{GC} and so would be suitably conservative if used as a correction factor, despite the small bias in the slope.

Simulations with confounding and polygenicity

To simulate a more realistic scenario where both polygenicity and bias contribute simultaneously to the inflation of test statistics, we obtained the genotypes for approximately 22,000 individuals throughout Europe from the Wellcome Trust Case Control Consortium 2 (ref. 13). We simulated polygenic phenotypes by drawing causal SNPs only from the first halves of chromosomes. All SNPs on the second halves of chromosomes were not causal. In addition, we included an environmental stratification component aligned with the first principal component of the genotype data, representing northern versus southern European ancestry. In this setup, the mean χ^2 statistic among SNPs on the second halves of chromosomes measures the average contribution of stratification. We performed similar simulations with cryptic relatedness using data from the Framingham Heart Study[14], which includes close relatives. In all simulation replicates, the LD Score regression intercept was approximately equal to the mean χ^2 statistic among null SNPs (**Supplementary Table 5**), which demonstrates that LD Score regression can partition the inflation in test statistics, even in the presence of both bias and polygenicity.

Finally, we modeled studies of a polygenic binary phenotype with case-control ascertainment using simulated genotypes and a liability threshold model, and we verified that LD Score regression was not noticeably biased by case-control ascertainment (**Supplementary Table 6**).

Frequency-dependent genetic architectures

LD Score regression works optimally when the variance explained per SNP is uncorrelated with the LD Score (this means that rare variants have larger effect sizes than common variants, which may be an appropriate model for a disease phenotype under moderate negative selection). A potential limitation of LD Score regression is that the variance explained by each SNP may be correlated with the LD Score for some phenotypes. For an example where this might occur, consider a phenotype that is selectively neutral, such that the per-allele effect size is uncorrelated with MAF (which means that the variance explained is positively correlated with MAF, as additive genetic variance is defined as $2pqa^2$, where p and q are the major and minor allele frequencies, respectively, and a is the additive genetic effect). Because the LD Score is also positively correlated with MAF, in this case we would expect the variance explained to be positively correlated with the LD Score, which will introduce downward bias into the LD Score regression intercept and upward bias into the LD Score regression slope, leading to underestimation of potential bias.

To quantify the magnitude of the bias that MAF-dependent genetic architectures could introduce, we simulated a frequency-dependent genetic architecture where the effect size was uncorrelated with MAF (Online Methods). For most phenotypes, this model should represent a reasonable bound on the genetic architecture. We observed minimal bias: in these simulations, the mean LD Score regression intercept was 0.994 (**Supplementary Fig. 6** and **Supplementary Table 7**). Nevertheless, there exist extreme genetic architectures where LD Score regression is not effective: for instance, if all causal variants are rare (MAF < 1%; which may be an appropriate model for a phenotype under extreme negative selection), then LD Score regression will often generate a negative slope, and the intercept will exceed the mean χ^2 statistic (**Supplementary Fig. 7**).

Table 1 LD Score regression results

Phenotype	Mean χ^2	λ_{GC}	Intercept (SE)	Type[a]	GC[b]	Ref.
Inflammatory bowel disease	1.247	1.164	1.095 (0.010)	Mega	0	26
Ulcerative colitis	1.174	1.128	1.079 (0.010)	Mega	0	26
Crohn's disease	1.185	1.122	1.059 (0.008)	Mega	0	26
Schizophrenia	1.613	1.484	1.070 (0.010)	Mega	0	32
Attention deficit/ hyperactivity disorder	1.033	1.033	1.008 (0.006)	Mega	0	18
Bipolar disorder	1.154	1.135	1.030 (0.008)	Mega	0	23
PGC cross-disorder analysis	1.205	1.187	1.018 (0.008)	Mega	0	29
Major depression	1.063	1.063	1.009 (0.006)	Mega	0	30
Rheumatoid arthritis	1.063	1.033	0.980 (0.007)	Mega	2	20
Coronary artery disease	1.125	1.096	1.033 (0.008)	Meta	1	24
Type 2 diabetes	1.116	1.097	1.025 (0.008)	Meta	1	28
BMI-adjusted fasting insulin	1.088	1.072	1.015 (0.007)	Meta	1	27
Fasting insulin	1.079	1.067	1.021 (0.007)	Meta	1	27
College (yes/no)	1.207	1.180	1.046 (0.009)	Meta	1	31
Years of education	1.220	1.188	1.041 (0.009)	Meta	1	31
Cigarettes per day	1.047	1.047	0.998 (0.008)	Meta	1	21
Ever smoked	1.097	1.083	1.008 (0.006)	Meta	1	21
Former smoker	1.050	1.048	0.999 (0.007)	Meta	1	21
Age of onset (smoking)	1.025	1.030	0.998 (0.006)	Meta	1	21
FN-BMD	1.163	1.109	1.001 (0.009)	Meta	2	25
LS-BMD	1.174	1.112	1.032 (0.009)	Meta	2	25
Waist-hip ratio	1.417	1.330	1.040 (0.008)	Meta	2	16
Height	1.802	1.478	1.149 (0.021)	Meta	2	17
BMI	1.130	1.090	1.033 (0.012)	Meta	2	19

LD Score regression results for all studies analyzed that either did not apply meta-analysis–level genomic control correction or listed λ_{GC} in the relevant publication. For GWAS that applied meta-analysis–level genomic control correction and listed λ_{GC}, we reinflated all test statistics by λ_{GC}. LD Score regression performed on genomic control–corrected summary statistics will generally yield an intercept less than one. Note that genomic control correction at the level of the individual study will also push the expected intercept, in the absence of confounding, slightly below one (**Supplementary Note**). Standard error (SE) estimates were obtained via a block jackknife over blocks of ~2,000 adjacent SNPs, providing a robust estimate of standard error in the presence of correlated, heteroskedastic error terms. BMI, body mass index; FN-BMD, femoral neck bone mineral density; LS-BMD, lumbar spine bone mineral density.
[a]Type of study: mega-analysis (with raw genotypes shared between studies) or meta-analysis (with only summary statistics shared between all contributing studies). [b]The number of rounds of genomic control correction that were performed.

Real data

Finally, we applied LD Score regression to summary statistics from GWAS representing more than 20 different phenotypes[15–32] (**Table 1** and **Supplementary Fig. 8a–w**; metadata about the studies in the analysis are presented in **Supplementary Table 8a**,**b**). For all studies, the slope of the LD Score regression was significantly greater than zero and the LD Score regression intercept was substantially less than λ_{GC} (mean difference of 0.11), suggesting that polygenicity accounts for a majority of the increase in the mean χ^2 statistic and confirming that correcting test statistics by dividing by λ_{GC} is unnecessarily conservative. As an example, we show the LD Score regression for the most recent schizophrenia GWAS, restricted to ~70,000 European-ancestry individuals (**Fig. 2**)[32]. The low intercept of 1.07 indicates at most a small contribution of bias and that the mean χ^2 statistic of 1.613 results mostly from polygenicity. LD Score plots for all other GWAS included in **Table 1** can be found in **Supplementary Figure 8a–w**. As with any inference procedure that relies on a model of genetic architecture, it is possible that our results may be biased by model misspecifications other than those we have simulated directly (for example, if independent effect sizes are a poor model, perhaps because coupled alleles have a tendency to have effects in the same

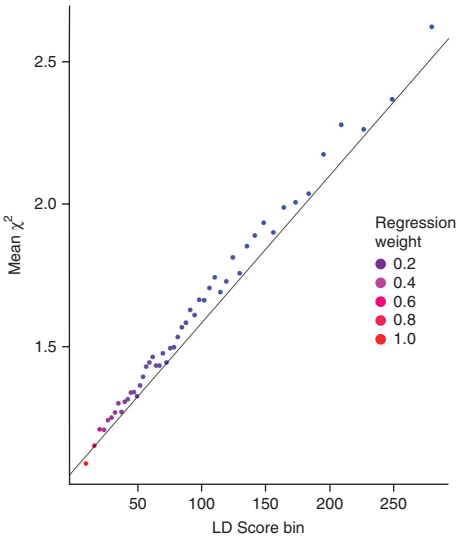

Figure 2 LD Score regression plot for the most recent schizophrenia meta-analysis. Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ^2 statistic of variants in that quantile in the most recent schizophrenia meta-analysis[33]. Colors correspond to regression weights, with red indicating large weight and blue indicating small weight. The black line is the LD Score regression line. The line appears to fall below the points on the right because this is a weighted regression in which the points on the left receive the largest weights (Online Methods).

direction). Such bias might explain the moderate inflation in the LD Score regression intercept that we observed in some large GWAS that are likely well calibrated. Note that upward bias in the LD Score regression intercept means only that the intercept may be conservative as a correction factor.

DISCUSSION

Whenever possible, it is preferable to obtain all relevant genotype data and correct for confounding biases directly[33–37]; *post-hoc* correction of test statistics is no substitute for diligent quality control. However, in the event that only summary data are available or if a conservative correction is desired, we propose that the LD Score regression intercept provides a more robust quantification of the extent of the confounding bias from inflation than λ_{GC} (or intergenic λ_{GC}; **Supplementary Table 9**). Because λ_{GC} increases with sample size in the presence of polygenicity (even without confounding bias)[3], the gain in power obtained by correcting test statistics with the LD Score regression intercept instead of λ_{GC} will become even more substantial for larger GWAS. Extending this method to non-European populations such as East Asians or West Africans is straightforward given appropriate reference panels, but extension to admixed populations is the subject of future research.

In conclusion, we have developed LD Score regression, a method to distinguish between inflated test statistics from confounding bias and polygenicity. Application of LD Score regression to over 20 complex traits confirms that polygenicity accounts for the majority of inflation in test statistics for GWAS results, and this approach can be used to generate a correction factor for GWAS that retains more power than λ_{GC}, especially with large sample sizes. We have made available for download a Python command line tool for estimating LD Score and performing LD Score regression and a database of LD Scores suitable for European-ancestry samples (see URLs). Research in progress aims to apply this method to the estimation of components of heritability, genetic correlation and the calibration of mixed-model association statistics.

URLs. Software tool for LD Score estimation and estimation of variance components from summary statistics, https://github.com/bulik/ldsc/; 1000 Genomes Project genetic map and haplotypes, http://mathgen.stats.ox.ac.uk/impute/data_download_1000G_phase1_integrated.html; LD Score, database: ftp://atguftp.mgh.harvard.edu/brendan/1k_eur_r2_hm3snps_se_weights.RDS; GIANT Consortium (anthropometric traits) summary statistics, http://www.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files; Psychiatric Genomics Consortium (PGC) and TAG (tobacco) summary statistics, https://pgc.unc.edu/Sharing.php#SharingOpp; IIBDGC (inflammatory bowel disease) summary statistics (note that these summary statistics are from meta-analysis of Immunochip data, which are not appropriate for LD Score regression), http://www.ibdgenetics.org/downloads.html; CARDIoGRAM (coronary artery disease) summary statistics, http://www.cardiogramplusc4d.org/downloads/; DIAGRAM (type 2 diabetes) summary statistics, http://diagram-consortium.org/downloads.html; rheumatoid arthritis summary statistics, http://www.broadinstitute.org/ftp/pub/rheumatoid_arthritis/Stahl_etal_2010NG/; blood pressure summary statistics, http://www.georgehretlab.org/icbp_088023401234-9812599.html; MAGIC (glycemic traits) summary statistics, http://www.magicinvestigators.org/downloads/; GEFOS (bone mineral density) summary statistics, http://www.gefos.org/?q=content/data-release; SSGAC (educational attainment) summary statistics, http://ssgac.org/Data.php.

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

We would like to thank P. Sullivan for helpful discussion. This work was supported by US National Institutes of Health grants F32 HG007805 (P.-R.L.), R01 HG006399 (A.L.P.), R03 CA173785 (H.K.F.) and R01 MH094421 (PGC) and by the Fannie and John Hertz Foundation (H.K.F.). Data on coronary artery disease and myocardial infarction were contributed by CARDIoGRAMplusC4D investigators and were downloaded from Psychiatric Genomics Consortium.

AUTHOR CONTRIBUTIONS

B.K.B.-S. conceived the idea, analyzed the data, performed the analyses and drafted the manuscript. B.M.N. conceived the idea and drafted the manuscript. M.J.D. conceived the idea and supplied reagents. N.P. conceived the idea and supplied reagents. A.L.P. conceived the idea and supplied reagents. P.-R.L. analyzed the data and performed the analyses. H.K.F. analyzed the data and performed the analyses. S.R. analyzed the data and performed the analyses. J.Y. provided software. All authors provided input and revisions for the final manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Pritchard, J.K. & Przeworski, M. Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* **69**, 1–14 (2001).
2. Sham, P.C., Cherny, S.S., Purcell, S. & Hewitt, J.K. Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *Am. J. Hum. Genet.* **66**, 1616–1630 (2000).
3. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
4. Voight, B.F. & Pritchard, J.K. Confounding from cryptic relatedness in case-control association studies. *PLoS Genet.* **1**, e32 (2005).

5. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
6. Lin, D.Y. & Sullivan, P.F. Meta-analysis of genome-wide association studies with overlapping subjects. *Am. J. Hum. Genet.* **85**, 862–872 (2009).
7. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
8. Yin, P. & Fan, X. Estimating R^2 shrinkage in multiple regression: a comparison of different analytical methods. *J. Exp. Educ.* **69**, 203–224 (2001).
9. Ralph, P. & Coop, G. The geography of recent genetic ancestry across Europe. *PLoS Biol.* **11**, e1001555 (2013).
10. Bersaglieri, T. *et al.* Genetic signatures of strong recent positive selection at the lactase gene. *Am. J. Hum. Genet.* **74**, 1111–1120 (2004).
11. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet.* **5**, e1000471 (2009).
12. Price, A.L. *et al.* The impact of divergence time on the nature of population structure: an example from Iceland. *PLoS Genet.* **5**, e1000505 (2009).
13. International Multiple Sclerosis Genetics Consortium & Wellcome Trust Case Control Consortium 2. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214–219 (2011).
14. Splansky, G.L. *et al.* The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. *Am. J. Epidemiol.* **165**, 1328–1335 (2007).
15. Sullivan, P.F. *et al.* Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol. Psychiatry* **14**, 359–375 (2009).
16. Heid, I.M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* **42**, 949–960 (2010).
17. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
18. Neale, B.M. *et al.* Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J. Am. Acad. Child Adolesc. Psychiatry* **49**, 884–897 (2010).
19. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
20. Stahl, E.A. *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.* **42**, 508–514 (2010).
21. Tobacco & Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42**, 441–447 (2010).
22. International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
23. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near *ODZ4*. *Nat. Genet.* **43**, 977–983 (2011).
24. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).
25. Estrada, K. *et al.* Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat. Genet.* **44**, 491–501 (2012).
26. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012).
27. Manning, A.K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **44**, 659–669 (2012).
28. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
29. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371–1379 (2013).
30. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18**, 497–511 (2013).
31. Rietveld, C.A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467–1471 (2013).
32. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
33. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
34. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
35. Kang, H.M. *et al.* Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* **42**, 348–354 (2010).
36. Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nat. Methods* **8**, 833–835 (2011).
37. Korte, A. *et al.* A mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat. Genet.* **44**, 1066–1071 (2012).

ONLINE METHODS

Estimation of LD Score. We estimated European LD Scores from 378 phased European-ancestry individuals (excluding one individual from a pair of cousins) from the 1000 Genomes Project reference panel using the --ld-mean-rsq option implemented in the GCTA[38] software package (with flags --ld-mean-rsq --ld-rsq-cutoff 0 --maf 0.00001. We implemented a 1-cM window using the --ld-wind flag and modified .bim files with physical coordinates replaced with genetic coordinates as described below; note that a 1-cM window can be achieved more conveniently using the flags --l2 and --ld-wind-cm in the LDSC software package by the authors). The primary rationale for using a sequenced reference panel containing several hundred individuals for LD Score estimation rather than a genotyped GWAS control panel with several thousand individuals was that, even after imputing off-chip genotypes, the variants available from a genotyping array only account for a subset of all variants. Using only a subset of all variants for estimating LD Score produces estimates that are biased downward.

We used a window size of 1 cM around the index variant for the sum of r^2 values (using the genetic map and phased genotypes from the IMPUTE2 website; see URLs), did not set an r^2 cutoff and excluded singletons (MAF < 0.13%). The standard estimator of the Pearson correlation coefficient has an upward bias of approximately $1/N$, where N is sample size, so we employed an approximately unbiased estimator of LD Score given by:

$$r^2_{\text{adj}} = \hat{r}^2 - \frac{1-\hat{r}^2}{N-2}$$

where $\hat{r}^2$ denotes the standard, biased estimator of the squared Pearson's correlation. Note that it is possible to have $r^2_{\text{adj}} < 1$, which is a mathematically necessary feature of any unbiased estimator of r^2. Thus, some estimated LD Scores will be less than one. In practice, almost all variants with an estimated LD Score of less than one were rare: only 0.01% of variants with MAF > 5% had estimated LD Scores below one.

We examined the effect of varying the window size on our estimates of LD Score and found that our estimates of LD Score were robust to choice of window size. The mean difference in the LD Scores estimated with a 1-cM window and a 2-cM window was less than 1% of the mean LD Score (**Supplementary Fig. 9**), and all LD Scores estimated with window sizes greater than 1 cM had squared correlations of >0.99 (**Supplementary Table 10**). This observation also addresses concerns about inflation in the LD Score from the intra-European population structure in the 1000 Genomes Project reference panel. The mean inflation in the 1-cM LD Score from population structure can approximately be bounded by the mean difference between a 1-cM LD Score and a 2-cM LD Score. Because this difference is <1% of the mean LD Score, we conclude that bias from population structure does not significantly inflate our estimates of LD Score.

We estimated the standard errors for LD Scores via a delete-one jackknife over the 378 phased individuals in the 1000 Genomes Project EUR reference panel. We found that the standard error for LD Score was positively correlated with MAF and with LD Score itself. Jackknife estimates of LD Score standard error became extremely large for variants with MAF < 1%, so we excluded variants with 1000 Genomes Project EUR sample MAF < 1% from all LD Score regressions.

Intra-European LD Score differences. To quantify the magnitude of intra-European differences in LD Score, we estimated LD Scores using each of the 1000 Genomes Project EUR subpopulations CEU, GBR, TSI and FIN. The LD Scores from the four subpopulations were all highly correlated, but the mean LD Score was not constant across populations. The mean LD Scores (MAF > 1%) were as follows: EUR, 110; CEU, 109; GBR, 104; FIN, 117; TSI, 102. The observation that the mean LD Score in the FIN population was elevated is consistent with a recent bottleneck in the genetic history of Finland[39], and the observation that the mean LD Score in the southern European TSI population was lower is consistent with reports that southern European populations have gone through less severe bottlenecks than northern European populations.

Intra-European differences in LD Score can be a source of bias in the LD Score regression intercept. For instance, if one attempts to perform LD Score regression using the 1000 Genomes Project EUR LD Score on a GWAS with all samples from Finland, then the LD Score regression intercept may be biased upward. Similarly, if one attempts to perform LD Score regression using the 1000 Genomes Project EUR LD Score on a GWAS with all samples from Italy, the LD Score regression intercept may be biased downward. If we make the approximation that the intra-European differences in LD Score can be described by an additive term plus 5% noise (that is, if we assume that the FIN LD Score equals the pan-European LD Score plus seven, which is a worst-case scenario among the linear relationships between two LD Scores in terms of bias in the intercept), then the bias introduced into the LD Score regression intercept by using the pan-European LD Score to perform LD Score regression on a Finnish GWAS will be seven multiplied by the slope of the LD Score regression plus 5% of the mean χ^2 value minus one, where seven is the difference between the reference population LD Score and the GWAS population LD Score. Because all of the mean EUR subpopulation LD Scores that we have estimated are within ±8 of the mean pan-European LD Score, we estimate that the bias in the LD Score regression intercept from intra-European LD Score differences is at most ±10 times the LD Score regression slope. For the real GWAS analyzed in **Table 1**, this corresponds to a worst-case difference of approximately ±10% in the estimate of the proportion of the inflation in the mean χ^2 statistic that results from confounding bias, with a higher probability of upward bias (because the noise term in the relationship between target and reference LD Scores always causes upward bias in the LD Score regression intercept, whereas systematic directional differences in the target and reference LD Scores can bias the LD Score regression intercept in either direction).

Regression weights. To produce an efficient regression estimator, we must deal with two problems. First, χ^2 statistics at SNPs in LD are correlated. Second, the χ^2 statistics of variants with high LD Scores have higher variance than the χ^2 statistics of variants with low LD Scores (heteroskedasticity).

The statistically optimal solution to the correlation problem is to perform generalized least squares (GLS) with the variance-covariance matrix of χ^2 statistics. However, this matrix is intractable under our model. As an approximation, we correct for correlation by weighting variant j by the reciprocal of the LD Score of variant j, counting LD only with other SNPs included in the regression. Precisely, if we let S denote the set of the variants included in the LD Score regression, then the LD Score of variant j counting LD only with other SNPs included in the regression is as follows:

$$\ell_j(S) = 1 + \sum_{k \in S} r^2_{jk}$$

Weighting by $1/\ell_j(S)$ would be equivalent to GLS with the full variance-covariance matrix of χ^2 statistics if the genome consisted of LD blocks and r^2 (in the population) was either zero or one. We estimate $\ell_j(S)$ for the set of variants S described in "Real data" using the same procedure we used to estimate the full 1000 Genomes Project LD Score. Because our estimates of $\hat{\ell}_j(S)$ can be negative and regression weights must be positive, we weight by $1/\max(\hat{\ell}_j, 1)$.

To account for heteroskedasticity, we weight by

$$(1 + Nh^2_g \ell_j / M)^2$$

which is the reciprocal of the conditional variance function $\text{var}[\chi^2_j \mid \ell_j]$ under our model if we make the additional assumption that the effect sizes for each normalized genotype are normally distributed (note that violation of this assumption does not bias the regression but only increases the standard error; a derivation is provided in the **Supplementary Note**).

Attenuation bias. Standard least-squares and weighted least-squares regression theory assumes that the explanatory variable (also referred to as the independent variable, or X) is measured without error. If the explanatory variable is measured with error, then the magnitude of the regression slope will be biased toward zero. This form of bias is known as attenuation bias. If the explanatory variable is measured with error but the variance of this error is known, then it is possible to produce an unbiased regression slope by multiplying the slope by a disattenuation factor, which is equal to the squared weighted Pearson correlation between the noisy estimates of the explanatory variable and the true value of the explanatory variable.

doi:10.1038/ng.3211

Simulations. When performing simulations with polygenic genetic architectures using genotyped or imputed data, variants in the 1000 Genomes Project reference panel not included in the set of genotypes used for simulation cannot contribute to the simulated phenotypes and thus should not contribute to the LD Score used for simulations. Precisely, for the simulations with polygenicity and the simulations with polygenicity and bias, we used LD Scores where estimates of r^2 were derived from the 1000 Genomes Project EUR reference panel but the sum of the r^2 values was taken over only those SNPs included in the simulations. For the simulations with frequency-dependent genetic architecture, we estimated the LD Scores from the same genotypes used for the simulations because we wanted to quantify the bias introduced by frequency-dependent genetic architecture even when the LD Scores are estimated with little noise. For the simulations with pure population stratification, we used an LD Score estimated from all 1000 Genomes Project variants, as there was no simulated polygenic architecture in these simulations. For simulations with pure population stratification, the details of the cohorts used are given in **Supplementary Table 1**.

It is difficult to use real genotypes to simulate ascertained studies of a binary phenotype with low population prevalence: to obtain 1,000 cases for a phenotype with a simulated prevalence of 1%, one would need to sample on expectation of 100,000 genotypes, which is not feasible. We therefore generated simulated genotypes at 1.1 million SNPs with a mean LD Score of 110 and a simplified LD structure where r^2 was either 0 or 1 and with all variants having a MAF of 50%. We generated phenotypes under the liability threshold model with all effect sizes per normalized genotype (effects on liability) drawn independently and identically from a normal distribution and then sampled individuals at random from the simulated population until the desired number of cases and controls for the study had been reached.

Application to real data. The majority of the sets of summary statistics that we analyzed did not contain information about sample MAF or imputation quality. To restrict to a set of common, well-imputed variants, we retained only those SNPs in the HapMap 3 reference panel[40] for the LD Score regression. To guard against underestimation of LD Score from summing only LD with variants within a 1-cM window, we removed variants in regions with exceptionally long-range LD[41] from the LD Score regression (note that LD for these variants was included in the estimation of LD Score). Lastly, we excluded pericentromeric regions (defined as ±3 cM from a centromere) from the LD Score regression because these regions are enriched for sequence gaps, which may lead to the underestimation of LD Score, and depleted for genes, which may reduce the probability of association with a phenotype[42,43]. The final set of variants retained for LD Score regression on real data consisted of approximately 1.1 million variants.

38. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
39. Jakkula, E. *et al.* The genome-wide patterns of variation expose significant substructure in a founder population. *Am. J. Hum. Genet.* **83**, 787–794 (2008).
40. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
41. Price, A.L. *et al.* Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83**, 132–135, author reply 135–139 (2008).
42. Smith, A.V., Thomas, D.J., Munro, H.M. & Abecasis, G.R. Sequence features in regions of weak and strong linkage disequilibrium. *Genome Res.* **15**, 1519–1534 (2005).
43. She, X. *et al.* The structure and evolution of centromeric transition regions within the human genome. *Nature* **430**, 857–864 (2004).