

Local joint testing improves power and identifies missing heritability in association studies

Brielin C. Brown,^{*1} Alkes L. Price,² Nikolaos A. Patsopoulos,^{†3,4,5} and Noah Zaitlen^{†*6}

¹Department of Computer Science, University of California Berkeley, Berkeley, CA

²Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA

³Department of Neurology, Brigham & Women's Hospital, Boston, MA

⁴Division of Genetics, Department of Medicine, Brigham & Women's Hospital, Boston, MA

⁵Broad Institute, Cambridge, MA

⁶Department of Medicine, University of California San Francisco, San Francisco, CA

Abstract

There is mounting evidence that complex human phenotypes are highly polygenic, with many loci harboring multiple causal variants, yet most genetic association studies examine each SNP in isolation. While this has led to the discovery of thousands of disease associations, discovered variants account for only a small fraction of disease heritability. Alternative multi-SNP methods have been proposed, but issues such as multiple testing correction, sensitivity to genotyping error, and optimization for the underlying genetic architectures remain. Here we describe a local joint testing procedure, complete with multiple testing correction, that leverages a genetic phenomenon we call linkage masking wherein linkage disequilibrium between SNPs hides their signal under standard association methods. We show that local joint testing on the original Wellcome Trust Case Control Consortium dataset leads to the discovery of 29% more associated loci that were later found in followup studies containing thousands of additional individuals. These loci double the heritability explained by genome-wide significant associations in the WTCCC dataset, implicating linkage masking as a novel source of missing heritability. Furthermore, we show that local joint testing in a *cis*-eQTL study of the gEUVADIS dataset increases the number of genes discovered by 10.7% over marginal analyses. Our multiple hypothesis correction and joint testing framework are available in a python software package called *jester*, available at github.com/brielin/Jester.

Introduction

Genetic association studies typically take a marginal approach to analysis; investigating each SNP in isolation of all other SNPs for association with a phenotype of interest. While this method has led to the discovery of thousands of loci associated with hundreds of phenotypes,^{1,2} it fails to capture the additional signal available when multiple SNPs representing independent genetic signals are examined simultaneously,³ or when SNPs are imperfectly imputed.⁴ Furthermore, the difference between the heritability due to genome-wide significant associations and heritability due to genotyped variants remains substantial.⁵ In this work we investigate a local joint testing approach to analysis of genetic data sets in which pairs of variants from the same locus are examined simultaneously for association with phenotype. The motivation for our approach comes from the mounting evidence that complex traits are highly polygenic,⁶ that causal variants are not evenly distributed across the genome,⁷ that known associated loci often harbor multiple causal variants,^{8,9,10,11,12,13} and that the underlying causal variants can be in linkage disequilibrium (LD) with each other.¹⁴

In fact, LD between underlying causal variants can result in additive associations that would be nearly impossible to detect using standard marginal methods. Consider the case of two SNPs: one risk-increasing

*Address correspondence to: brielin@berkeley.edu or noah.zaitlen@ucsf.edu

†Co-last author

for a disease, and the other protective. If these SNPs are correlated in the study population then marginal association methods will fail to detect the signal due to the large number of individuals carrying both variants (and therefore having little or no increased risk for the disease). In the context of this paper, we will refer to these SNP pairs as *linkage masked*, and we note that this phenomenon may be quite common due to the Bulmer effect,[?] or to correlated tagging of an untyped causal variant. Furthermore, Lappalainen et al.[?] give evidence that linkage masking between regulatory and coding variation may be common due to balancing selection. While linkage masked SNPs are difficult to uncover using standard marginal association methods, mixed model heritability is determined by a simultaneous fit of all SNPs while accounting for LD and therefore includes signal from linkage masked SNPs, implicating them as a source of missing heritability which has not been widely considered.[?]

Pairwise (joint) testing may help unmask these associations and, more generally, improve power in the presence of statistical interactions, multiple causal variants or multiple variants differentially tagging an untyped causal SNP. However, applying joint tests in practice has several problems. Because exact multiple testing correction is usually unknown, several studies have used joint testing for follow up and fine mapping of known associated loci, often revealing additional associated variants,^{?,?} and demonstrating the merits of joint testing in practice. Studies such as these are able to ignore multiple hypothesis correction issues due to their focus on known associated regions but do not have the potential to reveal novel loci. Other studies examining genome-wide joint testing approaches of all pairs of SNPs, including those with statistical interaction terms, pay such severe multiple hypothesis correction penalties that many loci found via standard marginal approaches would not reach genome-wide significance.^{?,?} Slavin and Elston[?] proposed testing all adjacent pairs and applied their approach in the WTCCC seven disease study. Howey and Cordell (SnipSnip)[?] proposed using a conditional test on 10 adjacent SNPs to choose a partner SNP for inclusion in the linear model. While these approaches reduce the multiple hypothesis correction penalty we show that they do not capture much of the available power gain. Furthermore, prior approaches have not accounted for a known issue with genotyping error and joint tests,[?] which we show impacts these methods.

By testing pairs of SNPs rather than individual SNPs, we improve power to detect loci containing multiple causal variants, including those containing linkage masked SNPs. Through local testing, we substantially reduce the multiple hypothesis correction penalty, while simultaneously enriching for situations in which joint tests are more powerful, i.e. when there are independent genetic signals contained in each of the SNPs in the test. Rather than employing the overly conservative Bonferroni procedure to account for multiple testing, we extend the work of Han et al.[?] to provide a method for estimating the null distribution of joint tests orders of magnitude faster than a permutation test, making application computationally efficient. We applied our method to the WTCCC[?] cohorts for bipolar disorder, coronary artery disease, Crohn's disease, hypertension, rheumatoid arthritis, type-1 diabetes and type-2 diabetes. We define *significant locus* as any 1 megabase region containing a statistically significant association at family-wise error rate (FWER) 5% and compare the number of significant loci discovered from the NHGRI database of replicated associations to the standard marginal method. We compare our approach to SnipSnip[?] and marginal analysis of imputed WTCCC genotypes. We also estimate the heritability explained by associated SNPs discovered in the marginal and joint approaches. Finally, we apply our method to gene expression data from the gEUVADIS project, comparing the number genes containing *cis*-eQTLs at various false discovery rate (FDR) thresholds under marginal and joint testing approaches.

We find: 1) Local joint testing provides significant power gains when multiple risk variants are proximal, reaching as high as 41% when they are linkage masked. 2) Local joint testing in the original WTCCC cohort discovers seven loci not found via the standard marginal approach, all of which were later replicated in more powerful followup studies. Marginal analysis of imputed genotypes discovers only two of these loci, as well as one locus not detectable in either genotype-based method. 3) New SNPs and loci discovered via local joint testing represent a substantial amount of the heritability of these diseases. Our method uncovers roughly double the heritability due to genome-wide significant associations on average across the diseases of interest, and 4) joint testing all pairs of *cis*-SNPs in gEUVADIS reveals 607 more genes at FDR 5%, an increase of 10.7% over our marginal approach.

Methods

We begin by describing the null distributions of the tests in our procedure in order to motivate the sampling approach used to determine the multiple testing correction. We then show that given the marginal Z -scores at two SNPs the joint χ^2 test statistic can be exactly determined, leading to a substantial speed-up after integration with the conditional sampling method SLIDE.[?] We then describe the local joint testing procedure, and the related joint-testing based QC protocol. For simplicity, we assume the phenotype and genotypes have been standardized.

Asymptotic Distribution of Marginal and Joint Tests

For many widely used statistical tests, the vector of test statistics over many markers asymptotically follows a multivariate normal distribution (MVN) under the null hypothesis of no association.^{?,?} In particular, let Y be the phenotype of interest and G the genotype matrix, with G_i the genotype at SNP i in a study with N individuals. Then the Wald test $Z_i = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}} = \sqrt{N} \frac{Cor(x,y)}{\hat{\sigma}}$ is asymptotically $\mathcal{N}(0,1)$. From this, one can derive the correlation structure for two tests under the null,[?] $\rho(Z_i, Z_j) = \rho(G_i, G_j) := \rho_{ij}$, so that the vector of marginal test statistics is asymptotically MVN with mean 0 and covariance matrix $\Sigma = \{\Sigma_{ij}\}_{i=1,j=1}^{N,N} = \{\rho_{ij}\}_{i=1,j=1}^{N,N}$. Next, consider the value of the likelihood ratio test statistic for a linear or logistic two-SNP association test. In the appendix we prove that the joint test statistic for SNPs i and j is a function only of their marginal Z -scores and correlation: $\chi^2_J = \frac{1}{1-\rho_{i,j}^2} (Z_i^2 + Z_j^2 - 2\rho_{i,j}Z_iZ_j)$. We verified computationally by simulating pairs of SNPs at all correlation levels that this equation is exact. Combining this with the result of Han et al.[?] that their conditional sampling approach is a close approximation to the null distribution of the marginal tests shows that our method is also well calibrated.

Estimating the Significance Threshold and Local Joint Testing

We use a conditional sampling method to sample the marginal test statistics from MVN. Since distal SNPs are likely to be independent, we choose a window size W_z and ‘slide’ along the genome, sampling SNP i conditional on the correlation with the previous W_z SNPs.[?] Specifically, the MVN factors as $f(Z_1, \dots, Z_L) \approx f(Z_1)f(Z_2|Z_1) \dots f(Z_{L-1}|Z_{L-2} \dots Z_{L-1-W_z})f(Z_L|Z_{L-1} \dots Z_{L-W_z})$ and we can sample $Z_i|Z_{i-1} \dots Z_{i-W_z}$ via the standard conditional MVN $Z_i|Z_p \sim \mathcal{N}(\Sigma_{i,p}\Sigma_{p,p}^{-1}Z_p, 1 - \Sigma_{i,p}\Sigma_{p,p}^{-1}\Sigma_{i,p}^\top)$ where $\Sigma_{i,p}$ is the vector of correlations between SNP i and the conditional SNPs, $\Sigma_{p,p}$ is the $W_z \times W_z$ correlation matrix for the conditional SNPs, and Z_p represents their sampled values.

Given the above, we are able to estimate the significance threshold for our joint testing framework. The algorithm, presented in Algorithm ?? (Appendix), takes as an input a desired FWER, number of samples (roughly analogous to the number of permutations), a window size, and a set of joint tests T , and outputs a multiple-testing corrected significance level corresponding to the desired FWER. In practice, we choose a new joint testing window size $W_j \leq W_z$ and a correlation threshold ρ_{min} and test all pairs with correlation exceeding ρ_{min} in the window. With the multiple testing correction in hand, the local joint testing procedure is a straightforward modification to the standard GWAS procedure. We choose a window size W_j , correlation cutoff ρ_{min}^2 and fit the model (??) for every pair of SNPs exceeding correlation ρ_{min}^2 within W_j markers of each other (Algorithm ??, Appendix). This procedure is implemented in a python package called *jester* (github.com/brielin/Jester).

Filtering False Positives

Lee et al.[?] describe an issue where genotyping errors that would go unnoticed in standard QC procedures can cause inflation in joint and conditional tests of association. When SNPs are highly correlated, mis-called bases in only the cases or controls induce rare haplotypes. As these haplotypes are only present in the cases or controls, this increases the association signal in the joint test.[?]

While performing our analysis, we found many highly correlated ($|\rho| > 0.9$) pairs of SNPs where neither SNP had much marginal signal but together showed an extremely strong association. We accounted for this in two ways: first, we considered only associations arising from pairs with correlation less than 0.9, second, we used imputation against 1000 genomes to reanalyze jointly significant SNPs.

For each pair of potentially significant SNPs, we replaced the raw genotype with the imputed genotype and re-computed the test statistic. When the signal was a true association, the joint test statistic remained significant after imputation (Table S2). However when the association appeared to be driven by genotyping error, the joint test statistic became insignificant after imputation (Table S7). In this way, we overcome the false positive error identified by Lee et al.[?]

Datasets

We analyzed the WTCCC phenotypes bipolar disorder, Crohns disease, coronary artery disease, hypertension, rheumatoid arthritis, type-1 diabetes and type-2 diabetes (CD, CAD, HT, RA, T1D, T2D). We chose this data set because it was one of the first GWAS performed and the phenotypes have been subsequently studied in independent large scale GWAS. Thus, we emphasize the potential of early discovery of true effect leveraging non-standard GWAS methods. We used a window size of $W_z = W_j = 100$ SNPs for estimating the null distribution, and a correlation cutoff of $\rho_{min}^2 = 0$ (all pairs in the window). We performed standard QC on the data, removing individuals with missingness > 0.1 , SNPs with missingness > 0.1 , markers failing a Hardy-Weinberg equilibrium (HWE) test at significance level 0.001, and SNPs with minor allele frequency < 0.05 . The number of cases, controls and SNPs left after QC for each dataset are presented in Table S8. To impute the WTCCC cohort, genotypes were split into 2 mega-base regions and pre-phased against the 1000 Genomes EUR reference panel using HAPI-UR, then imputed using Impute2 against the same reference panel. Non-biallelic SNPs and SNPs with reference panel frequency below 5% were not imputed. All imputed SNPs with info score below 0.5 were excluded from further analysis.

We also analyzed gene expression data for 16155 genes of the the gEUVADIS European dataset. Raw RNA-sequencing reads obtained from the European Nucleotide Archive were aligned to the transcriptome using UCSC annotations matching hg19 coordinates. RSEM was used to estimate the abundances of each annotated isoform and total gene abundance is calculated as the sum of all isoform abundances normalized to one million total counts or transcripts per million (TPM). Genotyping data was obtained from the 1000 Genomes Phase III public release. eQTL mapping was performed on a per-gene basis. The *cis* region of each eQTL was defined as all SNPs with MAF $> 5\%$ within 200KB of the transcription start site (TSS), which was chosen because the vast majority of eQTLs are known to be contained in this region.[?] Joint tests were performed between all pairs of SNPs in the *cis* region. In each analysis, 30 genotype principal components were included as covariates. Approximate permutation tests from our sampling procedure with 2500 samples were used to infer permuted p-values separately for the marginal and joint approaches, which were then independently analyzed to determine the number of significant genes at FDR 1%-25%.

Statement on Data Availability

All genotypes and phenotypes used in the analysis are publicly available. Software used to analyze genetic data is available at github.com/brielin/Jester. Supplemental Data include 8 tables in a excel spreadsheet. Table S1 contains details of each locus discovered in the marginal analysis. Table S2 is the same for jester with window size 100. Table S3 is the same for jester with window size 50 and correlation cutoff of 0.04. Table S4 contains SnipSnip results. Table S5 contains results of a marginal analysis of imputed WT genotypes. Table S6 contains imputation analysis of the pairs reported in Slavin and Elston19 that passed QC. Table S7 is a sampling of false positives. Table S8 describes the SNP and sample numbers remaining in our datasets after QC. Supplemental tables are available at github.com/brielin/jester_supplemental_tables. Scripts used in simulations available upon request.

Results

Multiple Testing Penalties for WTCCC and Power

We computed the multiple testing correction for the local joint testing method on WTCCC using *jester* for various window sizes and ρ_{min}^2 cutoffs. In this work we chose to use a window size of 100 and ρ_{min}^2 cutoff of 0 for our main analysis, increasing the effective number tests by a factor of 19.55 over the marginal test, even though we perform 100 times as many tests. We chose the window size based on the work of Han et al.,[?] and used no correlation cutoff because our power simulations showed an increase in power even in the

absence of LD (Figure 1, right). For smaller window sizes and cutoffs, the multiple testing burden decreases substantially. Using a modest ρ_{min}^2 cutoff of 0.004, for example, increases the number of tests by a factor of 8.90 over the marginal test.

We used these significance thresholds to estimate the power of local joint testing in the case of 1) two correlated variants in LD and 2) a single causal variant.. We found a substantial increase in power for modest window sizes and small correlation cutoffs (Figure 1, right) when there were multiple causal variants. Using a window size 50 and cutoff of 0.004 lead to an increase in power of up to 41.0% when there were multiple correlated causal variants, while a window size of 100 and cutoff of 0.0 saw an increase in power of 35.4%. In the absence of multiple causal variants, the increase in multiple testing burden and degree-of-freedom penalty gave a decrease in power of 15% and 20%, respectively, for all correlation levels. Note that while our method shows its most substantial gain when SNPs are linkage masked, we also see a 25% increase in power when the causal variants are uncorrelated.

Significant Loci in WTCCC

We analyzed the WTCCC phenotypes using four different methods: 1) marginal tests at an FWER of 5% (2.3×10^{-7}), 2) joint tests with window size 100 and $\rho_{min}^2 = 0$ at the significance level estimated by *jester* corresponding to an FWER of 5% (1.18×10^{-8}), 3) SnipSnip⁷ with window size of 10 SNPs at significance level 5×10^{-8} and 4) marginal tests of imputed genotypes at significance level 5×10^{-8} . Since SnipSnip does not include an analysis of the multiple testing correction we chose to use their suggested significance threshold.

Local joint testing resulted in the discovery of 2.3 times as many associated SNPs over the marginal method, summarized in Table ???. The marginal test of genotyped SNPs revealed 17 significant loci. *jester* discovered seven additional loci while missing two of the original due to the increased multiple testing burden, for a total of 22 significant loci. For each of these seven newly significant loci, we searched the NHGRI GWAS database⁷ for reported associations and found that each had been reported in more powerful disease-specific follow-up studies. The significant SNP pairs in four of these seven novel loci were linkage masked, with correlations ranging from -0.74 to -0.26 (signed w.r.t same-direction SNP effects). Interestingly, SnipSnip discovered fewer SNPs and loci than the marginal method, but we emphasize that the set of loci it uncovered are not a strict subset of those discovered via the marginal method (Table S4). That is, it discovers some new loci while missing some that are marginally significant and thus remains useful as a secondary analysis tool.

We also compared our method to a GWAS of imputed genotypes, the current gold standard method. This allows us both to determine how our method compares in the number of discovered loci, and whether the linkage masked loci that *jester* uncovered were due to anti-correlated tagging of an untyped causal variant. The imputed GWAS discovered 20 loci in total: the 17 marginally discovered loci, two of the seven linkage masked loci (Table ??) , and one locus found by SnipSnip but not *jester* or the marginal GWAS (CD 18p11.21, Table S4-5). The two loci discovered by both *jester* and imputation (CD 5q31.1 and T1D 10p15.1) were linkage masked, but the discovery of a significant SNP after imputation implies this was due to anti-correlated tagging of an untyped causal. Of the remaining five loci, three (CD 10q21.3, RA 1p36.32, RA 10p15.1) were uncorrelated, supporting the presence of multiple causal variants. The final two loci (CD 6p21.32, T2D 9p21.3) were anti-correlated ($r = -0.26$ and $r = -0.51$, respectively) but did not contain a significant SNP after imputation. We conclude that these loci are strong candidates for followup study to validate the presence of linkage masking. For complete details of all associations discovered in each method, see Tables S1-S5.

We sought to quantify the effect of newly significant SNPs on the heritability due to genome-wide significant associations. We computed the genetic relationships matrix (GRM) for each disease using only SNPs identified as significant using either the classic marginal or local joint testing method. Since many of these SNPs are in the same locus and thus likely to be highly correlated, standard estimates of heritability can be biased.⁷ To counter this we used the LD residual correction implemented in Eigensoft 5.0 to compute the GRM.⁷ We find that LJT nearly doubles the heritability captured by significant SNPs (Figure 2).

gEUVADIS eQTL Analysis

We compared the number of genes containing an eQTL at an FDR of 1%-25% using standard marginal linear regression of all *cis*-SNPs against joint tests of all *cis*-SNPs (Table ??). At an FDR of 5%, we find that 5641 of 16155 genes contain an eQTL using the marginal test, and 6248 genes contain an eQTL using the joint test, an increase of 10.7%. As in our analysis of the Wellcome Trust data, the genes discovered using the marginal approach are not a strict subset of those discovered using the joint approach. For each level of FDR, we determined the proportion of genes uncovered in the marginal but not joint approach that appear to be linkage masked. At FDR 5%, the joint testing approach discovers 908 new genes. In 381 of those 908 genes the significant SNP pair have a correlation of greater than 0.2 (Table ??).

Discussion

In this work we described a local joint testing procedure, its multiple testing correction procedure, the genetic architecture for which it is well powered, a system for reducing susceptibility to genotyping error, and implications for heritability explained from GWAS SNPs. We have shown that when loci harbor multiple causal variants, the joint test can outperform the marginal test substantially. We observed that our method out-performs the standard marginal association method, lending further evidence that disease loci frequently harbor multiple causal mutations. The largest power gains for our method come when SNPs are linkage masked. The Bulmer effect implies linkage masking should be common; high fitness haplotypes are able to resist selective pressures, and may acquire fitness decreasing mutations without being eliminated from the population. SNPs of this kind are hypothesized as a source of missing heritability by Haig et al.⁷ and Lappalainen et al.⁷ argue that gene expression data support widespread linkage masking due to balancing selection.

Our gEUVADIS analysis lends further support to this claim, as we observe a substantial increase in the number of genes containing a *cis*-eQTL using *jester*. Of the 908 new genes containing an eQTL at FDR 5%, 381 of them have absolute correlation of greater than 0.2, implicating the presence of linkage masking in gene regulation. On the other hand, we find only weak evidence that disease loci harbor linkage masked SNPs, observing that two loci containing linkage masked SNPs do not harbor a significant SNP after imputation. It is not surprising that such effects are more difficult to find in a genotyped cohort. As the effects of such SNPs are already masked, the tagging SNP pair must be in tight LD with the causal SNP pair to prevent severe power loss. On top of this, the tagging SNP pair must be highly correlated to achieve the increased signal necessary to find linkage masked SNP pairs. Hemani et al.⁷ make a similar argument, showing that small reductions in LD can result in dramatic under-estimation of epistatic effects.

While linkage masked SNPs are difficult to uncover using standard marginal association methods their signal is included in mixed-model SNP heritability estimates. This shows that these SNPs represent a source of the difference between the heritability due to genome-wide significant associations and the heritability due to genotyped variation. Our result narrows this heritability gap by discovering new associations which increase the heritability explained due to statistically significant associations.

Our approach is not without drawbacks. When causal variants are sparse we see a reduction in power due to the degree of freedom and multiple test correction which only reaps benefits in the presence of multiple signals. While our results indicate this is a less common situation, there are two loci discovered by a marginal GWAS but not by *jester* (Table S1). Additionally, it can be difficult to distinguish heavy linkage masking from genotyping error. In our analysis of the WT disease associations, we used a window size of 100 SNPs, chosen based on results from Han et al.⁷ A logical question is whether or not a larger or smaller window size would lead to more results. We repeated the analysis with a window size of 50 and a correlation cutoff of 0.04, and discovered the same number of loci and slightly more SNPs than presented here (Table S3).

Yang et al.⁷ propose a mathematically similar approach to ours, determining joint test statistics from marginal summary test statistics (albeit only at genome-wide significant marginal loci), while using an external reference panel to estimate the pairwise LD. This approach in combination with our multiple hypothesis correction threshold could provide a way to apply our local method without access to genotype data. We caution that our proposed imputation based genotyping error correction method will not be applicable here and thus high LD SNPs should be avoided in such an analysis. Furthermore, the variance

of the correlation coefficient estimates can be large even when many hundreds of individuals are available in the reference panel⁷ which could lead to false positives. While the reference panel correlations may still be relatively accurate for controls from the same population, case individuals are more likely to harbor many disease-associated mutations and thus will not match the reference panels as well.^{7,8} Even with these caveats, however, the vast gain in power possible with the increased sample size of summary data makes this a tempting proposition, and we have implemented this method in our software package.

Figures and Tables

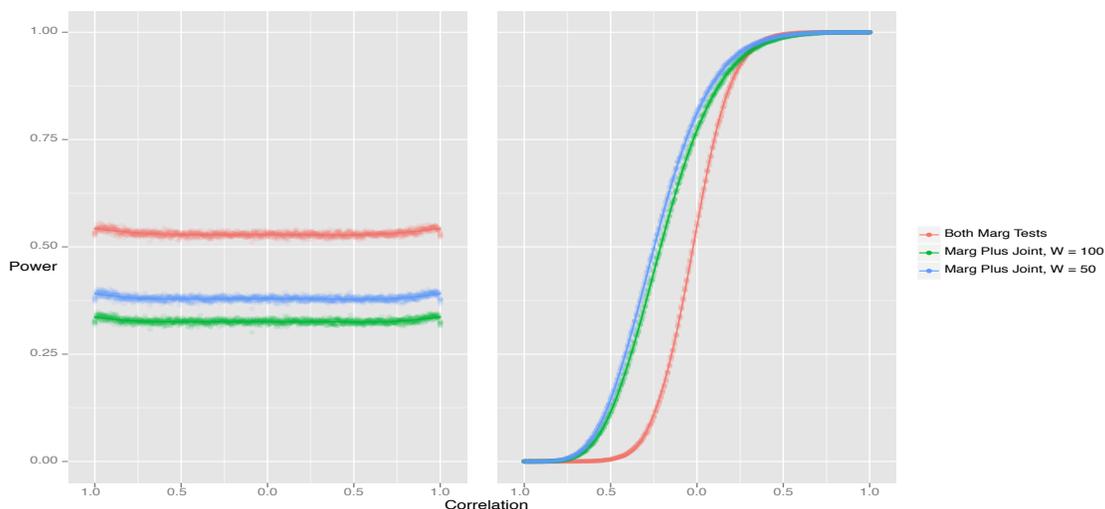


Figure 1: (Left) Joint testing genome-wide shows a power loss for all correlation structures when only one SNP affects the trait. (Right) Joint testing genome-wide shows a substantial power gain for anti-correlated SNPs that both affect the trait.

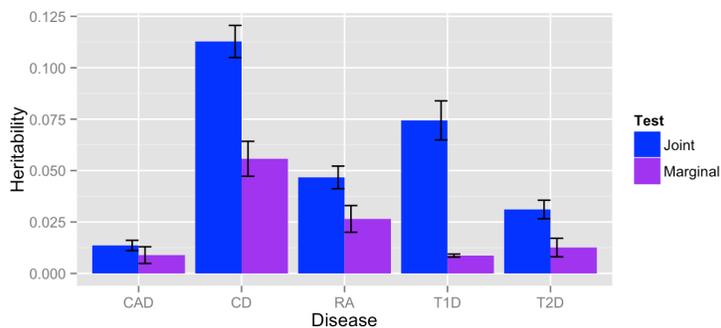


Figure 2: The heritability explained by significantly associated SNPs is nearly doubled by joint testing.

	marg	jester	snip-snip	imputed	marg	jester	snip-snip	imputed
BD	0	0	0	0	0	0	0	0
CAD	1	1	0	1	16	24	0	96
CD	7	8	5	8	58	89	56	587
HT	0	0	0	0	0	0	0	0
RA	2	4	3	3	6	29	6	88
T1D	5	6	3	6	14	82	14	264
T2D	2	3	2	2	16	25	24	76
Total	17	22	13	20	110	249	100	1111

Table 1: (Left) Total number of loci containing genome-wide significant SNPs discovered using standard marginal, local joint, and SnipSnip testing methods. (Right) Total number of genome-wide significant SNPs discovered using standard marginal, local joint, and SnipSnip testing methods. For our analysis of T1D and RA, we removed chromosome 6 because of the large effect HLA locus.

Dis	Locus	marginal		jester				imputation	
		SNP	PV	SNP1	SNP2	ρ	PV	SNP	PV
CD	5q31.1	rs6596075	5.97E-07	rs6596075	rs273913	-0.32	9.34E-09*	rs6897597	3.22E-8*
	6p21.32	rs9469220	9.09E-07	rs9469220	rs12524063	-0.26	2.81E-10*	rs210194	3.359E-6
	10q21.3	rs10761659	2.82E-07	rs10995271	rs10995271	-0.01	3.02E-09*	rs10761659	1.88e-07
RA	1p36.32	rs10910099	3.09E-06	rs12027041	rs10910099	0.0	1.02E-09*	rs867436	4.795e-07
	10p15.1	rs2104286	7.31E-06	rs1570527	rs2104286	-0.03	3.71E-09*	rs2181623	5.182e-05
T1D	10p15.1	rs2104286	8.28E-06	rs12722489	rs2104286	-0.74	1.17E-08*	rs12722563	6.39E-09*
T2D	9p21.3	rs523096	2.49E-04	rs10757283	rs10811661	-0.51	3.36E-09*	rs12555274	1.877e-07

Table 2: Loci that were not significant in the standard marginal approach but became significant using jester. ρ indicates correlation of SNPs. Results at these loci from imputation against 1000 genomes are also reported. P-values which are significant for a particular testing method are denoted by an asterisk. References for papers reporting significant associations at these loci can be found in Table S2.

FDR	1%	2%	3%	4%	5%	10%	15%	20%	25%
marginal	4528	4936	5240	5458	5641	6432	7072	7600	8123
joint	4884	5376	5726	6015	6248	7206	7947	8625	9277
new genes	597	698	781	850	908	1113	1263	1431	1565
LM genes	276	303	334	358	381	473	566	634	701

Table 3: Joint testing pairs of *cis*-SNPs improves the number of eQTL's detected by 10.7%. Many of the new genes discovered using the joint test appear to be linkage masked, with correlation between the significant SNP pair of above 0.2.

References

- [1] Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., et al. (2014). The nhgri gwas catalog, a curated resource of snp-trait associations. *Nucleic Acids Research* 42, D1001–D1006.
- [2] Eicher, J. D., Landowski, C., Stackhouse, B., Sloan, A., Chen, W., Jensen, N., Lien, J.-P., Leslie, R., and Johnson, A. D. (2015). Grasp v2.0: an update on the genome-wide repository of associations between snps and phenotypes. *Nucleic Acids Research* 43, D799–D804.

- [3] Yang, J., Ferreira, T., Morris, A. P., Medland, S. E., Madden, P. A. F., Heath, A. C., Martin, N. G., Montgomery, G. W., Weedon, M. N., Loos, R. J., et al. (2012). Conditional and joint multiple-snp analysis of gwas summary statistics identifies additional variants influencing complex traits. *Nat Genet* *44*, 369–375.
- [4] Wood, A. R., Hernandez, D. G., Nalls, M. A., Yaghootkar, H., Gibbs, J. R., Harries, L. W., Chong, S., Moore, M., Weedon, M. N., Guralnik, J. M., et al. (2011). Allelic heterogeneity and more detailed analyses of known loci explain additional phenotypic variation and reveal complex patterns of association. *Human Molecular Genetics* *20*, 4082–4092.
- [5] Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., and Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* *11*, 446–450.
- [6] Visscher, P. M., Brown, M. A., McCarthy, M. I., and Yang, J. (2012). Five years of gwas discovery. *American Journal of Human Genetics* *90*, 7–24.
- [7] Gusev, A., Bhatia, G., Zaitlen, N., Vilhjalmsson, B. J., Diogo, D., Stahl, E. A., Gregersen, P. K., Worthington, J., Klareskog, L., Raychaudhuri, S., et al. (2013). Quantifying missing heritability at known gwas loci. *PLoS Genet* *9*, e1003993 EP –.
- [8] Patsopoulos, N. A., Barcellos, L. F., Hintzen, R. Q., Schaefer, C., van Duijn, C. M., Noble, J. A., Raj, T., Gourraud, P.-A., Stranger, B. E., Oksenberg, J., et al. (2013). Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: Hla and non-hla effects. *PLoS Genet* *9*, e1003926 EP –.
- [9] Fellay, J., Thompson, A. J., Ge, D., Gumbs, C. E., Urban, T. J., Shianna, K. V., Little, L. D., Qiu, P., Bertelsen, A. H., Watson, M., et al. (2010). Itpa gene variants protect against anaemia in patients treated for chronic hepatitis c. *Nature* *464*, 405–408.
- [10] Liu, J. Z., Almarri, M. A., Gaffney, D. J., Mells, G. F., Jostins, L., Cordell, H. J., Ducker, S. J., Day, D. B., Heneghan, M. A., Neuberger, J. M., et al. (2012). Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. *Nat Genet* *44*, 1137–1141.
- [11] Trynka, G., Hunt, K. A., Bockett, N. A., Romanos, J., Mistry, V., Szperl, A., Bakker, S. F., Bardella, M. T., Bhaw-Rosun, L., Castillejo, G., et al. (2011). Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* *43*, 1193–1201.
- [12] Udler, M. S., Meyer, K. B., Pooley, K. A., Karlins, E., Struewing, J. P., Zhang, J., Doody, D. R., MacArthur, S., Tyrer, J., Pharoah, P. D., et al. (2009). Fgfr2 variants and breast cancer risk: fine-scale mapping using african american studies and analysis of chromatin conformation. *Human Molecular Genetics* *18*, 1692–1703.
- [13] Pickrell, J. K. (2014). Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *American Journal of Human Genetics* *94*, 559–573.
- [14] Corradin, O., Saiakhova, A., Akhtar-Zaidi, B., Myeroff, L., Willis, J., Cowper-Sallari, R., Lupien, M., Markowitz, S., and Scacheri, P. C. (2014). Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Research* *24*, 1–13.
- [15] Bulmer, M. G. (1971). The effect of selection on genetic variability. *The American Naturalist* *105*, 201–211.
- [16] Lappalainen, T., Montgomery, S. B., Nica, A. C., and Dermitzakis, E. T. (2011). Epistatic selection between coding and regulatory variation in human evolution and disease. *American Journal of Human Genetics* *89*, 459–463.

- [17] Prabhu, S. and Pe'er, I. (2012). Ultrafast genome-wide scan for snp–snp interactions in common complex disease. *Genome Research* *22*, 2230–2240.
- [18] Arkin, Y., Rahmani, E., Kleber, M. E., Laaksonen, R., März, W., and Halperin, E. (2014). Epiq—efficient detection of snp–snp epistatic interactions for quantitative traits. *Bioinformatics* *30*, i19–i25.
- [19] Slavin, T. P., Feng, T., Schnell, A., Zhu, X., and Elston, R. C. (2011). Two-marker association tests yield new disease associations for coronary artery disease and hypertension. *Human genetics* *130*, 725–733.
- [20] Howey, R. and Cordell, H. J. (2014). Imputation without doing imputation: A new method for the detection of non-genotyped causal variants. *Genetic Epidemiology* *38*, 173–190.
- [21] Lee, S. H., Nyholt, D. R., Macgregor, S., Henders, A. K., Zondervan, K. T., Montgomery, G. W., and Visscher, P. M. (2010). A simple and fast two-locus quality control test to detect false positives due to batch effects in genome-wide association studies. *Genetic Epidemiology* *34*, 854–862.
- [22] Han, B., Kang, H. M., and Eskin, E. (2009). Rapid and accurate multiple testing correction and power estimation for millions of correlated markers. *PLoS Genet* *5*, e1000456 EP –.
- [23] Consortium, W. T. C. C. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* *447*, 661–678.
- [24] Lappalainen, T., Sammeth, M., Friedlander, M. R., /t Hoen, P. A. C., Monlong, J., Rivas, M. A., Gonzalez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P. G., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* *501*, 506–511.
- [25] Seaman, S. R. and Muller-Myhsok, B. (2005). Rapid simulation of p values for product methods and multiple-testing adjustment in association studies. *American Journal of Human Genetics* *76*, 399–408.
- [26] Lin, D. Y. (2005). An efficient monte carlo approach to assessing statistical significance in genomic studies. *Bioinformatics* *21*, 781–787.
- [27] Stranger, B. E., Montgomery, S. B., Dimas, A. S., Parts, L., Stegle, O., Ingle, C. E., Sekowska, M., Smith, G. D., Evans, D., Gutierrez-Arcelus, M., et al. (2012). Patterns of cis regulatory variation in diverse human populations. *PLoS Genet* *8*, e1002639.
- [28] Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* *38*, 904–909.
- [29] Chang, C., Chow, C., Tellier, L., Vattikuti, S., Purcell, S., and Lee, J. (2015). Second-generation plink: rising to the challenge of larger and richer datasets. *GigaScience* *4*, 7.
- [30] Haig, D. (2011). Does heritability hide in epistasis between linked snps? *European Journal of Human Genetics* *19*, 123.
- [31] Hemani, G., Knott, S., and Haley, C. (2013). An evolutionary perspective on epistasis and the missing heritability. *PLoS Genet* *9*, e1003295 EP –.
- [32] Zaitlen, N., Min Kang, H., and Eskin, E. (2009). Linkage effects and analysis of finite sample errors in the hapmap. *Human Heredity* *68*, 73–86.
- [33] Zaitlen, N., Paşaniuc, B., Patterson, N., Pollack, S., Voight, B., Groop, L., Altshuler, D., Henderson, B. E., Kolonel, L. N., Marchand, L. L., et al. (2012). Analysis of case–control association studies with known risk variants. *Bioinformatics* *28*, 1729–1737.
- [34] Zaitlen, N., Lindström, S., Pasaniuc, B., Cornelis, M., Genovese, G., Pollack, S., Barton, A., Bickeböller, H., Bowden, D. W., Eyre, S., et al. (2012). Informed conditioning on clinical covariates increases power in case-control association studies. *PLoS Genet* *8*, e1003032 EP –.

Appendix

Algorithms

Algorithm 1: Method for sampling from the null distribution to determine significance threshold

Input: significance level α , number of samples n , window size W_z , a set of joint tests T .

Output: Significance threshold

Sample marginal test statistics using a conditional normal approximation

Compute the p-values associated to the marginal tests

for $J_t \in T$ **do**

 | Compute $S = \chi_{J_t}^2$ using (??)

 | Compute the p-value associated to S

end

Sort the p-values from all tests performed

return $(1 - \alpha) \times n$ th smallest p-value

Algorithm 2: *jester's* GWAS pipeline

Input: A matrix of genotypes G and a vector of phenotypes Y for N individuals

Output: A set of pairs of variants meeting genome-wide significance

Perform standard QC on G and Y

Estimate the joint test significance threshold α_{LJT}

for $SNP G_i \in G$ **do**

 | Test G_i for association with Y

 | Jointly test G_i with any of the preceding W_j markers correlated with G_i at level ρ_{min}^2 or greater

end

return SNP pairs with p-value $< \alpha_{LJT}$

Proof of Observation 1

Consider the linear models

$$N : Y \sim 1 \tag{1}$$

$$J : Y \sim 1 + G_i + G_j \tag{2}$$

The likelihood ratio test $2(\mathcal{L}_J - \mathcal{L}_N)$ is asymptotically χ_2^2 under the null hypothesis of no association. In fact, it is possible to compute the value of this test statistic directly, without fitting the model against a null phenotype.

Observation 1. Let Z_i, Z_j be the Z-values of test statistics for the SNPs (i, j) against a phenotype Y . Let the correlation between G_i and G_j be $\rho_{i,j}$. Then the likelihood ratio test statistic for the model J against the null N is

$$\chi_J^2 = \frac{1}{1 - \rho_{i,j}^2} (Z_i^2 + Z_j^2 - 2\rho_{i,j}Z_iZ_j) \tag{3}$$

and is asymptotically χ_2^2 distributed.

Proof. The equality follows from setting up the normal equations and solving them. Let X be the $N \times 3$ matrix of regressors $X = (\vec{1} | G_i | G_j)$. Assume that in the linear model $Y = X\beta + \epsilon$, the distribution of the error terms is $\epsilon \sim \mathcal{N}(0, \sigma^2)$. Then the normal equations for the β -coefficients are

$$\beta = (X^\top X)^{-1} X^\top Y$$

solving this and simplifying yields

$$\begin{pmatrix} \beta_1 \\ \beta_i \\ \beta_j \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{N^3}{D}(\rho_{Y,i} - \rho_{ij}\rho_{Y,j}) \\ \frac{N^3}{D}(\rho_{Y,j} - \rho_{ij}\rho_{Y,i}) \end{pmatrix}$$

where

$$D = N^3(1 - \rho_{ij}^2)$$

is the determinant of $X^\top X$.

The log-likelihood for a linear model is $\mathcal{L} = -\frac{1}{2\sigma^2} \sum_{i=1}^N (Y_i - \beta^\top X_i)$. Using the above calculation of β , we can find the model log-likelihoods and compute the likelihood ratio

$$\begin{aligned} \mathcal{L}_N &= -\frac{N}{2\sigma^2} \\ \mathcal{L}_J &= -\frac{N}{2\sigma^2}(1 - 2\rho_{Y,i}\beta_i - 2\rho_{Y,j}\beta_j + \beta_i^2 + \beta_j^2 + \rho_{i,j}\beta_i\beta_j) \\ \chi_J^2 &= 2(\mathcal{L}_J - \mathcal{L}_N) \end{aligned}$$

Simplifying the above yields,

$$\begin{aligned} \chi_J^2 &= \frac{N}{\sigma^2} \frac{1}{1 - \rho_{i,j}} (\rho_{Y,1}^2 + \rho_{Y,2}^2 - 2\rho_{i,j}\rho_{Y,1}\rho_{Y,2}) \\ &= \frac{1}{(1 - \rho_{i,j})} (Z_{Y,1}^2 + Z_{Y,2}^2 - 2\rho_{i,j}Z_{Y,1}Z_{Y,2}) \end{aligned}$$

□

Thus, given the marginal test statistics and the sample correlation of the genotype pair, we can compute the joint test statistic under the null without computationally fitting the model (similar results derived in other contexts can be found in[?] and[?]). This, when combined with MVN sampling of the marginal test statistics, allows a substantial speedup over a permutation test.

While the above is derived in the context of a continuous disease phenotype, it is straightforward to conclude that the framework is extensible to case-control (binary) phenotypes. While we assumed for simplicity the phenotype was standardized, the result is independent of the scale of Y and thus holds for the diseases on the underlying liability scale. Since the least squares model does not rely on the assumption that the error terms are normally distributed (only that they are spherical) the result extends to logistically distributed residuals (logistic regression). In this case the β 's are log odds-ratios.