An exact, unifying framework for region-based association testing in family-based designs, including higher criticism approaches, SKATs, multivariate and burden tests

Julian Hecker[1,2*], F. William Townes[3], Priyadarshini Kachroo[1], Jessica Lasky-Su[1], John Ziniti[1], Michael H. Cho[1], Scott T. Weiss[1], Nan M. Laird[2], Christoph Lange[2]

[1]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

[2]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[3]Department of Computer Science, Princeton University, Princeton, New Jersey, USA

***Correspondence: rejhe@channing.harvard.edu**

Abstract

Analysis of rare variants in family-based studies remains a challenge. To perform a region/set-based association analysis of rare variants in family-based studies, we propose a general methodological framework that integrates higher criticism, maximum, SKATs, and burden approaches into the family-based association testing (FBAT) framework. Using the haplotype algorithm for FBATs to compute the conditional genotype distribution under the null hypothesis of Mendelian transmissions, virtually any association test statistics can be implemented in our approach and simulation-based or exact p-values can be computed without the need for asymptotic settings. Using simulations, we compare the features of the proposed test statistics in our framework with the existing region-based methodology for family-based studies under various scenarios. The tests of our framework outperform the existing approaches. We provide general guidelines for which scenarios, e.g., sparseness of the signals or local LD structure, which test statistic will have distinct power advantages over the others. We also illustrate our approach in an application to a whole-genome sequencing dataset with 897 asthmatic trios.

Introduction

In family-based association studies, the concept of Mendelian transmissions can be utilized to construct association tests that are robust against genetic confounding (Transmission Disequilibrium Tests (TDTs) [1] or Family-based Association Tests (FBATs) [2]). This feature of family-based association tests was fundamental to establish them as a popular tool in association mapping since the days of candidate gene studies. The robustness of the approach largely out-weighted the requirement of family data, i.e., having to recruit additional related study subjects, and reduced statistical power compared to population-based cohorts with the same sample size. Testing strategies have been proposed that allowed the incorporation of the association information at the population-level in family-based designs without compromising the robustness of the test statistic [3–7].

As genome-wide association studies (GWAS) became a standard research tool, and the multiple testing problem had to be addressed at a genome-wide level, researchers started to emphasize statistical power in their choice for study designs. Power and study design considerations substantially contributed to making population-based designs approaches the most popular choice in association analysis.

Now, as whole-genome sequencing (WGS) studies are replacing chip-based GWAS, region-based rare variant analysis approaches have moved to the center of the statistical methodology development, as, even for very large sample sizes, the power of single locus association analysis will be too small when the minor allele is rare. Region-based approaches are motivated by the idea that, if we can combine "association signals" across a pre-defined region in a suitable way, a stronger genetic signal could be assessed by a suitable test statistic, and the resulting region-based association test would have increased statistical power. At the same time, the multiple testing problem could become less severe as fewer association tests

are computed. The major statistical challenges for region-based tests are to identify suitable ways to combine the genetic information across the pre-specified region, to incorporate/estimate the correlation between the selected rare variants in the region-based test statistic, and to select a suitable test statistic.

For population-based designs, based on assumptions about the alternative hypothesis/distribution of disease susceptibility loci (DSLs) and their effect directions, numerous region-based tests have been proposed, e.g., burden tests and variance component tests [8,9]. However, population stratification is a potential problem in population-based designs that can be even more severe in the analysis of rare variants.

For family-based designs, the popular burden tests and the SKAT approach have been translated to the FBAT framework [10,11]. As with their population-based equivalents, these two approaches estimate the correlation between the genetic loci empirically. This, especially for rare variant data, can be problematic, as the rare variant allele counts are small, and the empirical estimates can be affected by numerical instabilities. Furthermore, the application of asymptotic theory may not provide accurate results here. Other recent approaches are based on mixed models and can, theoretically, analyze unrelated and related samples [12].

As an other alternative to the transmission-based FBAT approach, methods that compare the allele frequencies between affected and unaffected individuals within families have been suggested, e.g., Generalized Disequilibrium Test (GDT) for single variant analysis [13]. As they can incorporate all available phenotypic and genetic information within one pedigree, they can be more powerful than the corresponding FBAT analysis. However, they require the assumption that the allele frequencies and the genetic variance for all members of a pedigree are equal under the null hypothesis. This assumption can be violated in the presence of

population substructure within the founders of the families or when there are departures from Hardy Weinberg equilibrium [14]. None of these assumptions is required for the FBAT approach to be valid. For region-based analysis, the Rare-Variant Generalized Disequilibrium Test (RV-GDT) extension has been proposed [15].

In this communication, we will exclusively focus on transmission-based analysis approaches to construct association tests that are robust against confounding. We propose a general framework for region-based rare variant analysis in extended pedigree/nuclear families that is based on the FBAT approach. Multiple offspring per family may be available, founder/phase information can be missing, and phenotypes can be dichotomous or quantitative. In contrast to previous approaches for region-based analysis in population- and family-based designs, the joint distribution of the rare variants in the region is obtained analytically under the null hypothesis, conditioning on the sufficient statistic, using the haplotype algorithm for FBATs [16,17]. Based on the conditional genotype distribution, it is straightforward to implement region-based FBATs, e.g., multivariate tests, burden tests, SKAT [9,11], and higher criticism approaches [18–20]. As the joint distribution of the rare variants is obtained analytically and can efficiently be sampled from, the significance of the test statistics in our framework can be obtained either by simulations or the construction of the exact distribution[21]. This flexibility of our approach enables the implementation of virtually any region/set-based test without the need for any asymptotic assumptions or approximations. We illustrate the implementation of higher criticism approaches, maximum statistics, SKATs and burden tests in our framework.

For different scenarios, e.g., regions with sparse signals, varying local Linkage Disequilibrium (LD) structure, we compare our proposed FBAT framework to existing methodology, using extensive simulation studies. Our simulation results support our theoretical considerations that our testing framework provides a substantial improvement over the existing

methodology in terms of statistical power and robustness against population substructure. We also develop general recommendations for which choice of the test statistic is preferable for which scenario. Furthermore, we also applied our methodology framework to a whole-genome sequencing study for childhood asthma with 897 trios.

Methods

In a family-based WGS association study, genotype data for rare variants are available for a set of marker loci that are in close physical proximity and define a genomic segment that is suitable for region-based association analysis. The genotype information may be available for multiple offspring as well as for the parents. For the i-th nuclear family, we introduce the $p \times n_i$ genotype matrix X_i and the n_i dimensional phenotype vector T_i, where n_i denotes the number of offspring in the i-th nuclear family, and p denotes the number of variants in the analysis region. We regard X_i as random while T_i is fixed in the FBAT approach. Below, we propose a possible set of test statistics that can capture the potential association between the offspring genotype data and the phenotypes under various conditions.

Simulation-based significance testing

For each region, using the haplotype algorithm for FBAT [16,17], we derive the conditional distribution of offspring genotypes X_i in the i-th nuclear family under the null hypothesis, given the sufficient statistic S_i for the possible missing founder genotypes [22]. The sufficient statistic approach utilizes parental genotypes if they are available. The knowledge about the conditional genotype distribution allows constructing association tests that are robust against population stratification and admixture. Based on this conditional genotype distribution, it is straightforward to compute the first two moments of commonly used test statistics under the null hypothesis of no association and derive the asymptotic distribution. However, as the

analysis of rare variant data leads to scenarios where the application of asymptotic theory does not provide reliable approximations, simulation-based or even exact p-values [21] are preferred. Here, we propose to evaluate association p-values based on a sufficiently large number of simulated draws from the null distribution. This procedure can be combined with adaptive permutation/simulation-based p-value techniques. In this context, we recommend using stopping rules that are nearly optimal in terms of the required number of simulations [23].

This approach, therefore, also allows our testing framework to incorporate test statistics where the (asymptotic) distribution is intractable or cumbersome, e.g., maximum statistics or higher criticism approaches.

Test statistics

All test statistics under consideration are based on the following two objects. For the i-th family, we define the p-dimensional vector of Mendelian residuals $U_i = (X_i - E[X_i|S_i])T_i$. Also, we define the corresponding $p \times p$ variance matrix $V_i = Var(U_i|S_i, T_i)$. For both objects, the moments are computed under the null hypothesis, based on the sufficient statistic S_i.

Burden-based approaches

Burden-type FBATs can be implemented by specifying a p-dimensional weight vector W that collapses/summarizes the rare variant information of the region into a single scalar value. The specification of the weight vector W requires assumptions about the effect direction and its effect size. In this context, the contribution to the FBAT statistic of the i-th family is then given by

$$U_i^* = W^T(X_i - E[X_i|S_i])T_i$$

The corresponding FBAT-statistic for the simulation-based testing is computed by $FBAT_{burden} = (\sum_i U_i^*)^2$. We note that for this burden test, it would be possible as well to compute an asymptotic p-value by also computing/estimating the corresponding variance.

Variance component/SKAT approaches

As an alternative to burden/collapsing association tests, SKAT/variance-component based region tests have been developed for rare variant data [9]. They have the advantage that they do not require any assumptions about the effect configuration at the rare variant loci under the alternative hypothesis, but they are not as powerful as burden/collapsing approaches if one is certain about the alternative hypothesis. We define the general statistic

$$FBAT_{vc} = U^T W U$$

where $U = \sum_i U_i$ and W is a fixed $p \times p$ weight matrix. While, in the scenario of affected offspring trios and a diagonal weight matrix W, this test statistic equals the FB-SKAT statistic [11], Ionita-Laza et al. assess significance based on asymptotic results that require the empirical estimation of the variance/covariance matrix for the rare variants, which, given the sparseness

of rare variant data, can become problematic. In our framework, the p-value of the test statistic is obtained based on simulations from the conditional genotype distribution.

If we set $W = V^{-1}$, where $V = \sum_i V_i$, we obtain the multivariate FBAT [24]. The multivariate FBAT was designed for common variants, and the asymptotic p-values also require an empirical estimate of the correlation matrix. Again, for rare variants, this can lead to unreliable results, making the implementation of the multivariate FBAT in our proposed framework preferable.

Higher criticism and maximum statistic

Besides the commonly used burden and variance component approaches, we introduce the higher criticism and maximum statistic for region-based analysis in family-based studies. Both approaches are designed to identify sparse alternatives and have been introduced to genetic association studies of unrelated individuals recently [18,20].

Define the normalized residuals $\frac{U_j}{\sqrt{V_{jj}}}, j = 1, \dots, p$ and denote the corresponding association p-value based on the asymptotic marginal normal distribution by q_j. Based on the available amount of information per variant, e.g., the number of informative transmissions/families, we restrict the set of variants to a subset of variants where the marginal variance is large enough (e.g., we require at least 5 informative nuclear families). Denote the number of variants in this subset by p'. Given the ordered p-values $q_{(1)} \leq q_{(2)} \leq \cdots \leq q_{(p')}$, we define the HC statistic as

$$FBAT_{HC} = \max_j \frac{\frac{j}{p'} - q_{(j)}}{\sqrt{q_{(j)}(1 - q_{(j)})}}$$

Here, the index set J can be $\{1, \dots, \frac{p'}{2}\}$ or $\{1, \dots, p'\}$, depending on assumptions about the underlying genetic architecture.

It is important to note that, while $FBAT_{HC}$ contains a transformation based on the single variant asymptotic distribution, the assessment of its significance based on simulations from the conditional distribution remains a valid approach regardless of whether the assumptions that motivated the transformation hold.

The second approach to detect spare signals in the tested genomic region is the MAX statistic which is simply defined as

$$FBAT_{MAX} = \max_{1 \le j \le p} |\frac{U_j}{\sqrt{V_{jj}}}|$$

The theory related to the higher criticism/max statistic in the setting of unrelated case-control data and sparse signals developed in Mukherjee et al. [20] can be transferred to family-based studies/FBATS. In Appendix A, we derive how the theory in Mukherjee et al. [20] can be applied to the FBAT framework in the scenario of affected offspring trios. The corresponding optimality results for sparse signal scenarios motivate the application of these test statistics to rare variants in sequencing studies.

Finally, we note that it is, of course, possible to set up an omnibus statistic that is based on the maximum of multiple test statistics described above.

Results

In this section, we describe the results of two simulation studies and an analysis of a whole-genome sequencing study for childhood-asthma with 897 affected offspring trios.

Simulation studies

We studied the performance of our proposed test statistics in two extensive simulation studies. In both studies, we compared the Type I error and power with the existing methodology for family-based region association analysis. We restricted all simulations to the scenario of trios with an affected offspring. However, it is important to note that our framework can be applied to any nuclear family and phenotype distribution. For the test statistics $FBAT_{burden}$ and $FBAT_{vc}$, we applied uniform weights. In the following, we will denote the test statistics $FBAT_{burden}$, $FBAT_{vc}$, $FBAT_{HC}$, and $FBAT_{MAX}$ by Burden, SKAT, HC, and MAX.

Genetic regions with unphased data

We extracted haplotypes for the CEU and the GBR subpopulations from the 1000 Genomes Project [25], consisting of 30 and 50 consecutive rare variants with a minor allele frequency (MAF) below 3%. Based on these haplotypes, we generated genotype data for trios using Mendelian transmissions. Using a standard logistic disease model with a disease prevalence of $\approx 10\%$, we simulated offspring affection status and collected $n = 1{,}000$ affected offspring trios. This simulation study is similar to the simulation studies described in the existing literature [15,26].

We compared our test statistics with the GTDT [26] and the RV-GDT [15]. The GTDT [26] offers five different test statistics for region-based affected offspring trio analysis, designed for different modes of inheritance. The test statistics require phased haplotype data. If the phase information is not available, this information is reconstructed up to small uncertainties. We considered the test statistics GTDT-AD, GTDT-DOM, and GTDT-CH in our study. The RV-GDT [15] describes a generalization of the single variant GDT [13] for multiple variants in a genetic region. The RV-GDT can be applied to arbitrary pedigrees where affected, and unaffected samples are

collected; members can be missing. The test statistic compares the genotype counts between affected and unaffected members and corrects for the relatedness using the estimated/reported kinship coefficients. We note that this implies that the phenotype information for parents must be available, whereas the classical TDT/FBAT test for offspring trios does not require this information. For comparison, we included the test MAX-BF that tests if at least one single variant FBAT statistic reached the Bonferroni-corrected significance level corresponding to the number of variants p in the region. The corresponding single variant p-values were evaluated using asymptotic theory due to computational reasons.

To check the Type I error rates and the robustness against population stratification; we considered a null hypothesis simulation where no genetic variant is associated with the affection status and three different population admixture scenarios (Table 1). In these admixture scenarios, we generated one fixed parent based on the CEU haplotypes and the other parent based on the GBR haplotypes. The affection status of the parents differed across the three admixture scenarios. For the power analysis, we simulated six different scenarios where the number of causal variants and corresponding effect sizes differ (Table 2). In scenario 5, we picked very rare and independent causal variants with a MAF below 1%, and in scenario 6 we chose causal variants that are in strong LD with multiple other variants. All results are based on 1,000 replicates.

In Table 1, we observe that all methods control the Type I error appropriately. The only exception is the RV-GDT in the scenario of population admixture with discordant parental phenotypes (adm2 and adm3, Table 1). This is expected, as the GDT/RV-GDT test compares the frequencies between affected and unaffected family members and cannot distinguish between association and admixture in the parents. We also note that the RV-GDT computes a one-sided p-value, which explains the deflation/inflation behavior, depending on the

parental phenotypes. The power results in Table 2 demonstrate the advantages of non-burden tests in specific scenarios. The power results are also illustrated in Figures 1 and 2. The SKAT statistic shows the highest power in the first three scenarios and outperforms the other tests. However, the MAX and HC statistics also show substantial power. The results for scenario 4 are comparable between SKAT, MAX, and HC. In scenario 5, the HC statistics achieves the highest power, which is supported by our theoretical considerations as well (see Appendix A). In the last scenario 6, all tests achieve substantial power, as expected, due to the LD structure that pushes power. The most powerful tests here are SKAT and RV-GDT, but MAX and HC test statistics achieve similar results. If we have different effect directions (scenario 2 and 4), the burden test loses power compared to the consistent effect direction scenarios 1 and 3, which is expected. The FBAT Burden test and GDT-AD have almost no power in scenarios 4 and 5. It is important to note that the GTDT-AD and the FBAT burden test are essentially based on the same test statistic idea, the only difference lies in the fact that the GTDT assigns haplotypes (with possible error) and our approach uses the robust conditional genotype distribution computed by the FBAT haplotype algorithm.

There is a substantial difference between the power of the MAX statistic and the MAX-BF statistic in all scenarios, because the Bonferroni correction is conservative, and the p-value of the MAX statistic is evaluated based on the joint conditional genotype distribution. We note that this difference could be even larger if the genome-wide significance levels for region- and single variant based testing are considered.

Dense genetic regions with phased data

For our second set of simulation studies, we utilized the 1006 EUR population haplotypes from the 1000 Genomes Project along 1,000 consecutive rare genetic variants with MAF below 3%. In this simulation study, we consider a large number of variants in combination with a sparse signal, which means a small subset of causal variants that are not in strong LD with any other variants. We simulated affected offspring trios as described in the first simulation study but also stored the phased haplotypes for all members of the trio. In the scenario where the haplotypes are observed, the conditional distribution identified by the FBAT haplotype algorithm equals the distribution where both parents transmit one of the observed haplotypes with equal probability of 0.5. We compared the performance of the FBAT, the GTDT, and the RV-TDT BRV [27] statistics to demonstrate the potential advantage of non-burden tests in the presence of sparse signals. Again, we also included the MAX-BF test, where we considered the Bonferroni corrected significance level based on $p = 1{,}000$ tests. As mentioned above, the knowledge about the phased haplotypes is the preferred setting for the GTDT. Also, the RV-TDT requires phased haplotypes [27]. We considered a null hypothesis scenario and four different power scenarios (Table 3 and Figure 3). For the null hypothesis simulation and the first two power scenarios we simulated 1,000 trios; the last two power scenarios are based on 10,000 trios. The four power scenarios include causal variants that are in almost no LD with other variants, and the number of causal variants is small compared to the overall number of $p = 1{,}000$ variants. All results are based on 1,000 replicates, and the p-values for all test statistics were evaluated empirically based on the same 1,000 draws from the conditional haplotype distribution. The results for this simulation are also visualized in Figure 3.

In Table 3, we observe that all test statistics control the Type I error appropriately. Since all p-values are evaluated empirically based on the conditional haplotype distribution by

simulation, this is expected. In the first power scenario 1, the MAX test statistic achieves the highest power as we simulated a sparse and rare, but strong signal in the genetic region, consisting of two rare variants. The HC test statistic also achieves substantial power, whereas all other tests (except MAX-BF) are almost powerless in this scenario. In the second scenario, where the MAF of the two causal variants is much higher, the MAX test statistics still outperforms the other tests, but also the SKAT and the HC test statistics show good performances. In scenario 3, where many very rare causal variants have a relatively small effect size, the HC is the most powerful test. This is in line with the results in Mukherjee et al. [20] that describe a lower detection boundary in the mild sparse regime compared to the MAX test statistics. However, the RV-TDT BRV and MAX test statistic also achieve substantial power in this scenario. The power behavior differs more in the last scenario 4, where the effects are pointing in different directions. Here, the MAX and HC statistics have a significantly increased power compared to the other tests, while the HC test statistic is the most powerful one. We note that the FBAT burden and the GTDT-AD test are very close to the nominal level in scenarios 1, 2, and 4. Both tests are equivalent since the test statistics are the same.

Overall, again, the MAX test shows higher power than the MAX-BF tests, as described in the context of the first simulation study.

Real data analysis

To demonstrate the applicability and the advantages of our proposed framework, we analyzed a whole-genome sequencing dataset consisting of 897 complete asthmatic trios from Costa Rica [28]. After standard quality control, including Mendelian error rates, we excluded all variants with a MAF above 5%. The resulting 27,345,734 non-monomorphic variants were partitioned into approximately 547,000 consecutive windows of 50 rare variants. Other partitioning approaches could be considered here [29,30], but, as the focus of this data analysis

was to demonstrate the feasibility of our approach, we did not explore different window-strategies here.

For each window, we performed the Burden, SKAT, MAX, and HC test, using affection status as the phenotype. We evaluated the p-values by simulation, where we used an adaptive heuristic that increases the number of simulations if the estimated p-value is close to the minimum possible value. The smallest possible p-value was $p = 10^{-9}$, since the number of simulations was truncated at 10^9. In Figure 4, we plotted the corresponding quantile-quantile-plot. The plot indicates that the test statistics control the Type I error rate but can also identify potential findings.

Based on the approximately $4 * 547{,}000$ tests and a False Discovery Rate at $\alpha = 0.05$ [31], our approach identified three single significant regions on Chromosome 1, 12, and 21, as well as multiple consecutive significant regions on Chromosome 10. The significance of the three regions on Chromosomes 1, 12, and 21 was declared by the Burden test, whereas the single variant FBAT p-values within the regions were not in the range of genome-wide significance. The other regions on Chromosome 10 were identified by the MAX, HC, and SKAT tests. The lowest p-value of 10^{-9} was reached by the SKAT test. For all these regions, the Burden test did not reach the magnitude of genome-wide significance. This shows the benefits of combining different test statistics to identify distinct genetic signal structures.

Discussion

In this manuscript, we propose a general framework for the region-based association analysis of sequencing datasets with family-based designs. The framework incorporates burden tests, SKAT, maximum and higher criticism approaches, and, given the flexibility of the framework,

any future approach can straightforwardly be implemented. In contrast to previously published approaches, the joint genotype distribution along the loci is not obtained by empirical estimates, but via the haplotype algorithm for FBATs [16,17]. This allows our proposed testing framework for FBATs to assess the significance of an arbitrary test statistic based on simulations or based on the exact distribution. This approach is enabled by the recent improvements in the FBAT haplotype algorithm [16], which reduces the computational burden of the original approach by several magnitudes. Our simulation results illustrate that the optimal test for the region-based analysis depends on the specific genetic architecture of the disease, and any WGS analysis relying on just one single test statistic may not detect all associations contained in the data. While dense signals with consistent effect directions can be captured by burden tests, different effect directions and less dense signals can be identified by SKAT approaches. If the signal becomes more separated and sparser, the MAX and HC approaches can be the most powerful tests.

The proposed implementation of the simulation-based p-values requires the user to pre-select the number of simulations that FBAT performs for each test. The computational burden can be decreased by adaptive strategies [23]. The applications of the proposed analysis framework to simulated and real data illustrate that the theoretically expected advantages are also of practical relevance and that simulation-based p-values are not prohibitive in WGS settings. A subject of future research will be to integrate the existing FBAT approaches to multivariate phenotypes, longitudinal data, age at onset [32,32–34], gene-environmental interactions, and testing strategies into the proposed framework [3,6,7].

Appendix A: Detection of sparse signals

We consider the scenario of an affected offspring trio. Both parental genotypes are observed along with the p variants in the analysis region. As noted in Chen et al. [26], if there is no variant where all three observed genotypes (mother, father, offspring) are heterozygous, the phase information can be recaptured from the observed unphased genotype data. However, as described in Hecker et al. [14], treating inferred haplotypes as observed haplotypes can lead to misspecification.

Nevertheless, more specifically, if there is no variant for which both parental genotypes are heterozygous, haplotypes can be phased, and the resulting conditional genotype distribution obtained by the FBAT haplotype algorithm equals the conditional distribution where we treat the haplotypes as observed. If we restrict the genetic data to rare variants, this is true for most nuclear families. In addition, with relatively high probability, at least 1 parent has only 1 minor allele in the genetic region.

Let us denote the phased parental mating type for such a trio by $G = (h_1^M, h_2^M)\ x\ (h_1^F, h_2^F)$. The possible offspring genotypes are denoted by $X_1 = (h_1^M + h_1^F)$, $X_2 = (h_1^M + h_2^F), X_3 = (h_2^M + h_1^F)$ and $X_4 = (h_2^M + h_2^F)$. We assume the following, commonly used, disease model that describes the conditional offspring genotype distribution

$$P(X_i|T = 1, G) = \frac{\exp(\beta^T X_i)}{\sum_{j=1}^{4} \exp(\beta^T X_j)}$$

where the p dimensional vector β describes the genetic effects of the variants in the region.

If we denote the inherited offspring haplotypes by (h^M, h^F), this model factors into the product of the two likelihoods

$$P\left(h^g = h_j^g | T = 1, G\right) = \frac{\exp\left(\beta^T h_j^g\right)}{\exp\left(\beta^T h_1^g\right) + \exp\left(\beta^T h_2^g\right)}, \quad j = 1,2, \qquad g = M, F$$

Since the haplotype data is sparse as described above, this setting matches the scenario of Weakly Correlated Designs that is described in the paper by Mukherjee et al. [20] about sparse binary regression (Definition 4.1). They showed that in the sparse regime, the higher criticism and the maximum statistic can identify sparse alternatives efficiently (see Theorem 7.4).

Although this motivates the application to affected offspring trios, the statistics can be applied in all FBAT scenarios. The Type I error is preserved because we utilize a simulation-based approach.

Acknowledgments

This work was supported by Cure Alzheimer's Fund; the National Human Genome Research Institute [R01HG008976]; and the National Heart, Lung, and Blood Institute [U01HL089856, U01HL089897, P01HL120839, P01HL132825].

Declaration of Interests

The authors declare no competing interests.

Web Resources

The FBAT software is available at https://sites.google.com/view/fbat-web-page. A new version that implements the described methodology is currently in preparation and will be available soon.

References

1. Spielman, R.S., McGinnis, R.E., and Ewens, W.J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am. J. Hum. Genet. *52*, 506–516.

2. Laird, N.M., and Lange, C. (2006). Family-based designs in the age of large-scale gene-association studies. Nat. Rev. Genet. *7*, 385–394.

3. Ionita-Laza, I., McQueen, M.B., Laird, N.M., and Lange, C. (2007). Genomewide weighted hypothesis testing in family-based association studies, with an application to a 100K scan. Am. J. Hum. Genet. *81*, 607–614.

4. Lange, C., Lyon, H., DeMeo, D., Raby, B., Silverman, E.K., and Weiss, S.T. (2003). A New Powerful Non-Parametric Two-Stage Approach for Testing Multiple Phenotypes in Family-Based Association Studies. Hum. Hered. *56*, 10–17.

5. Murphy, A., Weiss, S.T., and Lange, C. (2008). Screening and Replication using the Same Data Set: Testing Strategies for Family-Based Studies in which All Probands Are Affected. PLOS Genet. *4*, e1000197.

6. Steen, K.V., McQueen, M.B., Herbert, A., Raby, B., Lyon, H., DeMeo, D.L., Murphy, A., Su, J., Datta, S., Rosenow, C., et al. (2005). Genomic screening and replication using the same data set in family-based association testing. Nat. Genet. *37*, 683–691.

7. Won, S., Wilk, J.B., Mathias, R.A., O'Donnell, C.J., Silverman, E.K., Barnes, K., O'Connor, G.T., Weiss, S.T., and Lange, C. (2009). On the Analysis of Genome-Wide Association Studies in Family-Based Designs: A Universal, Robust Analysis Approach and an Application to Four Genome-Wide Association Studies. PLOS Genet. *5*, e1000741.

8. Li, B., and Leal, S.M. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am. J. Hum. Genet. *83*, 311–321.

9. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. Am. J. Hum. Genet. *89*, 82–93.

10. De, G., Yip, W.-K., Ionita-Laza, I., and Laird, N. (2013). Rare Variant Analysis for Family-Based Design. PLOS ONE *8*, e48495.

11. Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J.D., and Lin, X. (2013). Family-based association tests for sequence data, and comparisons with population-based association tests. Eur. J. Hum. Genet. EJHG *21*, 1158–1162.

12. Zhou, W., Nielsen, J.B., Fritsche, L.G., LeFaive, J., Taliun, S.A.G., Bi, W., Gabrielsen, M.E., Daly, M.J., Neale, B.M., Hveem, K., et al. (2019). Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts. BioRxiv 583278.

13. Chen, W.-M., Manichaikul, A., and Rich, S.S. (2009). A Generalized Family-Based Association Test for Dichotomous Traits. Am. J. Hum. Genet. *85*, 364–376.

14. Hecker, J., Laird, N., and Lange, C. (2019). A comparison of popular TDT-generalizations for family-based association analysis. Genet. Epidemiol. *43*, 300–317.

15. He, Z., Zhang, D., Renton, A.E., Li, B., Zhao, L., Wang, G.T., Goate, A.M., Mayeux, R., and Leal, S.M. (2017). The Rare-Variant Generalized Disequilibrium Test for Association Analysis of Nuclear and Extended Pedigrees with Application to Alzheimer Disease WGS Data. Am. J. Hum. Genet. *100*, 193–204.

16. Hecker, J., Xu, X., Townes, F.W., Fier, H.L., Corcoran, C., Laird, N., and Lange, C. (2017). Family-based tests for associating haplotypes with general phenotype data. Genet. Epidemiol. *42*, 123–126.

17. Horvath, S., Xu, X., Lake, S.L., Silverman, E.K., Weiss, S.T., and Laird, N.M. (2004). Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. Genet. Epidemiol. *26*, 61–69.

18. Barnett, I., Mukherjee, R., and Lin, X. (2017). The Generalized Higher Criticism for Testing SNP-Set Effects in Genetic Association Studies. J. Am. Stat. Assoc. *112*, 64–76.

19. Donoho, D., and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. Ann. Stat. *32*, 962–994.

20. Mukherjee, R., Pillai, N.S., and Lin, X. (2015). HYPOTHESIS TESTING FOR HIGH-DIMENSIONAL SPARSE BINARY REGRESSION. Ann. Stat. *43*, 352–381.

21. Schneiter, K., Degnan, J.H., Corcoran, C., Xu, X., and Laird, N. (2007). EFBAT: exact family-based association tests. BMC Genet. *8*, 86.

22. Rabinowitz, D., and Laird, N. (2000). A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. Hum. Hered. *50*, 211–223.

23. Hecker, J., Ruczinski, I., Cho, M., Silverman, E., Coull, B., and Lange, C. (2017). A flexible and nearly optimal sequential testing approach to randomized testing: QUICK-STOP. Genet. Epidemiol. (accepted).

24. Rakovski, C.S., Xu, X., Lazarus, R., Blacker, D., and Laird, N.M. (2007). A new multimarker test for family-based association studies. Genet. Epidemiol. *31*, 9–17.

25. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. Nature *526*, 68–74.

26. Chen, R., Wei, Q., Zhan, X., Zhong, X., Sutcliffe, J.S., Cox, N.J., Cook, E.H., Li, C., Chen, W., and Li, B. (2015). A haplotype-based framework for group-wise transmission/disequilibrium tests for rare variant association analysis. Bioinforma. Oxf. Engl. *31*, 1452–1459.

27. He, Z., O'Roak, B.J., Smith, J.D., Wang, G., Hooker, S., Santos-Cortez, R.L.P., Li, B., Kan, M., Krumm, N., Nickerson, D.A., et al. (2014). Rare-variant extensions of the transmission

disequilibrium test: application to autism exome sequence data. Am. J. Hum. Genet. *94*, 33–46.

28. (2016). NHLBI TOPMed: The Genetic Epidemiology of Asthma in Costa Rica.

29. Fier, H.L., Prokopenko, D., Hecker, J., Cho, M.H., Silverman, E.K., Weiss, S.T., Tanzi, R.E., and Lange, C. (2017). On the association analysis of genome-sequencing data: A spatial clustering approach for partitioning the entire genome into nonoverlapping windows. Genet. Epidemiol. *41*, 332–340.

30. He, Z., Xu, B., Buxbaum, J., and Ionita-Laza, I. (2019). A genome-wide scan statistic framework for whole-genome sequence data analysis. Nat. Commun. *10*, 1–11.

31. Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Ser. B Methodol. *57*, 289–300.

32. Ding, X., Lange, C., Xu, X., and Laird, N. (2009). New powerful approaches for family-based association tests with longitudinal measurements. Ann. Hum. Genet. *73*, 74–83.

33. Lange, C., Blacker, D., and Laird, N.M. (2004). Family-based association tests for survival and times-to-onset analysis. Stat. Med. *23*, 179–189.

34. Lange, C., Silverman, E.K., Xu, X., Weiss, S.T., and Laird, N.M. (2003). A multivariate family-based association test using generalized estimating equations: FBAT-GEE. Biostatistics *4*, 195–206.

Figures

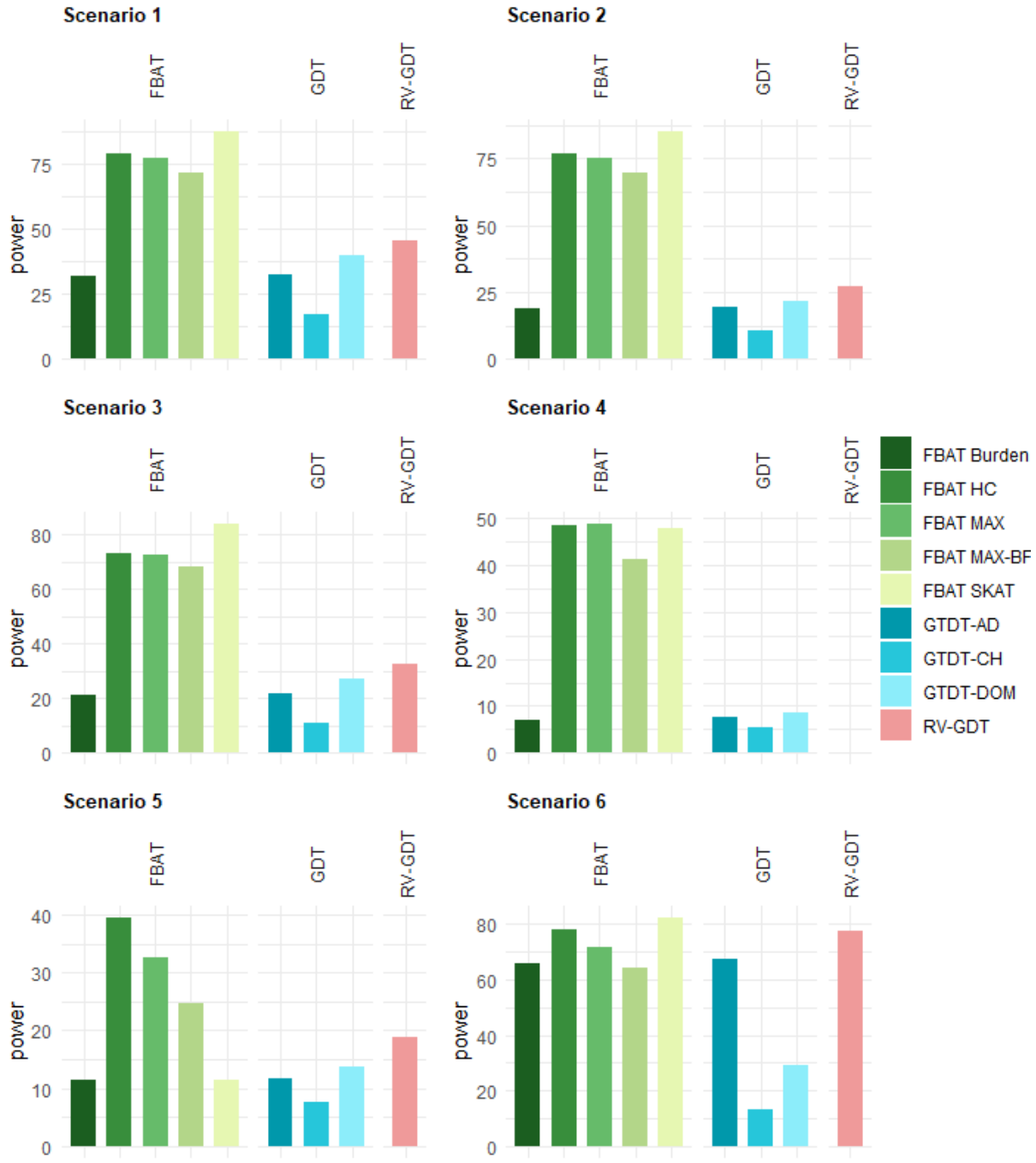

Figure 1. Power results in six different scenarios for genetic regions consisting of 30 variants at a significance level of $\alpha = 0.05$. All results based on 1,000 replicates.

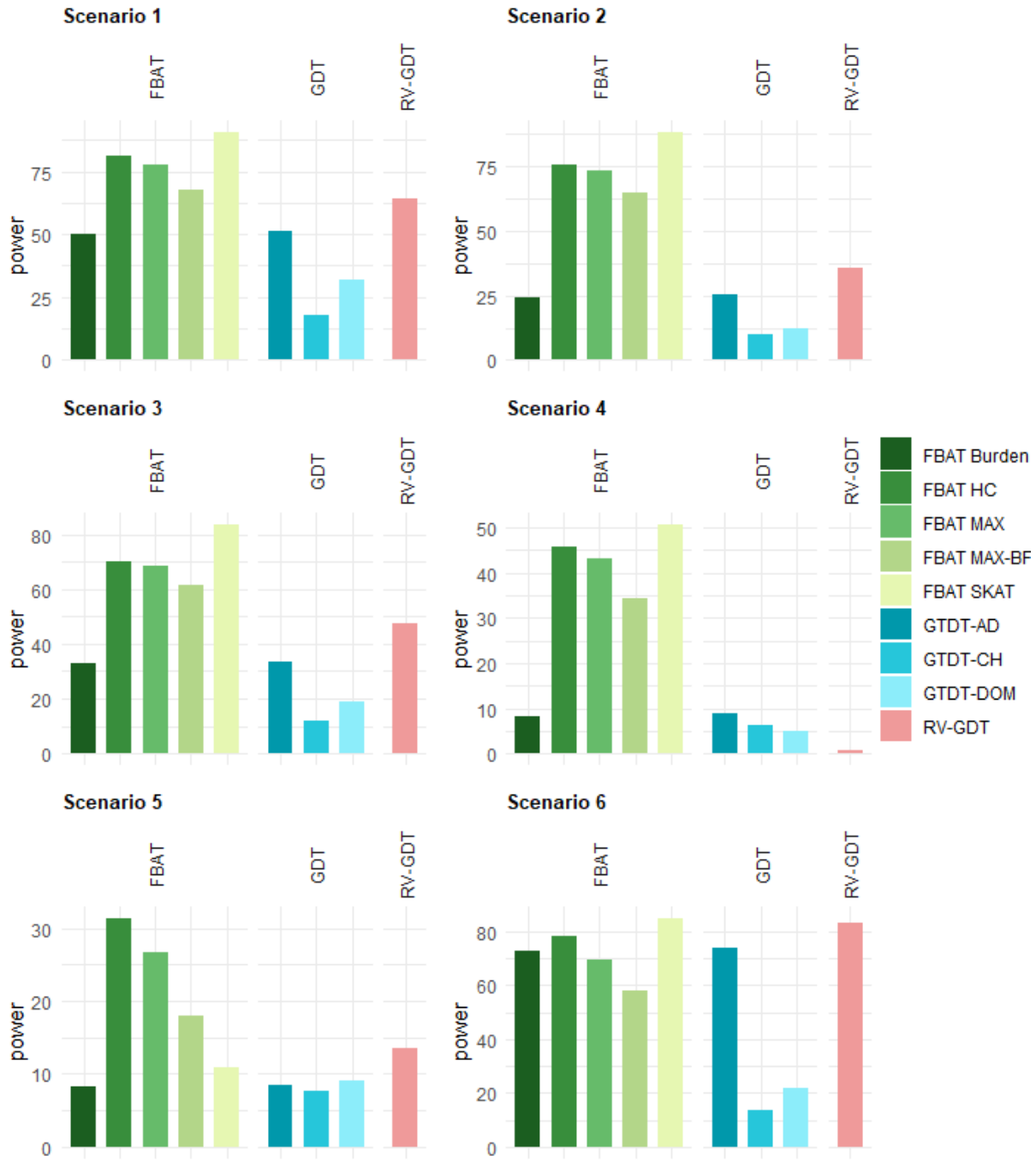

Figure 2. Power results in six different scenarios for genetic regions consisting of 50 variants at a significance level of $\alpha = 0.05$. All results based on 1,000 replicates.

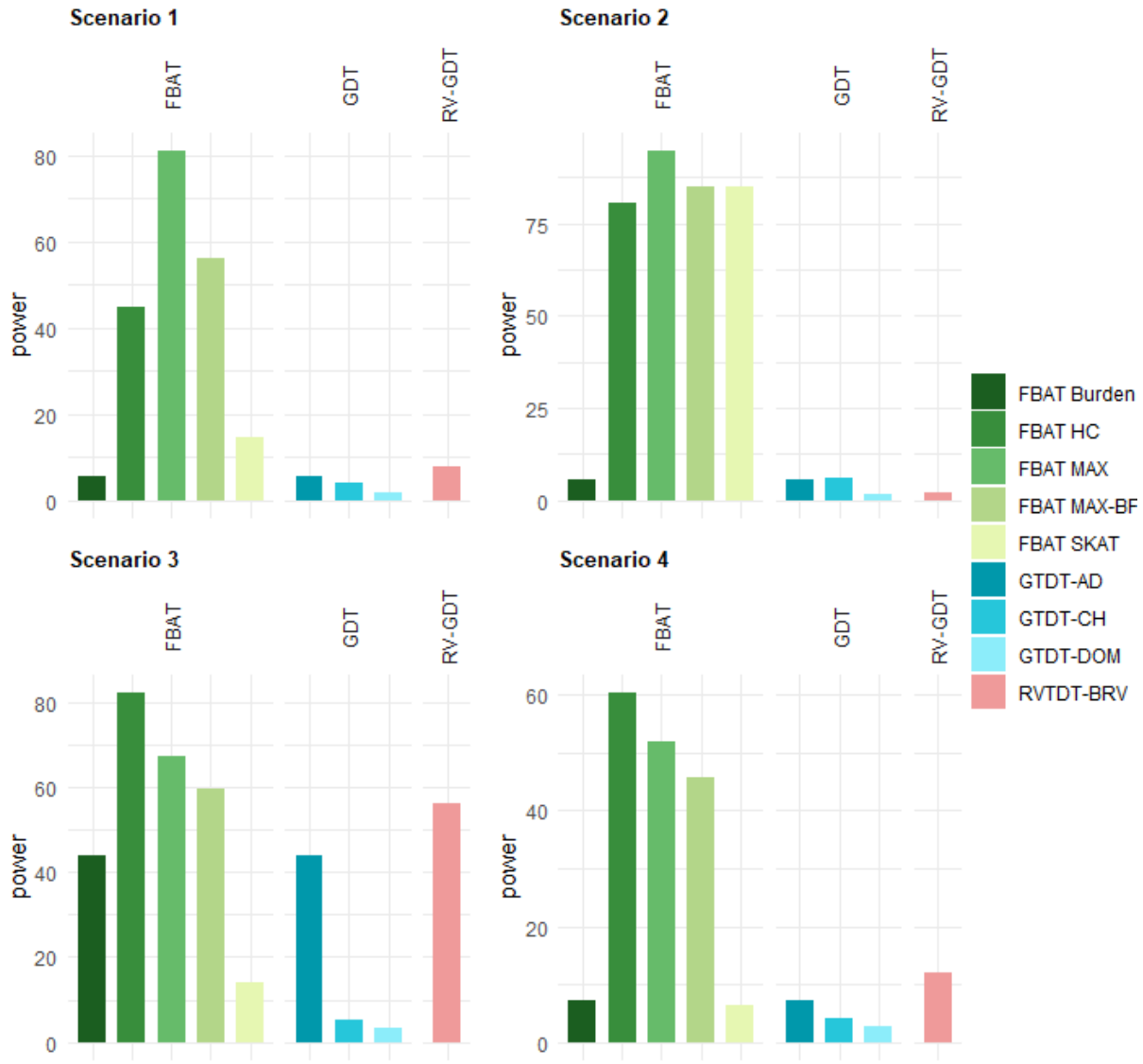

Figure 3. Power results for four different scenarios for genetic regions consisting of 1,000 variants at a significance level of $\alpha = 0.05$. All results based on 1,000 replicates.

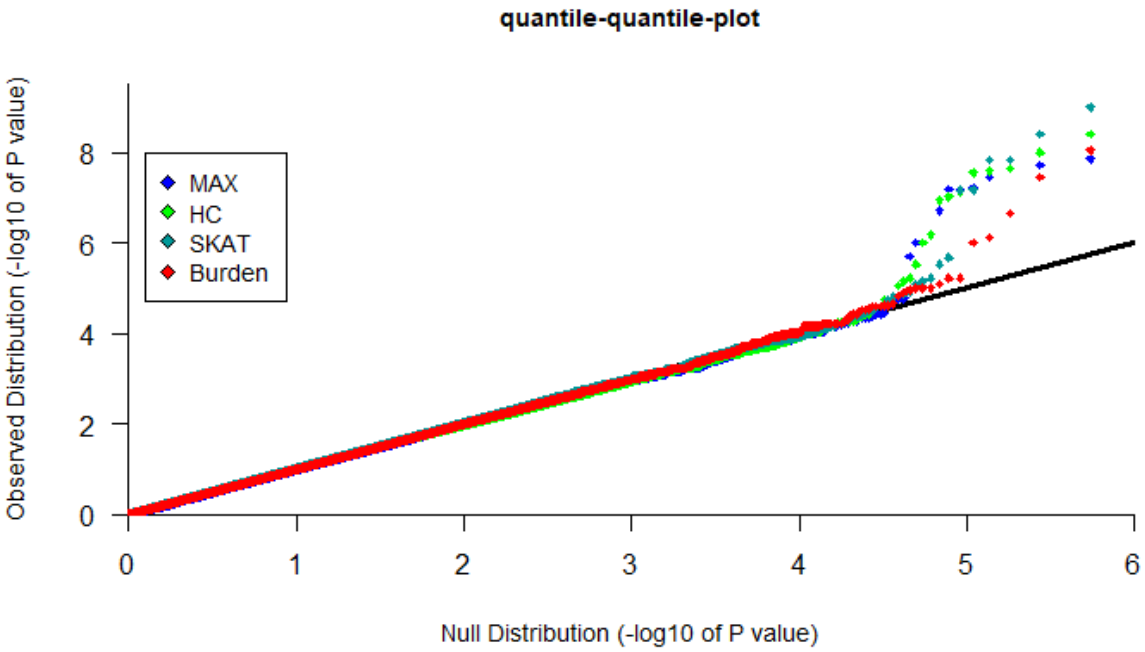

Figure 4. Real data analysis of 897 asthmatic offspring trios in a WGS study. Quantile-quantile plot for Burden, SKAT, MAX, and HC test statistics based on approximately 547,000 windows of 50 consecutive rare variants.

Tables

		FBAT					GTDT			RV-GDT
		Burden	SKAT	MAX	HC	MAX-BF	GTDT-AD	GTDT-DOM	GTDT-CH	RV-GDT
$p = 30$										
	null	4.7%	5.4%	5.7%	5.0%	2.9%	5.2%	5.6%	5.2%	4.4%
	adm1	5.7%	5.5%	4.4%	5.3%	3.0%	6.2%	5.3%	5.4%	5.7%
	adm2	4.8%	4.7%	3.9%	4.1%	2.3%	5.1%	6.2%	5.0%	0.0%
	adm3	4.2%	4.6%	3.2%	4.0%	2.0%	4.5%	5.1%	4.9%	99.6%
$p = 50$										
	null	3.1%	3.6%	3.7%	4.0%	2.0%	3.4%	3.5%	3.8%	4.4%
	adm1	4.3%	5.2%	4.0%	3.5%	1.2%	4.7%	4.8%	5.2%	4.0%
	adm2	4.1%	4.4%	4.8%	4.7%	2.4%	4.0%	5.3%	4.9%	0.0%
	adm3	4.5%	5.4%	6.4%	5.1%	2.7%	4.7%	4.5%	5.2%	35.4%

Table 1. Type I errors at a significance level of 5% for the FBAT, GTDT and RV-GDT statistics. We considered four scenarios, separately for $p = 30$ and $p = 50$ variants. All results based on 1,000 replicates.
null: no association between genetic variants and phenotype, no population admixture
adm1: 1 parent generated from CEU haplotypes, 1 parent generated from GBR haplotypes. Parents unaffected, offspring affected.
adm2: 1 parent generated from CEU haplotypes, 1 parent generated from GBR haplotypes. CEU parent affected, GBR parent unaffected, offspring affected.
adm3: 1 parent generated from CEU haplotypes, 1 parent generated from GBR haplotypes. CEU parent unaffected, GBR parent affected, offspring affected.

		FBAT					GTDT			RV-GDT
		Burden	SKAT	MAX	HC	MAX-BF	GTDT-AD	GTDT-DOM	GTDT-CH	RV-GDT
$p = 30$	1	32.2%	87.9%	77.6%	79.2%	72.0%	32.7%	39.7%	17.2%	45.5%
	2	19.2%	85.4%	75.1%	76.9%	69.7%	19.7%	21.5%	10.7%	27.4%
	3	20.9%	83.7%	72.5%	73.1%	68.2%	21.9%	27.0%	11.0%	32.7%
	4	7.1%	48.0%	48.8%	48.7%	41.4%	7.7%	8.7%	5.3%	0.17%
	5	11.3%	11.5%	32.6%	39.5%	24.8%	11.6%	13.7%	7.5%	18.8%
	6	65.9%	82.3%	71.5%	77.8%	64.2%	67.5%	29.1%	13.4%	77.2%
$p = 50$	1	50.2%	90.8%	77.8%	80.9%	67.9%	51.3%	31.8%	17.9%	64.1%
	2	24.0%	88.3%	73.5%	75.3%	64.9%	25.2%	12.1%	9.9%	35.6%
	3	32.9%	83.4%	68.6%	70.0%	61.3%	33.4%	18.9%	11.7%	47.6%
	4	8.2%	50.6%	43.3%	45.8%	34.3%	8.7%	4.9%	6.4%	0.7%
	5	8.3%	11.0%	26.8%	31.4%	17.9%	8.5%	9.0%	7.6%	13.5%
	6	72.6%	84.9%	69.3%	78.5%	58.0%	73.7%	21.9%	13.5%	83.3%

Table 2. Power estimates at a significance level of 5% for the FBAT, GTDT and RV-GDT statistics. We considered six scenarios, separately for $p = 30$ and $p = 50$ variants. All results based on 1,000 replicates.
scenario 1: three causal variants, effect sizes $0.4|\log_{10}(MAF)|$, same direction
scenario 2: three causal variants, effect sizes $0.4|\log_{10}(MAF)|$, different direction
scenario 3: two causal variants, effect sizes $0.4|\log_{10}(MAF)|$, same direction
scenario 4: two causal variants, effect sizes $0.4|\log_{10}(MAF)|$, different direction
scenario 5: four very rare causal variants, effect size 1.0, same direction
scenario 6: three causal variants, effect sizes $0.4|\log_{10}(MAF)|$, same direction, in strong LD with other variants.

		FBAT					GTDT			RV-TDT
		Burden	SKAT	MAX	HC	MAX-BF	GTDT-AD	GTDT-DOM	GTDT-CH	RV-TDT BRV
$\alpha = 0.05$	null	5.2%	4.5%	3.9%	5.0%	0.7%	5.2%	2.5%	5.9%	5.5%
	1	5.6%	14.5%	81.2%	45.0%	56.0%	5.6%	1.7%	4.1%	8.0%
	2	5.6%	85.0%	94.9%	80.8%	85.2%	5.6%	1.5%	6%	2.1%
	3	43.8%	14.2%	67.4%	82.3%	59.7%	43.8%	3.3%	5.2%	56.2%
	4	7.1%	6.3%	51.9%	60.4%	45.8%	7.1%	2.8%	4.1%	12.0%
$\alpha = 0.01$	null	1.3%	1.5%	0.7%	1.2%	0.0%	1.3%	0.4%	1.1%	1.3%
	1	1.3%	3.4%	61.2%	39.9%	33.0%	1.3%	0.2%	0.8%	2.3%
	2	0.9%	54.6%	87.6%	77.7%	71.0%	0.9%	0.2%	1.3%	0.5%
	3	23.0%	2.7%	39.2%	48.3%	32.0%	23.0%	0.5%	1.3%	32.2%
	4	1.8%	0.9%	27.3%	31.9%	20.8%	1.8%	0.5%	0.8%	3.1%

Table 3. Type I error and power estimates at significance levels of 1% and 5%, all results based on 1,000 replicates. MAX-BF refers to the test that at least one single variant test statistic reached the Bonferroni-corrected significance level based on p =1,000 tests.
scenario 1: two causal variants, MAF~0.2%, almost no LD with other variants, effect size 1.8
scenario 2: two causal variants, MAF~2%, almost no LD with other variants, effect size 0.7
scenario 3: 16 causal variants, MAF~0.1%, almost no LD with other variants, effect size 0.7
scenario 4: 16 causal variants, MAF~0.1%, almost no LD with other variants, effect size 0.7, effect direction alternates