A Joint Location-Scale Test Improves Power to Detect Associated SNPs, Gene Sets, and Pathways

David Soave,[1,2] Harriet Corvol,[3,4] Naim Panjwani,[2] Jiafen Gong,[2] Weili Li,[1,2] Pierre-Yves Boëlle,[4,5] Peter R. Durie,[6,7] Andrew D. Paterson,[2,8] Johanna M. Rommens,[2,9] Lisa J. Strug,[1,2,*] and Lei Sun[1,10,*]

Gene-based, pathway, and other multivariate association methods are motivated by the possibility of GxG and GxE interactions; however, accounting for such interactions is limited by the challenges associated with adequate modeling information. Here we propose an easy-to-implement joint location-scale (JLS) association testing framework for single-variant and multivariate analysis that accounts for interactions without explicitly modeling them. We apply the JLS method to a gene-set analysis of cystic fibrosis (CF) lung disease, which is influenced by multiple environmental and genetic factors. We identify and replicate an association between the constituents of the apical plasma membrane and CF lung disease (p = 0.0099 and p = 0.0180, respectively) and highlight a role for the SLC9A3-SLC9A3R1/2-EZR complex in contributing to CF lung disease. Many association studies could benefit from re-analysis with the JLS method that leverages complex genetic architecture for SNP, gene, and pathway identification. Analytical verification, simulation, and additional proof-of-principle applications support our approach.

Introduction

Identifying the genetic architecture of complex traits requires analytic strategies that move beyond single-variant association tests. Multivariate analyses such as gene-based, gene-by-gene interaction (GxG), gene-by-environment interaction (GxE), gene-set, and pathway analyses are now commonly implemented,[1–5] yet, one rarely sees GxG or GxE explicitly accounted for within gene-set and pathway analyses.[6] The specification of interacting variables that probably differ between genes in gene sets and pathways is not straightforward. Interacting exposure variables (termed *E* hereafter) could include contributing environmental factors or SNPs or haplotypes from the same region or at other susceptibility loci. Missing or incorrect information on interacting factors, as well as associated computational burden, might also limit more comprehensive surveys of the whole genome for disease association. Here we provide an easy-to-implement, straightforward solution to exploit potential GxG and GxE in gene-set and pathway analyses, and we illustrate the power of such an approach in a gene-modifier study of lung disease in cystic fibrosis (CF [MIM: 219700]).

CF is a life-limiting genetic disease, with the majority of mortality due to lung disease. CF is caused by mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR* [MIM: 602421]). Individuals with the same loss-of-function *CFTR* mutations have variable disease severity across the lungs, intestine, and other CF-affected organs, suggesting that variation in other (modifier) genes might play a role in disease pathophysiology.[7] A major source of pathophysiology in CF is impaired fluid and electrolyte flux in CF-affected organ epithelia. Sun et al.[5] hypothesized that regulators of fluid, solute, and ion transport that physically co-localize with CFTR would play a role in determining CF disease severity. Using a composite tissue gene set that corresponds to 157 gene products localized to the apical plasma membrane of epithelia, a multivariate sum-test provided evidence that many of the constituent gene products contribute to early intestinal obstruction (called meconium ileus [MIM: 614665]) in CF. Meconium ileus is heritable (~88% estimated heritability) and presents at birth in ~15% of CF-affected individuals,[8] with limited opportunity for environmental involvement.

Lung disease in CF is distinct, being progressive and notably affected by environmental exposures such as infection history, secondhand smoke, air pollution, and ambient air temperature,[9–11] with a ~50% estimated heritability.[12] Similar to the intestinal epithelia, the apical membrane constituents of lung epithelia should likewise contribute to disease. Testing this, or any hypothesis in CF lung disease, however, has the added difficulty of being able to specify and collect accurate information on the contributing exposure variables and model the potential GxE. Therefore, an alternative association testing methodology is needed to account for the effects of interacting variables that are either unknown or unmeasured.

[1]Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON M5T 3M7, Canada; [2]Program in Genetics and Genome Biology, Research Institute, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada; [3]Assistance Publique-Hôpitaux de Paris (AP-HP), Trousseau Hospital, Pediatric Pulmonology Department; Institut National de la Santé et la Recherche Médicale (INSERM), UMR_S 938, CDR Saint-Antoine, 75012 Paris, France; [4]Sorbonne Universités, Université Pierre et Marie Curie (UPMC) Paris 06, 75005 Paris, France; [5]AP-HP, Saint-Antoine Hospital, Biostatistics Department, INSERM, UMR_S 1136, 75012 Paris, France; [6]Program in Physiology and Experimental Medicine, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada; [7]Department of Pediatrics, University of Toronto, Toronto, ON M5S 1A1, Canada; [8]Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Toronto, ON M5T 3M7, Canada; [9]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada; [10]Department of Statistical Sciences, University of Toronto, Toronto, ON M5S 3G3, Canada
*Correspondence: lisa.strug@utoronto.ca (L.J.S.), sun@utstat.toronto.edu (L.S.)

For a biallelic SNP, a genetic interaction (either GxG or GxE) on a quantitative phenotype will lead to differences in phenotypic distribution across the three genotypes, leading to differences in phenotypic variance (scale).[13] In light of this, Levene's scale test of equality of variance[14] has been proposed as a method of prioritizing SNPs for subsequent GxG and GxE studies,[13,15] in contrast to the standard mean (location) test (i.e., testing for phenotypic differences in mean across genotypes). The advantage of using variance testing to incorporate GxG or GxE is that exposures need not be specified, and the enormous multiple testing burden of formally examining all possible pair-wise interactions is removed. Variance-only testing, however, has limited power to detect SNPs displaying main effects only.

Focusing on single-SNP analysis, Aschard et al.[16] proposed a test that compares the percentiles of phenotypic values between genotypes, capturing mean, variance, and other differences. Although this approach comprehensively evaluates the phenotypic distribution between genotypes, it sacrifices statistical power when a (approximately) normally distributed trait is sufficiently summarized by its mean and variance. Furthermore, the distribution test statistic requires computationally intensive permutation-based methods for p value estimation that are challenging to implement on a genome-wide scale.

Cao et al.[17] considered a joint test of mean and variance differences by using a full likelihood approach based on linear regression models. The likelihood ratio test (LRT) statistic follows an asymptotic chi-square distribution under the null hypothesis. The LRT approach can increase power but is more sensitive to model assumptions such as normality. Therefore, Cao et al.[17] proposed a parametric bootstrap method to calculate "honest" p values at the cost of computational efficiency. For both the LRT and distribution methods, implementation difficulties arise for multivariate (e.g., gene-based, gene-set, and pathway) analyses (see Appendix A).

Here we propose an easy-to-implement joint location-scale (JLS) testing framework that simultaneously tests the null hypothesis of equal mean *and* equal variance between genotypes, by aggregating association evidence from the individual location-only and scale-only tests, focusing on Fisher's method of combining information (JLS-Fisher). The proposed method detects association in the presence of underlying genetic main and/or interaction effects, *without* specifying the interactors; it allows any type of individual location and scale tests to be combined, making it particularly useful for gene-based, gene-set, and pathway analyses.

Using two proof-of-principle examples in the single-variant testing setting, where GxE interactions have been shown to exist from previous studies, we demonstrate the power of the JLS method. The first example pertains to a locus that determines glycosylated hemoglobin (HbA1c) levels based on a genome-wide association study (GWAS) of type 1 diabetes (T1D [MIM: 222100]) complica-

tions.[18] The second pertains to determinants of CF lung disease severity.[19] For these examples, the known environmental interactors are not explicitly modeled, but we leverage their existence to identify genetic contributors to HbA1c and CF lung disease.

We then use the multivariate JLS method to identify and replicate an association between the apical gene set, and the SLC9A3 complex within the set, and lung disease severity in CF; this finding would have been missed by the traditional multivariate gene-set tests that do not account for (heterogeneous) GxG and GxE interactions. Through extensive simulation analyses, we demonstrate that the proposed JLS method has good type 1 error control with improved power compared to other testing options, and, in contrast to the recently proposed distribution[16] and LRT joint testing methods,[17] can be easily implemented in the context of gene-set and pathway analyses.

Material and Methods

Genetic Model and Notation

Let Y be a quantitative trait and G be the minor allele count for the SNP under investigation ($G = 0$, 1, or 2); the additive assumption is not critical to the method development. Let E be an exposure variable interacting with the genetic factor. This exposure E could reflect continuous or categorical measures of environmental or genetic background. The underlying true genetic model can include main effects of both G (β_G) and E (β_E) on Y, as well as the interaction effect (β_{GE}):

$$Y \sim \beta_0 + \beta_G G + \beta_E E + \beta_{GE} GE + \varepsilon. \qquad \text{(Equation 1)}$$

We assume that the trait Y is normally distributed with unit variance conditional upon G and E, in other words, $\text{Var}(Y \mid G = g, E = e) = 1$ and $\varepsilon \sim \mathcal{N}(0, 1)$.

When considering only G, the working model would reduce to

$$Y \sim \beta_0 + \beta_G G + \varepsilon_G. \qquad \text{(Equation 2)}$$

Pare et al.[13] showed that the conditional variance of Y conditional on G alone could be expressed as $\sigma_G^2 = \text{Var}(Y \mid G = g) = (\beta_E + \beta_{GE} g)^2 + 1$. Thus, if an interaction effect was present (i.e., $\beta_{GE} \neq 0$), the trait variance would differ between genotypes.

Joint Location-Scale Testing Procedure for Single-SNP Analysis

Our proposed JLS testing framework, based on the working model of Equation 2, tests the following null hypothesis:

$$H_0^{joint} : \beta_G = 0 \text{ and } \sigma_i = \sigma_j \text{ for all } i \neq j, \text{ i, j} = 0, 1, 2.$$

The alternative hypothesis of interest is

$$H_1^{joint} : \beta_G \neq 0 \text{ or } \sigma_i \neq \sigma_j \text{ for some } i \neq j.$$

For a SNP under study, different JLS test statistics can be considered. Let p_L be the p value for the location test of choice (i.e., testing $H_0^{location} : \beta_G = 0$ using, for example, ordinary least-squares regression), and p_S be the p value for the scale test of choice (i.e., testing $H_0^{scale} : \sigma_i = \sigma_j$ for all $i \neq j$ using, for example, Levene's test). We first consider Fisher's method (JLS-Fisher) to combine

the association evidence from the individual location and scale tests. The JLS-Fisher statistic is defined as

$$W_F = -2(\log(p_L) + \log(p_S)).$$

Large values of W_F correspond to small values of p_L and/or p_S and provide evidence against the null H_0^{joint}. If p_L and p_S are independent under H_0^{joint}, W_F is distributed as a χ_4^2 random variable. Although Fisher's method here is used to combine evidence from two tests applied to the same sample, the assumption of independence between p_L and p_S under H_0^{joint} holds theoretically for a normally distributed trait (Appendix A, Lemma 1), as well as empirically for approximately normally distributed traits in finite samples (Figures S1 and S2, Tables S1 and S2).

One can also consider the minimum p value (JLS-minP) approach, or various alternatives based on combining the individual test statistics themselves with or without weights.[20–23] The JLS-minP statistic is defined as

$$W_M = \min(p_L, p_S).$$

If p_L and p_S are independent under H_0^{joint}, W_M is distributed as a Beta random variable (with shape parameters 1 and 2) where small values of W_M correspond to small values of p_L and/or p_S and evidence against the null.

Joint Location-Scale Testing Procedure for Gene-Set Analysis

The chosen JLS test statistic (e.g., W_F) for single-SNP analysis can then be used for implementing gene-based, gene-set, or pathway analysis in a direct fashion.

Assume that J SNPs have been annotated to a gene or gene-set of interest. For each SNP j, the JLS-Fisher test statistic (e.g., W_{Fj}) is first obtained and then the association evidence can be aggregated across the SNPs by considering, for example, the sum statistic,[5] $\sum_j W_{Fj}$. To account for LD between SNPs, the overall association evidence can be evaluated by a phenotype-permutation approach where the empirical p value is the proportion of K permutation replicates with sum statistics more extreme than the observed value. Because this multivariate method analyzes all J SNPs simultaneously, the number of permutations need not be exceedingly large and $K = 10,000$ provides accurate estimates for p values in the range of 0.05. If multiple genes or gene-sets are of interest, more replicates would be required to adjust for the corresponding number of hypothesis tests.

To compare strength of association evidence between sets of variants within the same gene or across different genes, an extension of the gene-set approach can be implemented. Sum statistics are obtained as previously described for each group of variants, then calibrated by the respective number of variants. The difference between the two proportional sum statistics is the test statistic of interest,

$$D_F = \frac{1}{J}\sum_j W_{Fj} - \frac{1}{K}\sum_k W_{F,k},$$

where the $j = 1,\ldots,J$ and $k = 1,\ldots,K$ subscripts index the competing sets of variants, and the F subscript indicates that the variant-specific joint location-scale statistics are obtained by Fisher's method (although other methods of combining such as minP could be used as well). The significance of D_F can be evaluated with the phenotype-permutation approach as described above.

In all applications, genotypes were coded additively ($G = 0, 1$ or 2), and for X chromosome SNPs, female and male genotypes were analyzed together and coded as $G = 0, 1$, or 2 and $G = 0$ or 2, respectively.

Application Data: HbA1c Levels in Type 1 Diabetes and Cystic Fibrosis Lung Disease

We tested the proposed JLS approach with applications to genetic association studies of HbA1c levels in type 1 diabetes and lung disease in cystic fibrosis.

The T1D application used the Diabetes Control and Complications Trial (DCCT) sample in which 667 individuals were conventionally treated and 637 intensively treated.[18] In this sample, an earlier GWAS of HbA1c levels in type 1 diabetes[18] showed that rs1358030 near *SORCS1* (10q25.1 [MIM: 606283]) interacts with treatment type (conventional versus intensive) on HbA1c levels (quarterly measured values spanning 6.5 years). To demonstrate that the JLS testing framework could leverage the interaction effect without knowledge of the interacting variable, we analyzed the association of rs1358030 with the average, inverse normal transformed HbA1c value, assuming the treatment variable was not available.

The CF application involves association studies of an averaged lung function measure, forced expiratory volume in 1 s, adjusted for sex, age, height, and mortality, and normalized (SaKnorm),[24] using data from the Canadian Cystic Fibrosis Gene Modifier Study (CGS).[25] To reduce the duration of heterogeneous environmental exposures that were not measured, and recognizing that age can serve as a surrogate for these exposures, we[19] previously restricted our (location-only) analysis to lung function measures from pediatric ages (<18 years, n = 815 subjects from 753 unique families), analyzing eight SNPs in three genes, *SLC9A3* (MIM: 182307), *SLC6A14* (MIM: 300444), and *SLC26A9* (MIM: 608481), previously identified as associated with meconium ileus in a hypothesis-driven GWAS (GWAS-HD).[5] This approach involved a 46% reduction in sample size, and that the age restriction be fixed for all variants, despite the possibility that the optimal exposure (here, age) is gene specific. Here we re-analyzed the SNPs with the individual location-only and scale-only tests, as well as the joint JLS-Fisher, JLS-minP, LRT,[17] and distribution[16] tests, removing the age restriction and using the full CGS sample (n = 1,409 unrelated subjects). For comparison, the location-only analyses restricted to the pediatric population using different age cut-off points were also investigated.

With the full CGS sample, we further tested the hypothesis that multiple proteins present on the apical plasma membrane contribute to lung disease severity as measured by SakNorm; the hypothesis was considered previously for meconium ileus susceptibility in CF.[5] In total, 3,814 GWAS SNPs (MAF > 0.02) were annotated to within ± 10 kb of 155 apical genes obtained from the Gene Ontology project.[5] The JLS-Fisher test was first applied to each SNP, and the SNP-specific test statistic was then aggregated across all SNPs to perform the multivariate apical gene-set analysis. We then used an independent French sample (n = 1,232) for replication. Imputation based on 1000 Genomes[26,27] (as outlined in the Online Methods of Sun et al.[5]) was used to assess regional association within the SLC9A3R1, SLC9A3R2, and EZR binding sites of SLC9A3.

Institutional review committees at all participating institutions for the DCCT T1D study as well as all Canadian CF clinics approved this study. Consent was also obtained for participants from France with procedural approval (CPP 2004/15) and information collection approval by CNIL (04.404). Data collection, genotyping, and quality-control procedures are reported elsewhere for the T1D[18] and CF (Canadian and French) studies.[5,25]

JLS Method Evaluation by Simulation

We conducted extensive simulation analyses to evaluate the performance of the proposed JLS-Fisher and JLS-minP tests for single-variant analysis and compared them with the individual location-only and scale-only tests, as well as the distribution test[16] and the LRT.[17] We also conducted a sensitivity analysis on the various tests, studying the impact of poorly captured genotypes that can be expected from imputed datasets. Simulation details are provided in Appendix A.

Results

Type 1 Diabetes Complications

Traditional analysis using the individual location-only test did not yield a genome-wide significant association of $p < 5 \times 10^{-8}$ (as previously proposed by Dudbridge and Gusnanto[28]) between rs1358030 and HbA1c (p = 2.3×10^{-7}), whereas the JLS-Fisher method did (p = 4.9×10^{-8}). Genome-wide significance was also not achieved by the scale-only test (p = 0.01), the distribution test (p = 1.7×10^{-7}, estimated from 10^8 permutation replicates), the LRT (p = 2.0×10^{-7}), or the JLS-minP test (p = 4.6×10^{-7}).

Knowing the treatment information, Paterson et al.[18] was able to explicitly model this interactor and identified rs1358030 as genome-wide significantly associated with HbA1c levels and interacting with treatment type (p = 3.8×10^{-10} and p = 0.013 for the main and interaction effect, respectively).

Cystic Fibrosis Lung Disease: Single-Variant Association Analysis

We first noted in Table 1 that all methods (the joint and individual tests using the full sample or various sub-samples) consistently highlighted that variants in both SLC9A3 and SLC6A14 are associated with the lung phenotype, but not those in SLC26A9. For the individual location-only test, although the age cut-off of <18 years appears to be ideal for rs17563161 in SLC9A3 resulting in the smallest p value (p = 3.3×10^{-5}) even with the smaller sample size (n = 753), a different cut-off (<20 years, n = 830) yields the most significant location-only test result for rs3788766 in SLC6A14 (p = 0.0002). This illustrates the challenge of specifying and modeling interacting exposures in the context of multiple SNPs and genes of interest.

With the entire sample of CF-affected individuals (n = 1,409) encompassing a wider age range and thus greater variability of environmental exposures, we indeed observed evidence for scale differences in the lung phenotype (Levene's test p = 0.01–0.08) for variants in SLC9A3 and SLC6A14, revealing the possibility of GxG or GxE interaction. (Table S3 provides evidence for SNP-by-age interaction effects from regression models directly incorporating age.) Compared with the distribution and LRT joint tests, the JLS-Fisher test consistently provides the smallest p values for variants in SLC9A3 and SLC6A14.

Cystic Fibrosis Lung Disease: Gene-Set Association Analysis

We observed evidence of an association between the apical gene set and SakNorm (JLS-Fisher permutation p = 0.0099; Figures 1A and 1B and Table 2). Comparison with the individual location-only and scale-only tests showed that this association does not reach statistical significance using the conventional gene-set location test (regression permutation p = 0.0876; Figures 1C and 1D) whereas the variance component contributed a significant result (Levene's permutation p = 0.0222; Figures 1E and 1F).

Examination of the SNP- and gene-specific JLS-Fisher p values and rankings showed SLC9A3 to be the top ranked (Table S4), with three of the top four ranked SNPs being annotated to SLC9A3. This provided consistent support for the relationship between SLC9A3 and lung function as previously reported (Table 1).[19,29] In the independent French sample (n = 1,232), we observed replication of the apical hypothesis via the JLS-Fisher test (permutation p = 0.0180; Figures 2A and 2B and Table 2). Again, the standard location testing approach, by itself, was not powerful enough to detect the association (regression permutation p = 0.2058; Figures 2C and 2D), and the added contribution from the scale-testing component was beneficial (Levene's permutation p = 0.0077; Figures 2E and 2F). After excluding all ten genotyped SNPs annotated to SLC9A3, the apical gene-set test in both the CGS and French samples remained significant (Table 2), suggesting that the JLS-Fisher method identified multiple additional associations within the gene set, beyond the known SLC9A3 contribution.

Three of the five top-ranked genes in the apical gene set were SLC9A3, SLC9A3R2 (MIM: 606553), and EZR (MIM: 123900), with ten genes in total displaying JLS-Fisher p < 0.05 (Table S4). Of these top five, protein product interactions are known between (1) SLC9A3 (also known as NHE3) and SLC9A3R2 ("SLC9A3 regulator 2," also known as E3KARP or NHERF2), (2) SLC9A3 and EZR, and (3) SLC9A3R2 and EZR. SLC9A3R1, or "SLC9A3 regulator 1" (also known as EBP50 or NHERF1), is recognized as a paralog of SLC9A3R2 with comparable binding sites to both SLC9A3 and EZR (reviewed in Donowitz and Li[30]); SLC9A3R1 (MIM: 604990) was the 21st ranked. Based on interaction investigations typically involving intestinal and kidney tissues, a current paradigm is that EZR provides anchorages for the SLC9A3 regulators to the actin cytoskeleton, and probably also facilitates early trafficking of SLC9A3 from the Golgi to the cell periphery, with eventual "hand-off" to the SLC9A3 regulators. The SLC9A3 regulators help to maintain SLC9A3 at the apical membrane and facilitate its dynamic shuffling with endosomes and internal vesicles.[30]

Notably, variants within the SLC9A3 complex component genes are significantly associated as a set in the CGS sample using the JLS-Fisher or regression tests (p < 0.0001 and p < 0.0001, respectively; Table 2). The French sample provided replicated support for the SLC9A3

Table 1. Candidate-Gene Study of Cystic Fibrosis Lung Disease Severity as Measured by SaKnorm in the Canadian Gene Modifier Study Sample

					Full CGS Sample (n = 1,409)							CGS Pediatric Subsample		
												Location/Regression		
							Joint Tests				<16 years	<18 years[d]	<20 years
Chr	Gene	SNP	BP[a]	MAF[b]	Location/Regression	Scale/Levene	JLS-Fisher	JLS-minP	LRT	Distribution[c]	(n = 653)	(n = 753)	(n = 830)
1	SLC26A9	rs7512462	204,166,218	0.41	0.30	0.58	0.48	0.52	0.64	0.84	0.41	0.58	0.79
1	SLC26A9	rs4077468	204,181,380	0.42	0.53	0.61	0.69	0.78	0.65	0.75	0.14	0.20	0.32
1	SLC26A9	rs12047830	204,183,322	0.49	0.55	0.15	0.29	0.28	0.38	0.43	0.09	0.12	0.21
1	SLC26A9	rs7419153	204,183,932	0.37	0.50	0.06	0.14	0.12	0.18	0.42	0.25	0.33	0.51
5[e]	SLC9A3	rs17563161	550,624	0.26	0.0004	0.02	0.0001	0.0008	0.0008	0.002	0.0001	3.3×10^{-5}	8.0×10^{-5}
X	SLC6A14	rs12839137	115,479,578	0.24	0.02	0.08	0.01	0.05	0.03	0.04	0.02	0.04	0.02
X[e]	SLC6A14	rs5905283	115,479,909	0.49	0.009	0.07	0.005	0.02	0.02	0.03	0.02	0.01	0.01
X[e]	SLC6A14	rs3788766	115,480,867	0.40	0.001	0.01	0.0002	0.002	0.0005	0.002	0.0004	0.0005	0.0002

SaKnorm[24] (normalized, averaged sex-, age-, height-, mortality-adjusted forced expiratory volume in 1 s; FEV1) was used as the lung function measure. Analysis was performed on eight SNPs in three candidate genes previously identified as associated with meconium ileus[5] and subsequently studied for association with SaKnorm in a pediatric subsample.[19] Individual location-only and scale-only tests, joint location-scale Fisher (JLS-Fisher), joint location-scale minimum-P (JLS-minP), the likelihood ratio test (LRT) of Cao et al.,[17] and the distribution test of Aschard et al.[16] were performed in the full sample of 1,409 unrelated individuals. For comparison, the location-only test was performed in three CGS pediatric subsamples using different age cutoffs of 16, 18, and 20 years.
[a]hg18 assembly (March 2006; NCBI36).
[b]MAF is similar across the pediatric subsets (see Table S8).
[c]Distribution test p values were estimated from 10^5 permutation replicates.
[d]Pediatric subset cutoff (<18 years) used in Li et al.;[19] subset here includes only unrelated subjects (n = 753).
[e]Significant variants (p < 0.05 corrected for multiple testing) reported in Li et al.[19]

JLS-Fisher Test

A

B

Regression Location-only Test

C

D

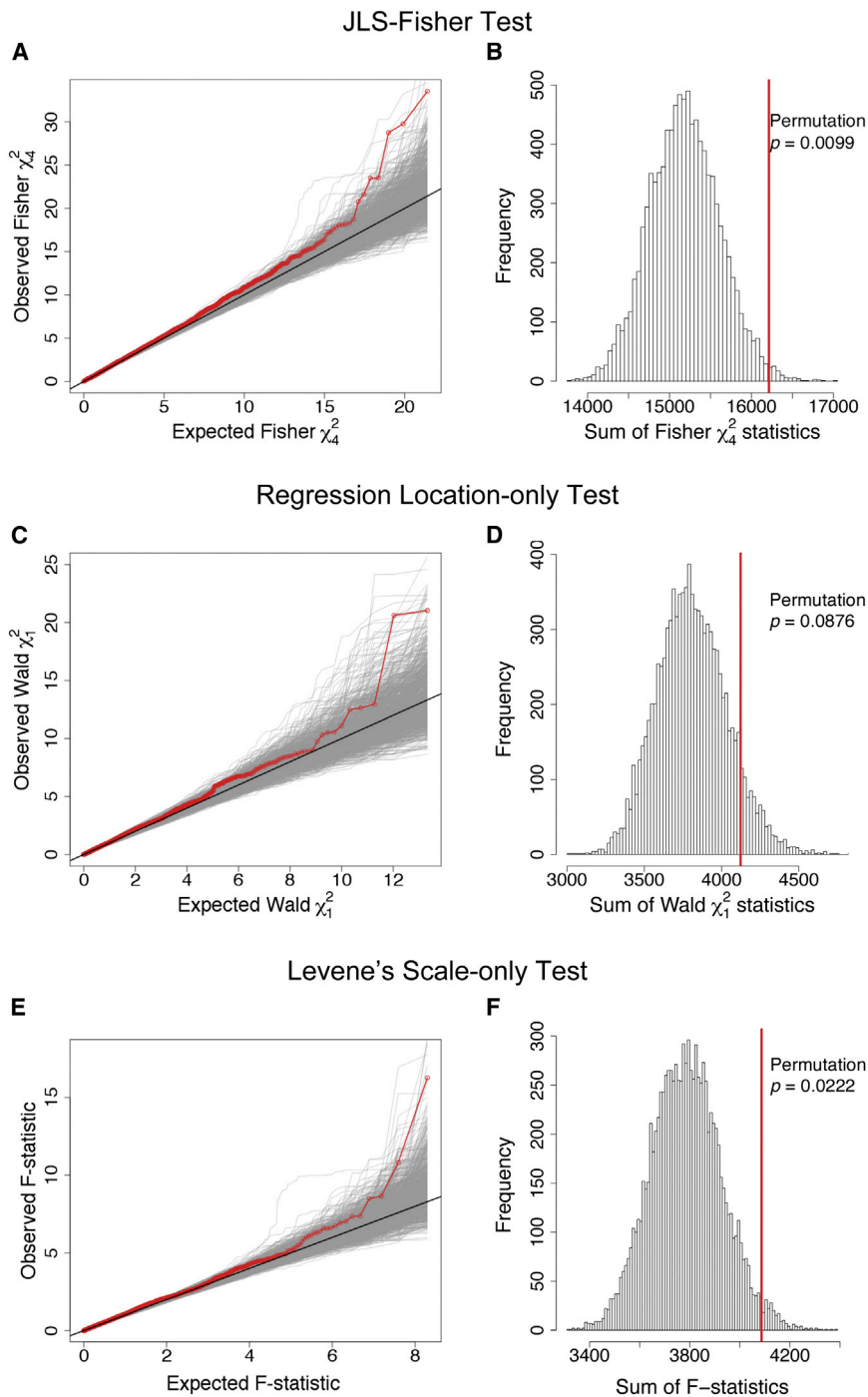## Levene's Scale-only Test

E

F

Figure 1. Apical Gene-Set Association Analysis of CF Lung Disease Severity as Measured by SaKnorm in the Canadian Gene Modifier Study Discovery Sample of 1,409 Subjects

SaKnorm[24] (sex-, age-, and mortality-adjusted, normalized forced expiratory volume in 1 s; FEV1) was used as the lung function measure. In total, 3,814 GWAS SNPs (MAF > 0.02) were annotated to within ± 10 kb of the 155 apical genes obtained from the Gene Ontology database.[5]

(A, C, E) Quantile-quantile plots (QQ-plots) of the SNP-specific association test statistics of the apical SNPs via (A) the proposed joint location-scale test (JLS-Fisher), (C) the regression based location-only test, and (E) Levene's scale-only test. QQ-plots of the observed association statistics are shown in red, and the QQ-plots of the statistics calculated from the 10,000 phenotype-permuted replicates are shown in gray.

(B, D, F) Corresponding gene-set association analysis results for (B) the JLS-Fisher test (permutation p = 0.0099), (D) the standard regression-based location-only test (permutation p = 0.0876), and (F) Levene's scale-only test (permutation p = 0.0222). For each figure, the observed sum statistic aggregated across all 3,814 SNPs, as described in the Material and Methods, is shown as a vertical line in red, and the sum statistics calculated from the 10,000 phenotype-permuted replicates is shown as a histogram.

complex using the JLS-Fisher test (p = 0.0415), although in this latter sample, *SLC9A3R1* is more highly ranked than *SLC9A3R2* and *EZR*. Removal of the four-gene set, *SLC9A3*, *SLC9A3R1/2*, and *EZR*, and the re-testing of the apical gene set with the remaining 151 genes suggests that association(s) beyond the SLC9A3 regulator complexes also exist (JLS-Fisher p = 0.0329 and 0.0201 in the CGS and French samples, respectively).

SLC9A3 contains multi-membrane spanning motifs in its amino portion to facilitate transport function, with a large cytosolic carboxyl portion that affords regulation with binding sites for multiple inter-actors. Although there is significant LD that extends throughout the gene (Figure 3), there was evidence suggesting *greater* association for variants as a group from the region corresponding to the regulatory portion, compared to the transporting portion (regression, Levene's, and JLS-Fisher permutation p = 0.0391, 0.312, and 0.1254, respectively). Amino acids 586–660 of SLC9A3 bind the SLC9A3 regulators where an exonic nucleotide variant (rs2230437) is associated with lung function (JLS-Fisher p = 7.6 × 10^{-6}). This synonymous change is in high LD ($r^2 > 0.8$) with four other variants, all with similar association evidence (rs11743825, rs41282625, 5:475625:GC_G, and rs11745923 with JLS-Fisher p = 1.4 × 10^{-6}, 2.2 × 10^{-6}, 2.3 × 10^{-6}, and 2.7 × 10^{-6}, respectively), and all in noncoding positions. There are also no associated coding variants in the major EZR binding site from amino acids 519–595. Similarly, there are no coding variants in the respective PDZ domains of the SLC9A3

Table 2. Apical Gene-Set and SLC9A3 Complex Association Analysis with CF Lung Disease Severity as Measured by SaKnorm in the Canadian Gene Modifier Study Discovery Sample and in the French Replication Sample

	Location-only Test	Scale-only Test	JLS-Fisher Test
Apical Gene-Set (3,814 SNPs)			
Canadian Discovery Sample	0.0876	0.0222	0.0099
French replication sample	0.2058	0.0077	0.0180
SLC9A3 Gene-Based (10 SNPs)			
Canadian Discovery Sample	<0.0001	0.0238	0.0001
French replication sample	0.0953	0.0415	0.0286
Apical Gene-Set Excluding SLC9A3 (3,804 SNPs)			
Canadian Discovery Sample	0.1319	0.0248	0.0161
French replication sample	0.2139	0.0076	0.0182
SLC9A3 Complex: SLC9A3, SLC9A3R1, SLC9A3R2, and EZR (38 SNPs)			
Canadian Discovery Sample	<0.0001	0.0016	<0.0001
French replication sample	0.2390	0.0245	0.0415
Apical Gene-Set Excluding the SLC9A3 Complex (3,776 SNPs)			
Canadian Discovery Sample	0.178	0.0329	0.0329
French replication sample	0.2173	0.0099	0.0201

Individual location-only and scale-only tests and the joint location-scale Fisher (JLS-Fisher) test were performed in the full CGS (n = 1,409) and French (n = 1,232) samples using unrelated individuals (see Material and Methods and Results). p values reported were estimated from 10,000 permutation replicates.

regulators or in the FERM domain of EZR that bind SLC9A3. The imputation analysis would have captured the major variation in this gene. Collectively, because the constituents or their direct physical interactions would not appear to be affected, disturbed expression with altered stoichiometry of components or dynamic positioning of the SLC9A3 complex might be contributing to lung disease severity.

Simulation

Results showed that, under the normal model and in most scenarios considered, the LRT and JLS-Fisher methods had similar power and were more powerful than the JLS-minP and distribution tests (Figures 4 and S3) and the individual location-only and scale-only tests (Figures S4 and S5).

Investigation of type 1 error (100,000 replicates) under the normal model showed that all individual and joint tests considered here maintained the nominal error rates at the 0.05, 0.005, and 0.0005 levels when MAF was at least 0.1 and there were ≥ 20 individuals within each genotype group (Tables S1 and S2). However, as MAF or the smallest genotype group size decreased, the LRT method demonstrated inflated type 1 error (Tables S1 and S2). Departure from normality also resulted in LRT having inflated type 1 error, as previously discussed in Cao et al.[17]

To study the type 1 error at the genome-wide level, we conducted association analysis of all 866,995 GWAS SNPs available in the T1D HbA1c example[18] with permuted phenotype. Results showed that all joint testing methods provide correct type 1 error control in a finite sample (Figures S1A, S1D, and S1G, respectively, for the JLS-Fisher, JLS-minP, and LRT methods; the distribution test by design has the correct type 1 error control). However, when stratified by minor allele frequency (MAF), we observed that, whereas the JLS-Fisher and JLS-minP tests are slightly conservative for SNPs with MAF < 0.1 (Figures S1C and S1F), the LRT method has increased type 1 error rate (Figure S1I) and would probably not be appropriate when the size of the genotype group is limited (n < 20). Permutation analysis of all of the 565,884 GWAS SNPs available in the CF application example[5] provided type 1 error results similar to those of the T1D HbA1c example (Figure S2).

For the sensitivity analysis, we observed that the individual location and scale tests and the JLS-Fisher and JLS-minP tests were not affected by poorly captured genotypes; they maintained correct type 1 error level irrespective of the proportion of incorrectly assigned genotypes (Table S5).

Discussion

Interactions between genes in gene-set and pathway analyses are assumed to exist, but they are, in general, not accounted for in current genetic analysis methods. Here we proposed a joint location-scale (mean-variance) testing method to account for these relationships as well as main effects, while avoiding explicit specification of the contributing interactions. We used this method to identify the apical membrane constituent gene set as contributing to CF lung disease and replicated this association in an independent sample of similarly ascertained CF-affected subjects. Further, we illustrated how the gene-set association notably benefits from consideration of phenotypic variance differences that might have been induced by GxG and/or GxE relationships. This apical gene-set association was not initially detected via the classical location-only testing as used in a typical GWAS and provides insight into interacting genes and exposure variables affecting lung function.

SNPs in SLC9A3 have been reported to associate with CF lung disease in two candidate gene studies.[19,29] However, a broader role for the SLC9A3/SLC9A3R2/EZR complex was not recognized. The JLS-Fisher test ranked these players as three of the top five contributing apical plasma membrane genes. Variation at SLC9A3R2 or EZR loci and possibly also SLC9A3R1 given its paralogous function (and supporting evidence in the French sample) could be expected to influence CF lung disease severity, presumably by affecting changes in net SLC9A3 activity. For example, modest changes in levels of either SLC9A3R2 or EZR could alter functional apical membrane SLC9A3 levels, just as
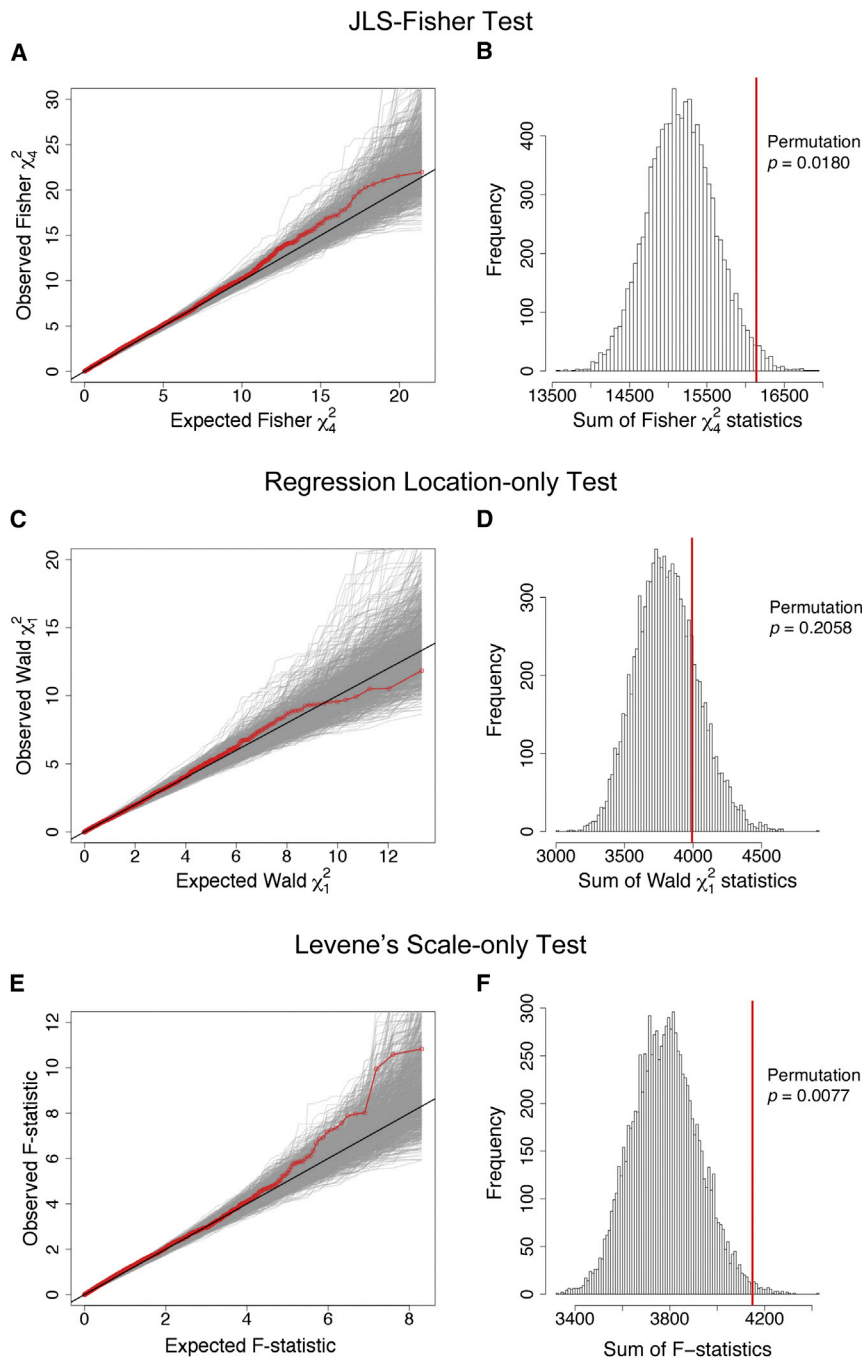
Figure 2. Apical Gene-Set Association Analysis of CF Lung Disease Severity as Measured by SaKnorm in the French Replication Sample of 1,232 Subjects

Analysis of the same SNPs as for the Canadian discovery sample (legend of Figure 1). The replication p values of the gene-set association analyses are permutation p = 0.0180 for the proposed joint location-scale test (JLS-Fisher) (A and B), permutation p = 0.2058 for the standard regression-based location-only test (C and D), and permutation p = 0.0077 for Levene's scale-only test (E and F). For other details see legend of Figure 1 and Material and Methods.

scale-only approach in practice. As proof of principle, we used examples from an association study of a complex trait (glycemic control in type 1 diabetes) and a modifier study of a Mendelian disease (lung function in CF). Both of these datasets provided examples of a genetic effect on a quantitative phenotype that varied based on another variable (i.e., treatment regimen and age, respectively). Our results showed that if the interacting exposure(s) were not known or available, the JLS method could have improved power over conventional approaches in detecting the association.

Recent literature cautions the investigation and interpretation of significant differences in phenotypic variability across genotypes as well as observed GxE and GxG interaction effects.[17,34–37] First, variance heterogeneity across genotypes can result from a mean-variance relationship induced by an inappropriate choice of measurement scale for the phenotype under investigation. Second, dependence between a causal locus and an exposure (G or E) can create synthetic interactions between the exposure and other tagged SNPs in incomplete LD with the causal locus.[36,37] Consequently, we might observe phenotypic variance heterogeneity across genotypes of a tagged SNP.[17] In these scenarios, however, because there is a true underlying genotype-phenotype relationship (main effect of G), it does not matter whether we detect this association through a location or a scale test, or both; the goal of our approach is to leverage complex genetic architecture to efficiently identify the associated G for follow-up investigation of causal interactors.

SLC9A3 locus variants might alter the amount of *SLC9A3* transcripts. It should also be noted that the SLC9A3 regulators, with their proximal localization at the apical membrane, could contribute a tethering function for a number of other apical membrane proteins, including CFTR.[31] Given loss of function in CF, SLC9A3 regulators might not be mediating their effects via CFTR; however, it is possible that some of the remaining significant apical membrane components, beyond SLC9A3, utilize the EZR and/or the SLC9A3 regulator scaffold.[32,33]

We demonstrated the potential advantage of using the JLS framework over the conventional location-only or

Challenges in understanding and being able to replicate original findings, particularly in the GxG, GxE, gene-set,
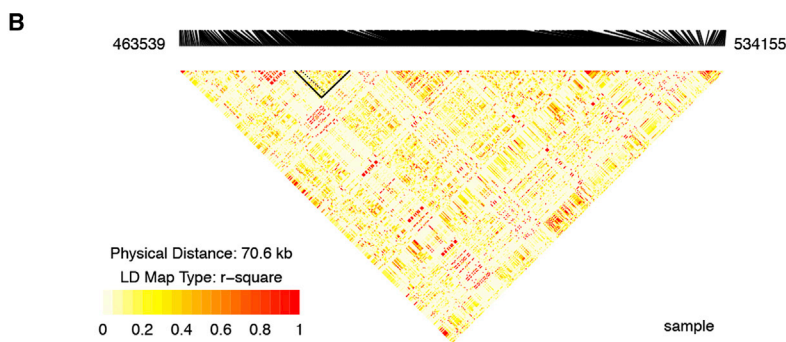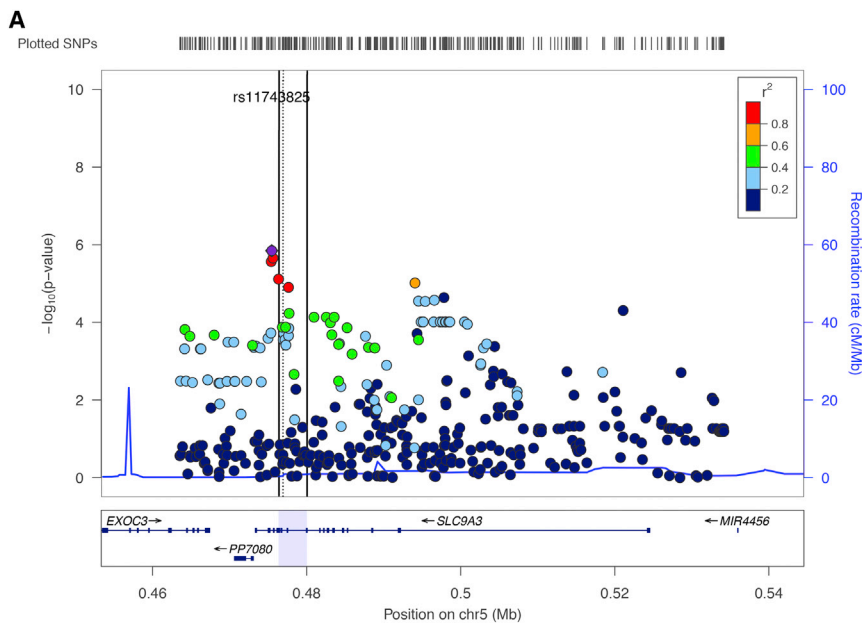
Figure 3. Association and Linkage Disequilibrium Information across *SLC9A3* in the Canadian Gene Modifier Study Sample

(A) Locus Zoom plot[50] displaying the genetic association of 374 imputed SNPs (see Online Methods of Sun et al.[5]) within ±10 kb of *SLC9A3* on chromosome 5. SNP colors indicate the pairwise correlation (r^2) with the top SNP, rs11743825 at bp 475463. The three vertical lines encompass the two adjacent binding regions of SLC9A3R1/SLC9A3R2 (left side of the dotted line: bp 476,353–476,778) and EZR (right side of the dotted line: bp 476,734–480,024). Both binding regions are bounded within the C terminus of the gene (bp 474,977–482,675).

(B) LD plot indicating pairwise correlation between imputed SNPs for the CGS sample using the "snp.plotter" package.[51] LD block colors indicate degree of pairwise LD (r^2). The triangles outlined with black lines indicate the SLC9A3R1/R2 and EZR binding regions as described in (A).

tified the SLC9A3 regulatory complex as an important contributor to CF lung disease, which was completely missed in previous analyses of the same data. The JLS association testing method should be considered for future gene-set, pathway, and whole-genome association scans. Further, JLS should be employed to re-examine datasets previously analyzed via the conventional location-only or scale-only testing approaches, for complex traits or complex secondary phenotypes of Mendelian diseases. The method will help researchers pinpoint susceptibility loci for additional analysis toward the understanding of the genetic architecture of a complex trait.

and pathway analysis setting, is well documented.[38] Given the potential heterogeneity between the Canadian discovery sample and the French replication sample (e.g., clinical treatment, climate, and air quality), interaction models could differ between the two samples. Therefore, even if the exposure variables were precisely measured, a significant multivariate interaction effect found in the discovery sample may not be observed in the replication sample. The proposed method allows different exposure variables between genes and between samples, as long as the corresponding underlying phenotype-genotype association mechanism results in phenotypic mean and/or variance differences between genotypes of the variants of interest. A step-by-step guide for application of the JLS framework, and a discussion of its advantages compared to alternative approaches, is included as additional considerations in Appendix A.

In conclusion, we have provided a robust joint location-scale (JLS) testing framework for the detection of single-variant, gene-set, or pathway associations involving either main or interaction effects, or both. This methodology applies to analysis of a quantitative trait, regardless of the biological interpretation of the chosen scale of phenotypic measurement. Application of the JLS approach iden-

Appendix A

Implementation: Steps for Applying the JLS Testing Framework

Note that the following implementation is appropriate for testing association of a phenotype with a genotyped genetic variant (e.g., SNP) or an imputed variant with a "hard call" (i.e., assign individual to the genotype with the highest posterior imputation probability; high genotype uncertainty does not affect type 1 error but will decrease power), using a sample of unrelated subjects.

1. Check the phenotype of interest for fit to a normal distribution. If required, adjust the phenotype using a suitable transformation, e.g., inverse normal transform. If the researcher proceeds using a non-normal
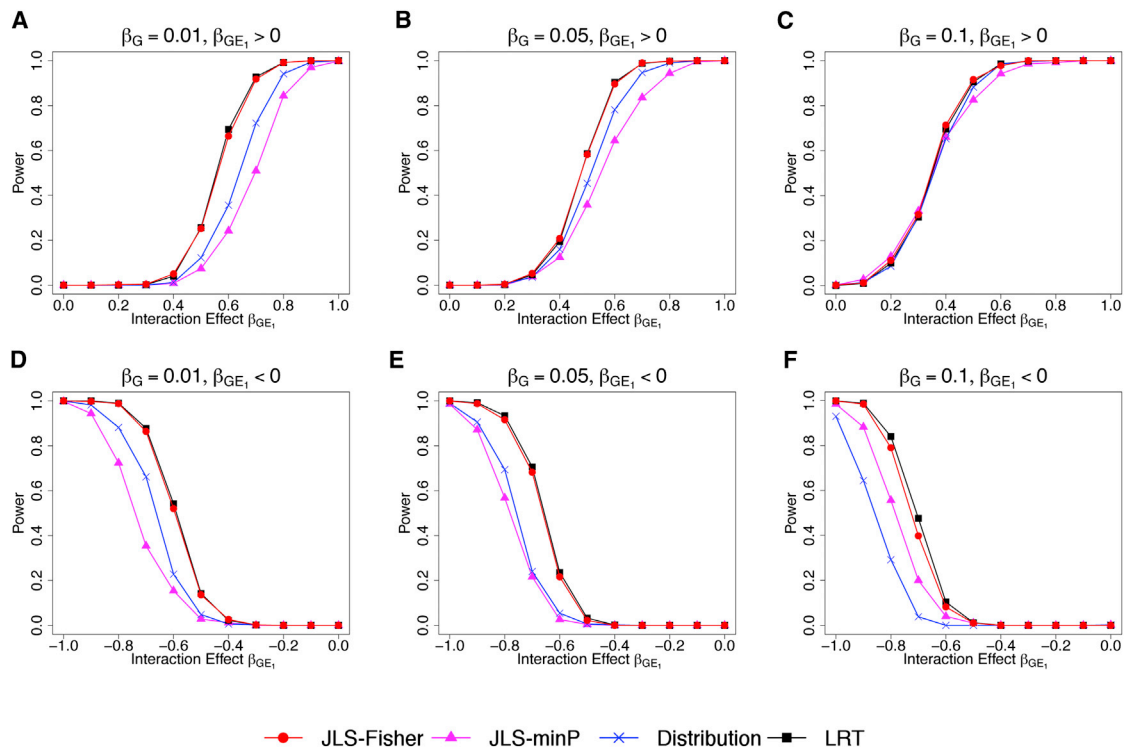
Figure 4. Power Comparison under Simulation Model 1

Four different testing methods are examined: the proposed JLS-Fisher (red) and JLS-minP (purple) methods, the distribution test (blue) of Aschard et al.,[16] and the LRT (black) of Cao et al.[17] Phenotype values for 2,000 independent subjects were simulated under Model 1, $E[Y] = \beta_G G + \beta_{E1} E_1 + \beta_{GE1} G \cdot E_1$, where the MAF of G was 0.3 and the exposure variable E_1 was simulated as a Bernoulli variable with frequency = 0.3. The effect of the exposure E_1, β_{E1}, was fixed at 0.3 while the other effects varied.

(A–C) Results when the main genetic effect β_G and the interaction effect β_{GE1} are in the same direction.

(D–F) Results when β_G and β_{GE1} are in opposing direction.

Power was calculated at the 5×10^{-8} level based on 500 replicates. For other simulation details, see Appendix A. Additional power results are in Figures S3–S5.

phenotype, only the permutation (resampling-based) p value analysis will be valid (see step 4b).

2. Choose the individual location and scale tests based on the distribution of phenotype (normal or non-normal) or preference (for example, parametric or non-parametric versions of each test). In the present paper, our phenotypes were normally distributed (after transformation) and we chose linear regression and Levene's test for the location and scale tests, respectively.

3. Choose a JLS testing method of combining information from the individual location and scale tests and calculate the JLS test statistic. We acknowledge that there is no "most powerful" method for all situations in practice. Based on our experience, we recommend the use of Fisher's method (JLS-Fisher) of combining the association evidence:

$$W_F = -2\big(\log(p_L) + \log(p_S)\big),$$

where p_L and p_S are the individual location and scale test p values, respectively.

4. Chose the p value estimation method for the JLS statistic.

(a) Based on the approximate asymptotic distribution of the JLS test statistic: For the JLS-Fisher example, W_F is distributed as a χ^2_4 random variable, if the chosen individual location and scale tests are independent of each other under the null hypothesis. This assumption is correct if the trait is normally distributed and if the location-only test statistic is a function of the complete sufficient statistic (e.g., linear regression t-statistic, ANOVA F-statistic) and the distribution of the scale-only test statistic does not depend on the model parameters (e.g., Levene's test or the F-test for equality of variances).

(b) Based on resampling methods such as permutation:

- Calculate the observed JLS test statistic, e.g., W_F
- Choose the number of permutation replicates, K, based on the desired p value accuracy.
- Permute the phenotype independently K times (not valid if subjects are correlated with each other), and for each replicate k, recalculate the JLS test statistic, W_F^k, $k = 1, ..., K$.
- Obtain the permutation p value as [the number of $W_F^K > W_F$]/K.

Simulation Models

The following three models, as previously considered by Aschard et al.,[16] were used to simulate the data:

Model 1 : $E[Y] = \beta_G G + \beta_{E1} E_1 + \beta_{GE1} G \cdot E_1$
Model 2 : $E[Y] = \beta_{E1} E_1 + \beta_{E2} E_2 + \beta_{GE1} G \cdot E_1 + \beta_{GE2} G \cdot E_2$
Model 3 : $E[Y] = \beta_{GE1} G \cdot E_1$

For all three models, the observed genetic variant (G) was coded additively with minor allele frequency (MAF) of 0.3. Y was simulated from models with varying effects (βs) and residual variation (ε) following a standard normal distribution (mean = 0, standard deviation = 1).

Model 1 is analogous to Equation 1 where Y depended on the main effects of both G and E_1 and an interaction effect between G and E_1. The unobserved exposure variable E_1 was binary with frequency 0.3. The main genetic effect β_G took on values of 0.01, 0.05, and 0.1, and the interaction effect β_{GE1} was varied between -1 and 1 by a grid of 0.1. The main exposure effect β_{E1} was fixed at 0.3 when β_{GE1} was positive and -0.3 when β_{GE1} was negative.

For Model 2, Y was a function of main effects due to two unobserved exposures (E_1 and E_2; both binary with frequency 0.3) and interaction effects between the exposures and G. β_{GE1} was always positive and less than 1, whereas β_{GE2} was varied between -1 and 1 by a grid of 0.1. β_{E1} was fixed at 0.3, whereas β_{E2} was fixed at 0.3 when β_{GE2} was positive and -0.3 when β_{GE2} was negative.

For Model 3, Y depended only on the interaction between G and E_1. For this model, the interaction effect β_{GE1} and exposure frequency were chosen such that the observed marginal effect of G was fixed at 10% of the trait standard deviation.

In all cases, the working association model corresponded to Equation 2 because information on E_1 and E_2 was assumed to be unavailable.

To assess the type 1 error level of the joint location-scale methods at 0.05, 0.005, and 0.0005 levels, we simulated 100,000 replicate samples of n = 2,000 subjects each from the null model with no genetic association (i.e., $\beta_G = 0$ and $\beta_{GE} = 0$). (Results of n = 1,000 and 4,000 are qualitatively similar.) To examine the behavior of the testing methods under small group sizes, we conducted additional simulations under varied MAF (0.3, 0.2, 0.1, 0.05, and 0.03) as well as under fixed genotype group sizes where the rare homozygote group size was small (2, 5, 7, 10, 15, or 20) with the other genotype group sizes determined with respect to Hardy-Weinberg equilibrium. For comparison, empirical type 1 error rates of the individual location-only and scale-only tests are also studied, in addition to the JLS-Fisher and JLS-minP tests, and the LRT of Cao et al.;[17] the distribution test of Aschard et al.[16] has the correct type 1 error by design. Type 1 error control at the genome-wide level was assessed by phenotype-permutation analysis of the 866,995 T1D GWAS SNPs and 565,884 CF GWAS SNPs.

For the sensitivity analysis of genotype imputation uncertainty, simulated true genotypes were converted to probabilistic genotype data using a Dirichlet distribution with scale parameters a for the correct genotype category and $(1 - a)/2$ for the other two;[39] a was fixed at values of 1, 0.9, 0.8, 0.7, 0.6, and 0.5. Based on the simulated posterior probabilities, the most-likely genotype for each subject was the genotype with the highest posterior probability (i.e., the "hard call"); the incorrect call rate under this Dirichlet model ranges from 0% to 50% on average. The most-likely genotypes were then used to assess type 1 error control at the 0.05, 0.005, and 0.0005 levels, using 100,000 simulated replicate samples of n = 2,000, under the null model of no genetic association (i.e., $\beta_G = 0$ and $\beta_{GE} = 0$), and MAF = 0.3 for each level of genotype imputation uncertainty (a).

For power evaluation, as in Aschard et al.,[16] the results presented focused on MAF = 0.3 and n = 2,000 for Models 1 and 2 and n = 4,000 for Model 3. Power (at the 5×10^{-8} level) was estimated from 500 replicates, based on asymptotic p values of the tests considered, with the exception of the distribution test. For the distribution test, p values required estimation by permutation, and corresponding power results were from Aschard et al.,[16] kindly provided by Drs. Aschard and Kraft.

Lemma 1: Independence of Location-only and Scale-only Test Statistics under the Null Hypothesis for Normally Distributed Traits

Let $T_{Location} = \widehat{\beta}_1/(S/\sqrt{S_{xx}})$ be our location-only test statistic, testing the linear effect of x on Y in a sample of size n, where $S^2 = (1/(n-2))\sum(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i)^2$, $\widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1\overline{x}$, $\widehat{\beta}_1 = S_{xy}/S_{xx}$, $S_{xy} = \sum(x_i - \overline{x})(y_i - \overline{y})$, and $S_{xx} = \sum(x_i - \overline{x})^2$ ($x = G$ in Equation 2), and let T_{Scale} be our scale-only test statistic, here defined as Levene's test statistic for equality of variances.[14]

Lemma 1: For the conditional normal model $Y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma^2_{x_i})$, where $x_i = 0, 1$ or 2, $T_{Location}$ and T_{Scale} are independent if $\sigma^2_0 = \sigma^2_1 = \sigma^2_2$.

Proof: For fixed x, Y is normally distributed with constant variance σ^2 and mean $E[Y|x] = \beta_0 + \beta_1 x$. The density of Y is

$$(2\pi\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}\sum\left(y_i - \beta_0 - \beta_1 x_i\right)^2\right].$$

This is an exponential family with three parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) = (\beta_1/\sigma^2, -1/2\sigma^2, \beta_0/\sigma^2)$ for which the sufficient statistics $\boldsymbol{T} = (T_1, T_2, T_3) = (\sum x_i y_i, \sum y_i^2, \sum y_i)$ are complete. If $\sigma^2_0 = \sigma^2_1 = \sigma^2_2$, T_{Scale} is approximately distributed as a $F_{3-1,\,n-3}$ variable, and it does not depend on $\boldsymbol{\theta}$ (i.e., T_{Scale} is ancillary for $\boldsymbol{\theta}$). Thus, T_{Scale} is independent of $\boldsymbol{T}$ (see page 152 in Lehmann and Romano[40]). Because $T_{Location}$ is a function of $\boldsymbol{T}$, $T_{Location}$ and T_{Scale} are therefore independent under the null.

Note that the proof of independence holds regardless of the version of Levene's test statistic chosen, provided that the approximation to the F distribution (or some other distribution not depending on $\boldsymbol{\theta}$) is justifiable. Similar

statements of independence with analogous proofs can be obtained for other choices of location test statistics such as the analysis of variance (ANOVA) F-statistic.

Additional Considerations for the JLS Framework

In the most extreme scenario of simulation, i.e., Model 3, when there were no main G or E effects, the interaction effect was large ($\beta_{GE1} = 2$), and the unobserved exposure was rare (prob($E_1 = 1$) = 0.05), the distribution test was observed to be more powerful than the JLS-Fisher test (0.916 versus 0.406) (row 1 of Table S6). This is because the resulting phenotype distributions across genotypes differed in shape and their differences were not well captured by only mean and variance parameters. (The JLS-Fisher test could, in theory, be extended to include the skewness parameter or higher moments of such a distribution. The gain in power for this particular setting, however, would be associated with power loss for other approximately normally distributed traits.) In this scenario, when the phenotype deviated from normality, the LRT method appeared to be most powerful (power = 0.996). However, further analysis using permutation estimation of p values showed that power of the LRT under asymptotic analysis was greatly inflated, whereas the proposed JLS methods were robust (Table S7).

The alternative distribution[16] and LRT[17] joint testing methods were proposed for analysis of single SNPs. In principle, they can be extended to gene-set or pathway analysis, but multiple issues arise in implementation. The distribution test statistic depends on the size of the genotype groups for each variant, so it is not clear what the best strategy is to combine the statistics across SNPs with different MAFs. The LRT method is sensitive to the normality assumption of the phenotype distribution and to small group size of the genotype distribution (Tables S1 and S2).

The proposed multivariate JLS testing method is also extremely relevant for single-variant association analysis. In GWAS or Next-Generation Sequencing (NGS) settings where millions or tens of millions of SNPs are investigated, rapid screening of the whole genome to correctly prioritize SNPs for further examination demands methods that are powerful, yet computationally efficient. The proposed JLS testing method is robust and easy-to-implement, suitable for large-scale whole-genome scans, and can reveal individual genetic variants with main and/or interaction effects without the need to explicitly specify the interacting genetic and/or environmental variables. Compared with the distribution test and LRT alternatives, our method combines both simplicity of implementation and robustness to small size of the rare homozygous genotype group.

Our simulation analyses also demonstrated that the location (regression) and scale (Levene's) tests, and consequently our JLS-Fisher and JLS-minP tests, are robust to poorly captured genotype data. These findings agree with the results of Kutalik et al.[41] where minimal to no bias in false positive rates and type 1 error was found for location-only association testing of quantitative phenotypes with uncertain genotypes, specifically when imputed genotype probabilities were converted to most-likely genotype categories and analyzed without explicitly accounting for the uncertainty.

Violations of the normal data assumption can affect the type 1 error level of the proposed JLS test method as it does the LRT approach (increased type 1 error rate is more severe with LRT). This is largely due to the assumption of independence between the individual location-only and scale-only tests, which is required for the χ^2_4 approximation of the JLS-Fisher statistic. To circumvent this issue, investigators might choose to rely on a permutation distribution when estimating p values. However, this aspect can impose computation challenges at the genome-wide level. An alternative approach is to model the dependency between the individual location and scale test statistics and obtain an adjusted distribution for the JLS statistic. These adjusted distributions have been explored elsewhere under complete specification of the dependency structure between the test statistics being combined.[42–44] However, it remains to be explored how to model the dependency between the individual location and scale association test statistics at different loci across the genome.

In consideration of the choices available for the individual location and scale tests and the methods of combining information from these individual components for each variant, we recognize that there is no single, most powerful method for all circumstances.[20,22] We showed that for normally distributed traits, the JLS-Fisher test statistic W_F follows a χ^2_4 distribution under the null hypothesis, as long as the location test statistic is a function of the complete sufficient statistic (e.g., linear regression t-statistic, ANOVA F-statistic) and the distribution of the scale test statistic does not depend on the model parameters (e.g., Levene's test or the F-test for equality of variances). In practice, when the normality assumption is violated or other JLS tests are preferred, permutation-based p value evaluation can be used with increased computational cost.

The proposed framework can be easily extended for meta-analysis, where the sample- or study-specific association test statistics or p values to be combined across samples are obtained from the JLS testing application instead of the typical location-only testing method. The sample-specific choices of the individual location and scale tests need not be identical across different studies, as long as p values of the JLS tests are valid within each study. However, choice of optimal weighting factors assigned to individual samples requires further investigation. Analyzing imputed SNPs (explicitly incorporating the genotype probabilities) or correlated subjects (i.e., pedigree family data) using the proposed JLS framework would require development of appropriate scale-only testing methods; location-only methods for these more complex settings are already available.[39,45–49]

Web Resources

The URLs for data presented herein are as follows:

GWAS data for the type 1 diabetes study, DCCT/EDIC, available through dbGaP, http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000086.v3.p1
JLS analysis tool, http://www.utstat.toronto.edu/sun/
OMIM, http://www.omim.org/

References

1. Cordell, H.J. (2009). Detecting gene-gene interactions that underlie human diseases. Nat. Rev. Genet. *10*, 392–404.
2. Cornelis, M.C., Tchetgen, E.J., Liang, L., Qi, L., Chatterjee, N., Hu, F.B., and Kraft, P. (2012). Gene-environment interactions in genome-wide association studies: a comparative study of tests applied to empirical studies of type 2 diabetes. Am. J. Epidemiol. *175*, 191–202.
3. Mukherjee, B., Ahn, J., Gruber, S.B., and Chatterjee, N. (2012). Testing gene-environment interaction in large-scale case-control association studies: possible choices and comparisons. Am. J. Epidemiol. *175*, 177–190.
4. Maciejewski, H. (2014). Gene set analysis methods: statistical models and methodological differences. Brief. Bioinform. *15*, 504–518.
5. Sun, L., Rommens, J.M., Corvol, H., Li, W., Li, X., Chiang, T.A., Lin, F., Dorfman, R., Busson, P.F., Parekh, R.V., et al. (2012). Multiple apical plasma membrane constituents are associated with susceptibility to meconium ileus in individuals with cystic fibrosis. Nat. Genet. *44*, 562–569.
6. Ritchie, M.D., Hahn, L.W., Roodi, N., Bailey, L.R., Dupont, W.D., Parl, F.F., and Moore, J.H. (2001). Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am. J. Hum. Genet. *69*, 138–147.
7. Welsh, M.J., Ramsey, B.W., Accurso, F.J., and Cutting, G.R. (2001). Cystic fibrosis. In The Metabolic and Molecular Bases of Inherited Disease, C.R. Schriver, A.L. Beaudet, W.S. Sly, and D. Valle, eds. (McGraw-Hill), pp. 5121–5188.
8. Blackman, S.M., Deering-Brose, R., McWilliams, R., Naughton, K., Coleman, B., Lai, T., Algire, M., Beck, S., Hoover-Fong, J., Hamosh, A., et al. (2006). Relative contribution of genetic and nongenetic modifiers to intestinal obstruction in cystic fibrosis. Gastroenterology *131*, 1030–1039.
9. Collaco, J.M., Vanscoy, L., Bremer, L., McDougal, K., Blackman, S.M., Bowers, A., Naughton, K., Jennings, J., Ellen, J., and Cutting, G.R. (2008). Interactions between secondhand smoke and genes that affect cystic fibrosis lung disease. JAMA *299*, 417–424.
10. Goss, C.H., Newsom, S.A., Schildcrout, J.S., Sheppard, L., and Kaufman, J.D. (2004). Effect of ambient air pollution on pulmonary exacerbations and lung function in cystic fibrosis. Am. J. Respir. Crit. Care Med. *169*, 816–821.
11. Collaco, J.M., McGready, J., Green, D.M., Naughton, K.M., Watson, C.P., Shields, T., Bell, S.C., Wainwright, C.E., Cutting, G.R., and Cutting, G.R.; ACFBAL Study Group (2011). Effect of temperature on cystic fibrosis lung disease and infections: a replicated cohort study. PLoS ONE *6*, e27784.
12. Vanscoy, L.L., Blackman, S.M., Collaco, J.M., Bowers, A., Lai, T., Naughton, K., Algire, M., McWilliams, R., Beck, S., Hoover-Fong, J., et al. (2007). Heritability of lung disease severity in cystic fibrosis. Am. J. Respir. Crit. Care Med. *175*, 1036–1043.
13. Paré, G., Cook, N.R., Ridker, P.M., and Chasman, D.I. (2010). On the use of variance per genotype as a tool to identify quantitative trait interaction effects: a report from the Women's Genome Health Study. PLoS Genet. *6*, e1000981.
14. Levene, H. (1960). Robust tests for equality of variances. In Contributions to Probability and Statistics; Essays in Honor of Harold Hotelling, I. Olkin, ed. (Stanford University Press), pp. 278–292.
15. Deng, W.Q., Asma, S., and Pare, G. (2014). Meta-analysis of SNPs involved in variance heterogeneity using Levene's test for equal variances. Eur. J. Hum. Genet. *22*, 427–430.
16. Aschard, H., Zaitlen, N., Tamimi, R.M., Lindström, S., and Kraft, P. (2013). A nonparametric test to detect quantitative trait loci where the phenotypic distribution differs by genotypes. Genet. Epidemiol. *37*, 323–333.
17. Cao, Y., Wei, P., Bailey, M., Kauwe, J.S., and Maxwell, T.J.; Alzheimer's Disease Neuroimaging Initiative (2014). A versatile omnibus test for detecting mean and variance heterogeneity. Genet. Epidemiol. *38*, 51–59.
18. Paterson, A.D., Waggott, D., Boright, A.P., Hosseini, S.M., Shen, E., Sylvestre, M.P., Wong, I., Bharaj, B., Cleary, P.A.,

Lachin, J.M., et al.; MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium); Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group (2010). A genome-wide association study identifies a novel major locus for glycemic control in type 1 diabetes, as measured by both A1C and glucose. Diabetes *59*, 539–549.

19. Li, W., Soave, D., Miller, M.R., Keenan, K., Lin, F., Gong, J., Chiang, T., Stephenson, A.L., Durie, P., Rommens, J., et al. (2014). Unraveling the complex genetic model for cystic fibrosis: pleiotropic effects of modifier genes on early cystic fibrosis-related morbidities. Hum. Genet. *133*, 151–161.

20. Derkach, A., Lawless, J.F., and Sun, L. (2013). Robust and powerful tests for rare variants using Fisher's method to combine evidence of association from two or more complementary tests. Genet. Epidemiol. *37*, 110–121.

21. Derkach, A., Lawless, J.F., and Sun, L. (2014). Pooled association tests for rare genetic variants: a review and some new results. Stat. Sci. *29*, 302–321.

22. Owen, A.B. (2009). Karl Pearson's meta-analysis revisited. Ann. Stat. *37*, 3867–3892.

23. Marozzi, M. (2013). Nonparametric simultaneous tests for location and scale testing: s comparison of several methods. Commun. Stat. Simul. Comput. *42*, 1298–1317.

24. Taylor, C., Commander, C.W., Collaco, J.M., Strug, L.J., Li, W., Wright, F.A., Webel, A.D., Pace, R.G., Stonebraker, J.R., Naughton, K., et al. (2011). A novel lung disease phenotype adjusted for mortality attrition for cystic fibrosis genetic modifier studies. Pediatr. Pulmonol. *46*, 857–869.

25. Wright, F.A., Strug, L.J., Doshi, V.K., Commander, C.W., Blackman, S.M., Sun, L., Berthiaume, Y., Cutler, D., Cojocaru, A., Collaco, J.M., et al. (2011). Genome-wide association and linkage identify modifier loci of lung disease severity in cystic fibrosis at 11p13 and 20q13.2. Nat. Genet. *43*, 539–546.

26. Li, Y., Willer, C., Sanna, S., and Abecasis, G. (2009). Genotype imputation. Annu. Rev. Genomics Hum. Genet. *10*, 387–406.

27. Li, Y., Willer, C.J., Ding, J., Scheet, P., and Abecasis, G.R. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet. Epidemiol. *34*, 816–834.

28. Dudbridge, F., and Gusnanto, A. (2008). Estimation of significance thresholds for genomewide association scans. Genet. Epidemiol. *32*, 227–234.

29. Dorfman, R., Taylor, C., Lin, F., Sun, L., Sandford, A., Paré, P., Berthiaume, Y., Corey, M., Durie, P., and Zielenski, J.; Members of Canadian Consortium for CF Genetic Studies (2011). Modulatory effect of the SLC9A3 gene on susceptibility to infections and pulmonary function in children with cystic fibrosis. Pediatr. Pulmonol. *46*, 385–392.

30. Donowitz, M., and Li, X. (2007). Regulatory binding partners and complexes of NHE3. Physiol. Rev. *87*, 825–872.

31. Sun, F., Hug, M.J., Lewarchik, C.M., Yun, C.H., Bradbury, N.A., and Frizzell, R.A. (2000). E3KARP mediates the association of ezrin and protein kinase A with the cystic fibrosis transmembrane conductance regulator in airway cells. J. Biol. Chem. *275*, 29539–29546.

32. Kato, A., and Romero, M.F. (2011). Regulation of electroneutral NaCl absorption by the small intestine. Annu. Rev. Physiol. *73*, 261–281.

33. Seidler, U., Singh, A.K., Cinar, A., Chen, M., Hillesheim, J., Hogema, B., and Riederer, B. (2009). The role of the NHERF family of PDZ scaffolding proteins in the regulation of salt and water transport. Ann. N Y Acad. Sci. *1165*, 249–260.

34. Sun, X., Elston, R., Morris, N., and Zhu, X. (2013). What is the significance of difference in phenotypic variability across SNP genotypes? Am. J. Hum. Genet. *93*, 390–397.

35. Struchalin, M.V., Dehghan, A., Witteman, J.C., van Duijn, C., and Aulchenko, Y.S. (2010). Variance heterogeneity analysis for detection of potentially interacting genetic loci: method and its limitations. BMC Genet. *11*, 92.

36. Dudbridge, F., and Fletcher, O. (2014). Gene-environment dependence creates spurious gene-environment interaction. Am. J. Hum. Genet. *95*, 301–307.

37. Wood, A.R., Tuke, M.A., Nalls, M.A., Hernandez, D.G., Bandinelli, S., Singleton, A.B., Melzer, D., Ferrucci, L., Frayling, T.M., and Weedon, M.N. (2014). Another explanation for apparent epistasis. Nature *514*, E3–E5.

38. Hutter, C.M., Mechanic, L.E., Chatterjee, N., Kraft, P., and Gillanders, E.M.; NCI Gene-Environment Think Tank (2013). Gene-environment interactions in cancer epidemiology: a National Cancer Institute Think Tank report. Genet. Epidemiol. *37*, 643–657.

39. Acar, E.F., and Sun, L. (2013). A generalized Kruskal-Wallis test incorporating group uncertainty with application to genetic association studies. Biometrics *69*, 427–435.

40. Lehmann, E.L., and Romano, J.P. (2005). Testing statistical hypotheses. In Springer Texts in Statistics (Springer), p. 784.

41. Kutalik, Z., Johnson, T., Bochud, M., Mooser, V., Vollenweider, P., Waeber, G., Waterworth, D., Beckmann, J.S., and Bergmann, S. (2011). Methods for testing association between uncertain genotypes and quantitative traits. Biostatistics *12*, 1–17.

42. Kost, J.T., and McDermott, M.P. (2002). Combining dependent P-values. Stat. Probab. Lett. *60*, 183–190.

43. Chuang, L.-L., and Shih, Y.-S. (2012). Approximated distributions of the weighted sum of correlated chi-squared random variables. J. Stat. Plan. Inference *142*, 457–472.

44. Hou, C.-D. (2005). A simple approximation for the distribution of the weighted combination of non-independent or independent probabilities. Stat. Probab. Lett. *73*, 179–187.

45. Aulchenko, Y.S., Struchalin, M.V., and van Duijn, C.M. (2010). ProbABEL package for genome-wide association analysis of imputed data. BMC Bioinformatics *11*, 134.

46. Schaid, D.J., Rowland, C.M., Tines, D.E., Jacobson, R.M., and Poland, G.A. (2002). Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am. J. Hum. Genet. *70*, 425–434.

47. Horvath, S., Xu, X., and Laird, N.M. (2001). The family based association test method: strategies for studying general genotype–phenotype associations. Eur. J. Hum. Genet. *9*, 301–306.

48. Lou, X.Y., Chen, G.B., Yan, L., Ma, J.Z., Mangold, J.E., Zhu, J., Elston, R.C., and Li, M.D. (2008). A combinatorial approach to detecting gene-gene and gene-environment interactions in family studies. Am. J. Hum. Genet. *83*, 457–467.

49. Yu, Z. (2012). Family-based association tests using genotype data with uncertainty. Biostatistics *13*, 228–240.

50. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J. (2010). LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics *26*, 2336–2337.

51. Luna, A., and Nicodemus, K.K. (2007). snp.plotter: an R-based SNP/haplotype association and linkage disequilibrium plotting package. Bioinformatics *23*, 774–776.

A Joint Location-Scale Test Improves Power to Detect

Associated SNPs, Gene-Sets, and Pathways

David Soave, Harriet Corvol, Naim Panjwani, Jiafen Gong, Weili Li, Pierre-Yves Boëlle,
Peter Durie, Andrew D. Paterson, Johanna M. Rommens, Lisa J. Strug, and Lei Sun
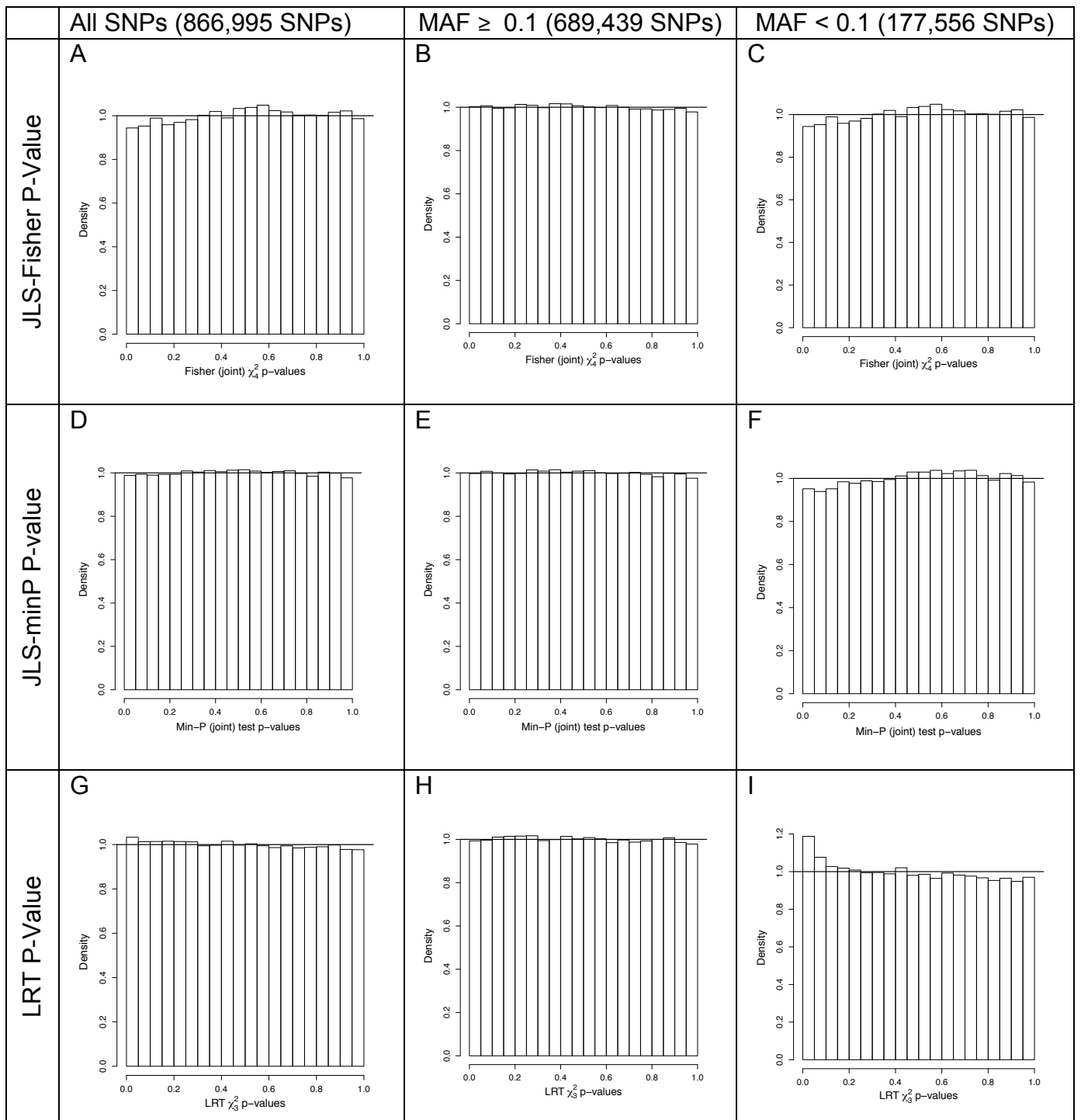
Figure S1. Type 1 error comparison of the joint testing methods – GWAS of HbA1c in type 1 diabetes (T1D). Following a permutation of the T1D HbA1c phenotype (Inverse Normal Transform on average of 38 quarterly measured values), GWAS was conducted on the DCCT/EDIC sample (n=1304 subjects) using (A-C) the JLS-Fisher test, (D-F) the JLS-minP test and (G-I) the LRT methods, across all SNPs (A,D,G), as well as stratifying by MAF ≥ 0.1 (B,E,H) and MAF < 0.1 (C,F,I).
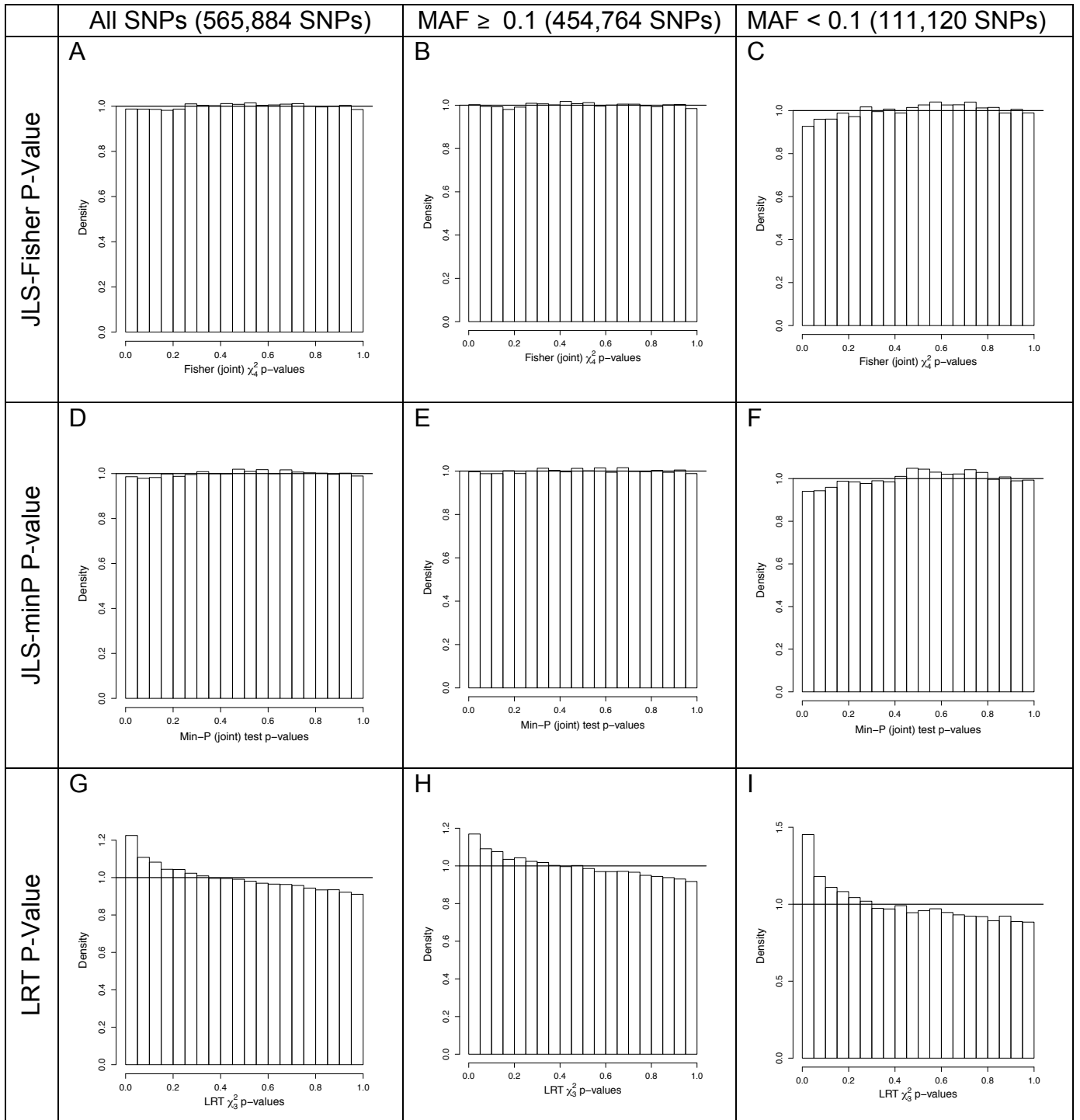
Figure S2. Type 1 error comparison of the joint testing methods – GWAS of cystic fibrosis (CF) lung disease as measured by SaKnorm. Following a permutation of SaKnorm in the CGS sample (n=1409 patients), GWAS was conducted using (A-C) the JLS-Fisher test, (D-F) the JLS-minP test and (G-I) the LRT methods across all SNPs (A,D,G), as well as stratifying by MAF ≥ 0.1 (B,E,H) and MAF<0.1 (C,F,I).
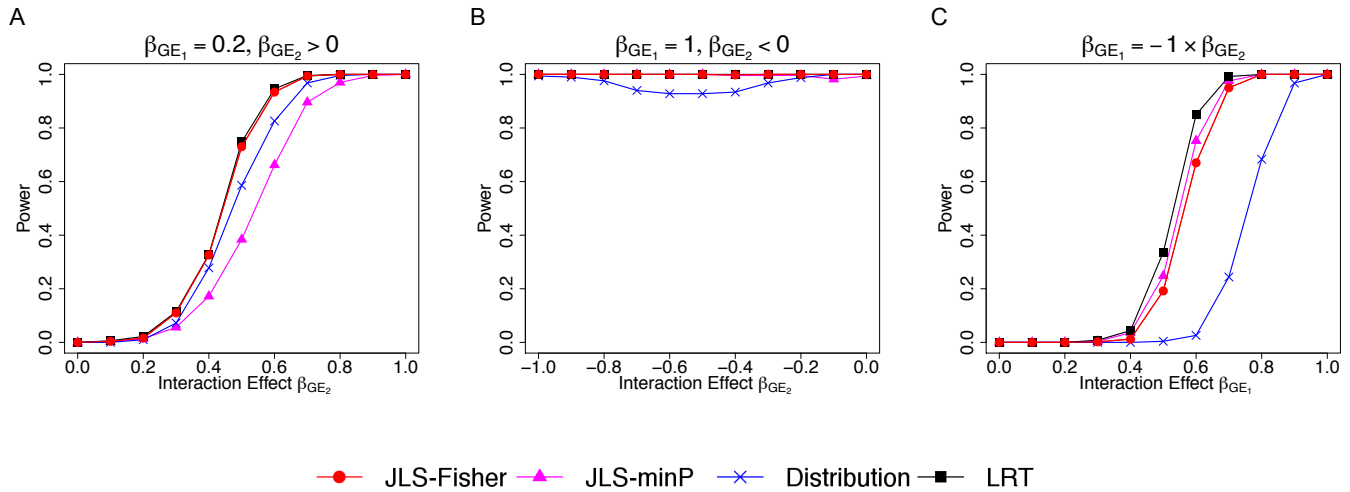
Figure S3. Power comparison under simulation Model 2 – Proposed and competing joint location-scale testing methods. Four different joint location-scale testing methods are examined: the proposed JLS-Fisher (red) and JLS-minP (purple) tests, and the distribution test (blue) of Aschard et al. [1] and the LRT (black) of Cao et al. [2]. Phenotype values for 2000 independent subjects were simulated under Model 2, $E[Y] = \beta_{E1}E_1 + \beta_{E2}E_2 + \beta_{GE1}G{\cdot}E_1 + \beta_{GE2}G{\cdot}E_2$, where the MAF of G was 0.3 and the exposure variables E_1 and E_2 were simulated as Bernoulli variables with frequency=0.3. The effect of the exposure E_1, β_{E1}, was fixed at 0.3 while the interaction effect β_{GE1} varied. The effect of exposure E_2, β_{E2}, was fixed at 0.3 when the interaction effect β_{GE2} was positive, and -0.3 when β_{GE2} was negative. Results are presented for models when effects of the two interaction effects, β_{GE1} and β_{GE2}, are in the same direction (A), and when the two interaction effects are in opposite direction having different amplitude (B) or the same amplitude (C). Power was calculated at the 5×10^{-8} level based on 500 replicates. For other details see Appendix A.
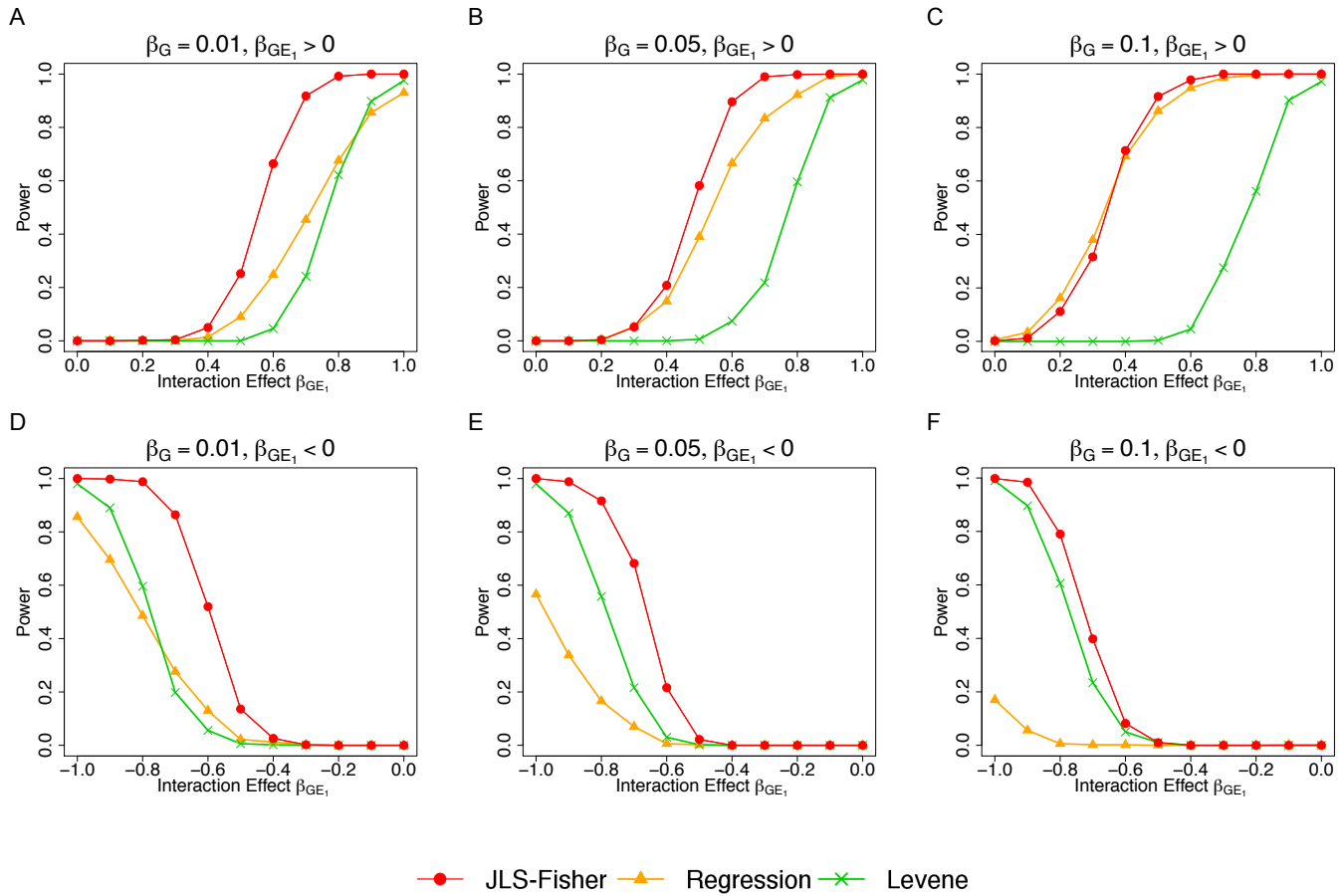
Figure S4. Power comparison under simulation Model 1 – Proposed joint location-scale testing method and individual location-only or scale-only testing methods. The proposed JLS-Fisher test (red) is compared to the individual regression location-only test (orange) and Levene's scale-only test (green). Phenotype values for 2000 independent subjects were simulated under Model 1, $E[Y] = \beta_G G + \beta_{E1} E_1 + \beta_{GE1} G \cdot E_1$, where the MAF of G was 0.3 and the exposure variable E_1 was simulated as a Bernoulli variable with frequency=0.3. The exposure effect β_{E1} was fixed at 0.3 while the other effects vary. Top panel (A)-(C) are results when the main genetic effect β_G and the interaction effect β_{GE1} are in the same direction, and the bottom panel (D)-(F) are results when β_G and β_{GE1} are in opposite direction. Power was calculated at the 5×10^{-8} level based on 500 replicates. For other details see Appendix A.

A $\beta_{GE_1} = 0.2, \beta_{GE_2} > 0$

B $\beta_{GE_1} = 1, \beta_{GE_2} < 0$

C $\beta_{GE_1} = -1 \times \beta_{GE_2}$

JLS-Fisher   Regression   Levene
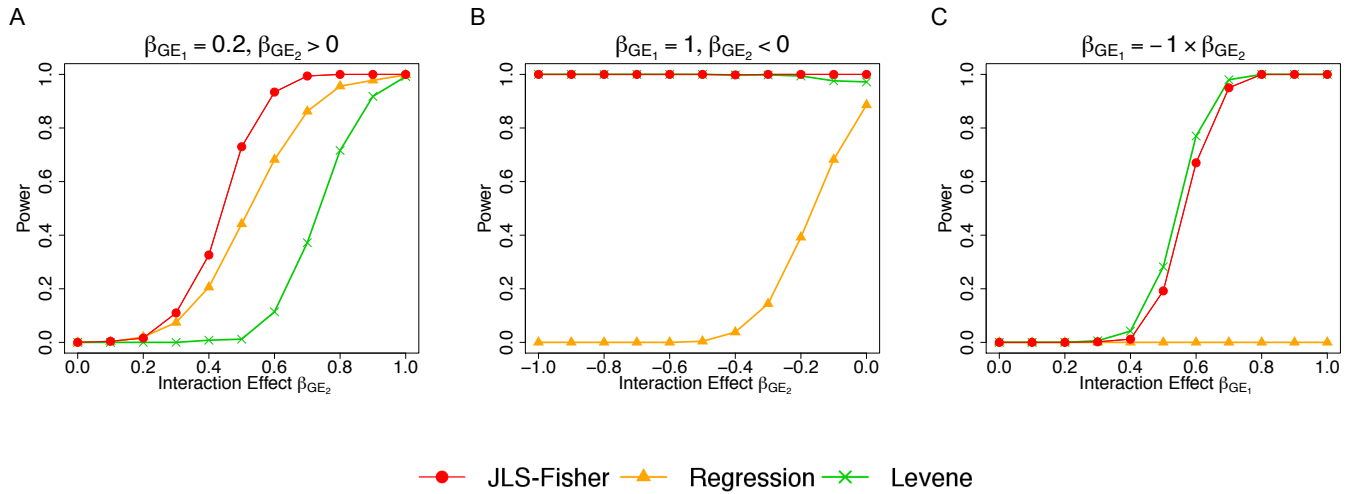
Figure S5. Power comparison under simulation Model 2 – Proposed joint location-scale testing method and individual location-only or scale-only testing methods. The proposed JLS-Fisher test (red) is compared to the individual regression location-only test (orange) and Levene's scale-only test (green). Phenotype values for 2000 independent subjects were simulated under Model 2, $E[Y] = \beta_{E1}E_1 + \beta_{E2}E_2 + \beta_{GE1}G \cdot E_1 + \beta_{GE2}G \cdot E_2$, where the MAF of G was 0.3 and the exposure variables E_1 and E_2 were simulated as Bernoulli variables with frequency=0.3. The exposure effect β_{E1} was fixed at 0.3 while the interaction effect β_{GE1} varied. The exposure effect β_{E2} was fixed at 0.3 when the interaction effect β_{GE2} was positive, and -0.3 when β_{GE2} was negative. Results are presented for models when the two interaction effects, β_{GE1} and β_{GE2}, are (A) in the same direction, and (B) when the two interaction effects are in opposite direction having different amplitude or (C) the same amplitude. Power was calculated at the 5×10^{-8} level based on 500 replicates. For other details see Appendix A.

Table S1. Type 1 error comparison with varied minor allele frequency (MAF).
Type 1 error is presented for the regression location test (Reg), Levene's scale test
(Levene), the proposed JLS-Fisher and JLS-minP joint location-scale tests, and the
LRT of Cao et al. [2]. The distribution test of Aschard et al. [1] has correct type 1 error by
design because it is a permutation-based testing method. Phenotype values for
2,000 independent subjects were simulated under the null model with no genetic
association (Model 1 with $\beta_G = 0, \beta_E = 0, \beta_{GE} = 0$) and varied MAF (0.03-0.3) with
residual variation from a normal distribution with mean 0 and standard deviation 1.
Empirical type 1 error rates were calculated at the 0.05, 0.005, and 0.0005 nominal
levels based on 100,000 replicates. For other details see Appendix A.

MAF=0.3					
	Individual		Joint		
level	Reg	Levene	JLS-Fisher	JLS-MinP	LRT
0.05	0.05005	0.04988	0.04931	0.04952	0.05144
0.005	0.00490	0.00497	0.00476	0.00488	0.00473
0.0005	0.00047	0.00056	0.00048	0.00045	0.00045

MAF=0.2					
	Individual		Joint		
level	Reg	Levene	JLS-Fisher	JLS-MinP	LRT
0.05	0.04983	0.04788	0.04869	0.04910	0.05060
0.005	0.00475	0.00473	0.00508	0.00481	0.00476
0.0005	0.00049	0.00054	0.00053	0.00058	0.00047

MAF=0.1					
	Individual		Joint		
level	Reg	Levene	JLS-Fisher	JLS-MinP	LRT
0.05	0.05101	0.04766	0.04888	0.04860	0.05254
0.005	0.00476	0.00435	0.00460	0.00452	0.00506
0.0005	0.00047	0.00058	0.00049	0.00049	0.00053

MAF=0.05					
	Individual		Joint		
level	Reg	Levene	JLS-Fisher	JLS-MinP	LRT
0.05	0.05026	0.04183	0.04603	0.04624	0.06130
0.005	0.00473	0.00498	0.00460	0.00524	0.00777
0.0005	0.00047	0.00064	0.00060	0.00063	0.00107

MAF=0.03					
	Individual		Joint		
level	Reg	Levene	JLS-Fisher	JLS-MinP	LRT
0.05	0.05026	0.04788	0.04900	0.04959	0.06722
0.005	0.00523	0.00583	0.00568	0.00592	0.01120
0.0005	0.00045	0.00110	0.00089	0.00087	0.00231

Table S2. Type 1 error comparison with varied genotypic group sizes. Type 1 error is presented for the regression location test (Reg), Levene's scale test (Levene), the proposed JLS-Fisher and JLS-minP joint location-scale tests, and the LRT of Cao et al. [2]. The distribution test of Aschard et al. [1] has correct type 1 error by design because it is a permutation-based testing method. Phenotype values for 2,000 independent subjects were simulated under the null model with no genetic association (Model 1 with $\beta_G = 0, \beta_E = 0, \beta_{GE} = 0$) with residual variation from a normal distribution with mean 0 and standard deviation 1. The genotype group sizes were fixed with respect to the smallest group size ($N_{smallest}= N_2 = 2, 5, 7, 10, 15$ or 20) for the rare homozygous group and the corresponding Hardy Weinberg equilibrium (N_0 and N_1) for the common homozygous and heterozygous groups. Empirical type 1 error rates were calculated at the 0.05, 0.005, and 0.0005 nominal levels based on 100,000 simulation replicates. For other details see Appendix A.

Group Sizes (N_0, N_1, N_2)=(1882,116,2) (MAF~0.03)					
	Individual		Joint		
level	Reg	Levene	JLS-Fisher	JLS-MinP	LRT
0.05	0.04890	0.05369	0.05350	0.05593	0.09134
0.005	0.00485	0.01002	0.00820	0.00862	0.02108
0.0005	0.00045	0.00221	0.00157	0.00166	0.00582

Group Sizes (N_0, N_1, N_2)=(1805,190,5) (MAF~0.05)					
	Individual		Joint		
level	Reg	Levene	JLS-Fisher	JLS-MinP	LRT
0.05	0.05039	0.03769	0.04263	0.04262	0.05869
0.005	0.00496	0.00356	0.00399	0.00411	0.00712
0.0005	0.00043	0.00043	0.00036	0.00042	0.00079

Group Sizes (N_0, N_1, N_2)=(1767,226,7) (MAF~0.06)					
	Individual		Joint		
level	Reg	Levene	JLS-Fisher	JLS-MinP	LRT
0.05	0.05087	0.04103	0.04578	0.04591	0.05677
0.005	0.00505	0.00406	0.00451	0.00488	0.00617
0.0005	0.00041	0.00061	0.00047	0.00063	0.00070

Group Sizes (N_0, N_1, N_2)=(1730,320,10) (MAF~0.07)					
	Individual		Joint		
level	Reg	Levene	JLS-Fisher	JLS-MinP	LRT
0.05	0.04999	0.04524	0.04776	0.04709	0.0542
0.005	0.00474	0.00506	0.00487	0.00502	0.00571
0.0005	0.00037	0.00071	0.00067	0.00065	0.00071

Group Sizes (N_0, N_1, N_2)=(1674,311,15) (MAF~0.085)					
	Individual		Joint		
level	Reg	Levene	JLS-Fisher	JLS-MinP	LRT

level	Reg	Levene	JLS-Fisher	JLS-MinP	LRT
0.05	0.05037	0.04539	0.04808	0.04845	0.05263
0.005	0.00486	0.00509	0.00471	0.00464	0.00553
0.0005	0.00046	0.00055	0.00050	0.00051	0.00061

Group Sizes (N_0, N_1, N_2)=(1620,360,20) (MAF~0.1)

	Individual		Joint		
level	Reg	Levene	JLS-Fisher	JLS-MinP	LRT
0.05	0.05090	0.04704	0.04953	0.04936	0.05319
0.005	0.00514	0.00414	0.00470	0.00472	0.00500
0.0005	0.00050	0.00057	0.00046	0.00060	0.00045

Table S3. SNP-by-Age interaction effect p-values for variants in Cystic Fibrosis SaKnorm susceptibility loci. Interaction effect p-values were obtained from linear regression models including the main effects of a SNP and the Age variable, and their interaction effect. Variants were previously identified as associated with meconium ileus[3]. Age (in years) was modeled as either a continuous or dichotomous variable (using different cut points of 16yrs, 18yrs and 20yrs).

Chr	Gene	SNP	BP[a]	Age Continuous	Age Dichotomous: Cut point		
					16yrs	18yrs	20yrs
1	*SLC26A9*	rs7512462	204,166,218	0.23	0.23	0.29	0.52
1	*SLC26A9*	rs4077468	204,181,380	0.04	0.08	0.10	0.17
1	*SLC26A9*	rs12047830	204,183,322	0.02	0.01	0.01	0.02
1	*SLC26A9*	rs7419153	204,183,932	0.03	0.09	0.07	0.11
5	*SLC9A3*	rs17563161	550,624	0.003	0.03	0.02	0.02
X	*SLC6A14*	rs12839137	115,479,578	0.26	0.23	0.34	0.26
X	*SLC6A14*	rs5905283	115,479,909	0.21	0.31	0.26	0.36
X	*SLC6A14*	rs3788766	115,480,867	0.009	0.03	0.03	0.02

[a] hg18 assembly (March 2006; NCBI36).

Table S4. Apical gene-specific association results in the Canadian Gene Modifier Study (CGS) sample (n=1409). A list of 155 genes was annotated as described in Sun et al. [3]. Genes with gene-based JLS-Fisher p-value < 0.1 are listed in rank order.

Gene	Chr	#SNPs[a]	Gene p-value[b]		
			JLS-Fisher	Location/ Regression	Scale/ Levene
SLC9A3	5	10	0.0001	<0.0001	0.0238
ACY3	11	7	0.0006	0.0003	0.3233
MREG	2	37	0.0076	0.0901	0.013
SLC9A3R2	16	10	0.012	0.0696	0.0263
EZR	6	10	0.0128	0.032	0.0635
SLC46A1	17	3	0.0136	0.0118	0.2588
CRB1	1	27	0.017	0.0671	0.0398
LMO7	13	56	0.0287	0.0462	0.1386
KCNMA1	10	225	0.033	0.293	0.0152
SI	3	10	0.0461	0.336	0.0207
PTK2	8	38	0.0506	0.1196	0.0668
LCT	2	11	0.0565	0.0119	0.9008
SLC34A2	4	12	0.0617	0.036	0.4264
STXBP3	1	12	0.0635	0.3105	0.0533
AJAP1	1	50	0.0703	0.022	0.6287
ATP6V0A4	7	39	0.0734	0.7735	0.0071
DPEP1	16	2	0.0738	0.0483	0.347
MUC1	1	3	0.0776	0.0656	0.2099
SLC9A4	2	55	0.0945	0.042	0.4244
SLC22A12	11	4	0.095	0.145	0.1178
SLC9A3R1	17	8	0.0956	0.0807	0.313

[a] The number of GWAS SNPs (MAF>0.02) within ±10 kb of the boundaries of indicated gene.

[b] Permutation-based p-values for the gene. See Material and Methods for details.

Table S5. Type 1 error control with genotype uncertainty. Type 1 error is presented for the regression location test (Reg), Levene's scale test (Levene), and the proposed JLS-Fisher and JLS-minP joint location-scale tests. Simulated true genotypes were converted to probabilistic genotype data using a Dirichlet distribution with scale parameters a for the correct genotype category and $(1-a)/2$ for the other two; a was fixed at values of 1, 0.9, 0.8, 0.7, 0.6 and 0.5, corresponding to group uncertainty ranging from 0% to 50%. Based on the simulated posterior probabilities, the most-likely genotype for each subject was the genotype with the highest posterior probability (i.e. the 'hard call'). The most-likely genotypes were then used to assess type 1 error control at the 0.05, 0.005 and 0.0005 levels, using 100,000 simulated replicate samples of n = 2000, under the null model of no genetic association (i.e. β_G = 0 and β_{GE} = 0), and MAF = 0.3 for each level of genotype imputation uncertainty (a). Average concordance is the average percent of agreement between true and most-likely genotypes across the simulation replicates. For other details see Appendix A.

Uncertainty=0% (a=1); Average Concordance=1				
	Individual		Joint	
level	Reg	Levene	JLS-Fisher	JLS-MinP
0.05	0.05039	0.05032	0.05005	0.05010
0.005	0.00489	0.00478	0.00498	0.00523
0.0005	0.00054	0.00052	0.00057	0.00057

Uncertainty=10% (a=0.9); Average Concordance=0.907				
	Individual		Joint	
level	Reg	Levene	JLS-Fisher	JLS-MinP
0.05	0.05012	0.05032	0.04964	0.04996
0.005	0.00528	0.00499	0.00535	0.00530
0.0005	0.00043	0.00053	0.00066	0.00050

Uncertainty=20% (a=0.8); Average Concordance=0.808				
	Individual		Joint	
level	Reg	Levene	JLS-Fisher	JLS-MinP
0.05	0.04862	0.04865	0.04780	0.04795
0.005	0.00459	0.00475	0.00424	0.00462
0.0005	0.00039	0.00049	0.00043	0.00047

Uncertainty=30% (a=0.7); Average Concordance=0.706				
	Individual		Joint	
level	Reg	Levene	JLS-Fisher	JLS-MinP
0.05	0.05082	0.05160	0.05081	0.05187
0.005	0.00540	0.00533	0.00534	0.00523
0.0005	0.00050	0.00051	0.00053	0.00043

Uncertainty=40% (a=0.6); Average Concordance=0.605				
	Individual		Joint	

0.05	0.04942	0.04943	0.04895	0.04987
0.005	0.00502	0.00489	0.00520	0.00505
0.0005	0.00054	0.00053	0.00050	0.00048

Uncertainty=50% (a=0.5); Average Concordance=0.508

	Individual		Joint	
level	Reg	Levene	JLS-Fisher	JLS-MinP
0.05	0.05025	0.04920	0.04967	0.04922
0.005	0.00495	0.00450	0.00500	0.00481
0.0005	0.00047	0.00040	0.00040	0.00047

Table S6. Power comparison under simulation Model 3. Power is presented for the regression location test (Reg), Levene's scale test (Levene), the proposed JLS-Fisher and JLS-minP joint location-scalue tests, the distribution (Dist.) test of Aschard et al. [1], and the LRT of Cao et al. [2]. Phenotype values for 4000 independent subjects were simulated under Model 3, $E[Y] = \beta_{GE1}G{\cdot}E_1$, where the MAF of G was 0.3. Results are presented for scenarios when the interaction effect, β_{GE1}, and exposure (E_1) frequency were chosen such that the observed marginal effect of G was fixed at 10% of the trait standard deviation. Power was calculated at the 5×10^{-8} level based on 500 replicates. For other details see Appendix A.

		Individual		Joint			
Freq-E_1	Int. Effect (β_{GE1})	Reg.	Levene	JLS-Fisher	JSL-minP	LRT	Dist.
0.05	2	0.010	0.110	0.406	0.090	0.996	0.916
0.1	1	0.040	0.040	0.378	0.054	0.740	0.178
0.2	0.5	0.038	0.000	0.098	0.026	0.110	0.064
0.3	0.33	0.110	0.000	0.086	0.084	0.084	0.028
0.5	0.2	0.080	0.000	0.046	0.064	0.036	0.054
1	0.1	0.076	0.000	0.042	0.060	0.036	0.040

Table S7. Power comparison under simulation Model 3 using p-values estimated based on the approximate asymptotic distribution of the test statistics vs. permutation-based method. Power is presented for the regression location-only test (Reg), Levene's scale-only test (Levene), the proposed JLS-Fisher and JLS-minP joint location-scale tests, the distribution (Dist.) test of Aschard et al. [1], and the LRT of Cao et al. [2]. Phenotype values for 1000 independent subjects were simulated under Model 3, $E[Y] = \beta_{GE1}G{\cdot}E_1$, where the MAF of G was 0.3 and the exposure variable E_1 was simulated as a Bernoulli variable with frequency=0.05, and the interaction effect β_{GE1} was fixed at 2. Power was calculated at the 0.01 significance level based on 500 replicates. For each replicate, permutation p-values were estimated from 10,000 iterations.

	Individual		Joint			
	Reg.	Levene	JLS-Fisher	JLS-minP	LRT	Dist.
P-value estimation method						
Asymptotic	0.190	0.266	0.390	0.288	0.932	NA
Permutation	0.192	0.268	0.364	0.288	0.698	0.812

Table S8. Minor allele frequency (MAF) of variants in Cystic Fibrosis SaKnorm susceptibility loci. Variants were previously identified as associated with meconium ileus[3]. Pediatric subsets (age cutoffs of 16, 18 and 20 years) included for comparison.

Chr	Gene	SNP	BP[a]	Full Sample MAF	Pediatric Subsets MAF		
				All ages (n=1409)	<16yrs (n=653)	<18yrs[b] (n=753)	<20yrs (n=830)
1	*SLC26A9*	rs7512462	204,166,218	0.41	0.38	0.39	0.39
1	*SLC26A9*	rs4077468	204,181,380	0.42	0.38	0.39	0.40
1	*SLC26A9*	rs12047830	204,183,322	0.49	0.48	0.48	0.48
1	*SLC26A9*	rs7419153	204,183,932	0.37	0.39	0.39	0.38
5	*SLC9A3*	rs17563161	550,624	0.26	0.26	0.26	0.26
X	*SLC6A14*	rs12839137	115,479,578	0.24	0.23	0.23	0.24
X	*SLC6A14*	rs5905283	115,479,909	0.49	0.51	0.50	0.49
X	*SLC6A14*	rs3788766	115,480,867	0.40	0.39	0.39	0.39

[a] hg18 assembly (March 2006; NCBI36).
[b] Pediatric subset cutoff (<18yrs) used in Li et al. [4]; subset here only includes unrelated subjects (n=753).

References

1. Aschard, H., Zaitlen, N., Tamimi, R.M., Lindstrom, S., and Kraft, P. (2013). A nonparametric test to detect quantitative trait Loci where the phenotypic distribution differs by genotypes. Genetic epidemiology 37, 323-333.
2. Cao, Y., Wei, P., Bailey, M., Kauwe, J.S., Maxwell, T.J., and Alzheimer's Disease Neuroimaging, I. (2014). A versatile omnibus test for detecting mean and variance heterogeneity. Genetic epidemiology 38, 51-59.
3. Sun, L., Rommens, J.M., Corvol, H., Li, W., Li, X., Chiang, T.A., Lin, F., Dorfman, R., Busson, P.F., Parekh, R.V., et al. (2012). Multiple apical plasma membrane constituents are associated with susceptibility to meconium ileus in individuals with cystic fibrosis. Nature genetics 44, 562-569.
4. Li, W., Soave, D., Miller, M.R., Keenan, K., Lin, F., Gong, J., Chiang, T., Stephenson, A.L., Durie, P., Rommens, J., et al. (2014). Unraveling the complex genetic model for cystic fibrosis: pleiotropic effects of modifier genes on early cystic fibrosis-related morbidities. Hum Genet 133, 151-161.