A selective inference approach for FDR control using multi-omics covariates yields insights into disease risk

Ronald Yurko[a], Max G'Sell[a], Kathryn Roeder[a,b], and Bernie Devlin[c]

[a]Department of Statistics & Data Science, Carnegie Mellon University
[b]Department of Computational Biology, Carnegie Mellon University
[c]Department of Psychiatry, University of Pittsburgh School of Medicine

Abstract

To correct for a large number of hypothesis tests, most researchers rely on simple multiple testing corrections. Yet, new methodologies of selective inference could potentially improve power while retaining statistical guarantees, especially those that enable exploration of test statistics using auxiliary information (covariates) to weight hypothesis tests for association. We explore one such method, adaptive p-value thresholding (Lei & Fithian 2018, AdaPT), in the framework of genome-wide association studies (GWAS) and gene expression/coexpression studies, with particular emphasis on schizophrenia (SCZ). Selected SCZ GWAS association p-values play the role of the primary data for AdaPT; SNPs are selected because they are gene expression quantitative trait loci (eQTLs). This natural pairing of SNPs and genes allow us to map the following covariate values to these pairs: GWAS statistics from genetically-correlated bipolar disorder, the effect size of SNP genotypes on gene expression, and gene-gene coexpression, captured by subnetwork (module) membership. In all 24 covariates per SNP/gene pair were included in the AdaPT analysis using flexible gradient boosted trees. We demonstrate a substantial increase in power to detect SCZ associations using gene expression information from the developing human prefontal cortex (Werling et al. 2019). We interpret these results in light of recent theories about the polygenic nature of SCZ. Importantly, our entire process for identifying enrichment and creating features with independent complementary data sources can be implemented in many different high-throughput settings to ultimately improve power.

***Keywords**:* Multiple hypothesis testing, False Discovery Rate, GWAS, eQTL, neuropsychiatric disorders

Large scale experiments, such as scanning the human genome for variation affecting a phenotype, typically result in a plethora of hypothesis tests. To overcome the multiple testing challenge, one needs corrections to limit false positives while maximizing power. Introduced by Benjamini & Hochberg (1995), false discovery rate (FDR) control has become a popular approach to improve power for detecting weak effects by limiting the expected false discovery proportion (FDP) instead of the more classical Family-Wise Error Rate. The Benjamini-Hochberg (BH) procedure was the first method to control FDR at target level α using a step-up procedure that is *adaptive* to the set of p-values for the hypotheses of interest (Benjamini & Hochberg 1995). Other methods for FDR control have led to improvements in power over BH by incorporating prior information, such as by the use of p-value weights (Genovese et al. 2006). In the "omics" world – genomics, epigenomics, proteomics, and so on – the challenge of multiple testing is burgeoning, in part because our ability to characterize omics features grows continually and in part because of the realization that multiple omics are required for describing phenotypic variation. One might imagine merging complementary omics data and tests using a priori hypothesis weights to improve power; however, until recently, it was not clear how to choose these weights in a data driven manner.

Recent methodologies have been proposed to account for covariates or auxiliary information while maintaining FDR control (Scott et al. 2015, Ignatiadis et al. 2016, Boca & Leek 2018, Li & Barber 2019, Zhang et al. 2019). We implement a selective inference approach, called adaptive p-value thresholding (Lei & Fithian 2018, AdaPT), to explore prior auxiliary information while maintaining guaranteed finite-sample FDR control. A recent review compared the performance of AdaPT with other covariate-informed methods for FDR control with off-the-shelf one and two-dimensional covariate examples (Korthauer et al. 2019). One of the weaknesses they ascribe to AdaPT is the unintuitive modeling framework for incorporating covariates; however, AdaPT is not a specific algorithm that one can simply apply to a dataset, but rather a meta-algorithm for marrying machine learning methods to multiple testing problems without compromising FDR control. We fully embrace AdaPT's flexibility via gradient boosted trees in a much richer, high-dimensional setting. Our boosting implementation of AdaPT easily scales with more covariates, enabling practitioners to capture interactions and non-linear effects from the rich resources of prior information available.

In this manuscript, we demonstrate our gradient boosted trees implementation of AdaPT on

results from genome-wide association studies (GWAS), incorporating covariates constructed from independent GWAS and gene expression studies. Specifically, we apply AdaPT to GWAS for detecting single nucleotide polymorphisms (SNPs) associated with schizophrenia (SCZ) using bipolar disorder (BD) GWAS results from an independent dataset as a covariate. Additionally, we incorporate results from the recent BrainVar study to identify a set of expression-SNPs (eSNPs) based on 176 neurotypical brains, sampled from pre- and post-natal tissue from the human dorsolateral prefrontal cortex (Werling et al. 2019). Along with the genetically correlated BD z-statistics, we create additional features from this complementary data source by summarizing the associated developmental gene expression quantitative trait loci (eQTL) slopes and membership in gene co-expression networks. We demonstrate that this process of identifying an enriched set of eSNPs and applying AdaPT with covariates summarizing gene expression from the developing human prefrontal cortex yield substantial improvement in power with each additional piece of information from the BrainVar study. Furthermore, we validate the replication of our results using more recent, independent SCZ studies.

This study had two goals, to explore the use of AdaPT in a realistic high-dimensional multi-omics setting and to determine what can be learned about the neurobiology of SCZ by this exploration. Our results revealed the power of incorporating auxiliary information with flexible gradient boosted trees. While each covariate independently provided at best a modest increase in power, our adaptive search discovered a more complex model with far greater power. These discoveries also led to increasing support for the polygenic basis of SCZ, complementing recent findings and suggesting that there are many physiological avenues to its underlying neurobiology. We emphasize that the process and analysis undertaken with this implementation of AdaPT can be extended to a variety of "omics" and other settings to utilize the rich contextual information that is often ignored by standard multiple testing corrections. We highlight this feature by analyzing two other sets of GWAS studies, type 2 diabetes (T2D) and body mass index (BMI), using results from these analyses to interpret findings from SCZ.

Results

Methodology overview

AdaPT is an iterative search procedure, introduced by Lei & Fithian (2018), for determining a set of discoveries/rejections, $\mathcal{R}$, with guaranteed finite-sample FDR control at target level α under conditions outlined below. We apply AdaPT to the collection of p-values and auxiliary information, $(p_i, x_i)_{i \in n}$, testing hypothesis H_i regarding SNP i's association with the phenotype of interest (e.g. SCZ). The covariates from some feature space, $x_i \in \mathcal{X}$, capture information collected independently of p_i, but potentially related to whether or not the null hypothesis for H_i is true and the effect size under the alternative. AdaPT provides a flexible framework to incrementally *learn* these relationships, potentially increasing the power of the testing procedure, while maintaining valid FDR control.

For each step $t = 0, 1, \ldots$ in the AdaPT search, we first determine the rejection set $\mathcal{R}_t = \{i : p_i \leq s_t(x_i)\}$, where $s_t(x_i)$ is the rejection threshold at step t that is *adaptive* to the covariates x_i. This provides us with both the number of discoveries/rejections $R_t = |\mathcal{R}_t|$, as well as a *pseudo*-estimate for the number of false discoveries $A_t = |\{i : p_i \geq 1 - s_t(x_i)\}|$ (i.e. number of p-values above the "mirror estimator" of $s_t(x_i)$). These quantities are used to estimate the FDP at the current step t,

$$\widehat{\mathrm{FDP}}_t = \frac{1 + A_t}{\max\{R_t, 1\}}. \tag{1}$$

If $\widehat{\mathrm{FDP}}_t \leq \alpha$, then the AdaPT search ends and the set of discoveries $\mathcal{R}_t$ is returned. Otherwise, we proceed to update the rejection threshold while satisfying two protocols: (1) the updated threshold must be more stringent $s_{t+1}(x_i) \leq s_t(x_i)$, and (2) p-values determining R_t and A_t are *partially* masked,

$$\tilde{p}_{t,i} = \begin{cases} p_i, \text{if } s_t(x_i) < p_i < 1 - s_t(x_i), \\ \{p_i, 1 - p_i\}, \text{ otherwise.} \end{cases} \tag{2}$$

Under these protocols, the rejection threshold can be updated using R_t, A_t, and $(x_i, \tilde{p}_{t,i})_{i \in [n]}$. The flexibility in how this update takes place is one of AdaPT's key strengths and allows it to easily incorporate other approaches from the multiple testing literature, such as a conditional version of

the two-groups model (Efron et al. 2001) with estimates for the probability of being non-null, π_1, and the effect size under the alternative, μ.

The algorithm proceeds by sequentially updating the threshold $s_{t+1}(x_i)$ to discard the most likely null element in the current rejection region, as measured by the conditional local false discovery rate (fdr): i.e., $i^* = \arg\max_{i \in \mathcal{R}_t} \text{fdr}_{t,i}$ is removed from $\mathcal{R}_t$. With the threshold updated, the AdaPT search repeats by estimating FDP and updating the rejection threshold until the target FDR level is reached $\widehat{\text{FDP}}_t \leq \alpha$ or $\mathcal{R}_t = 0$.

This procedure guarantees finite-sample FDR control under independence of the null p-values and as long as the null distribution of p-values is *mirror conservative*, i.e. the large "mirror" counterparts $1 - p_i \geq 0.5$ are at least as likely as the small p-values $p_i \leq 0.5$. To address the assumption of independence, we select a subset of weakly correlated SNPs detailed in *Data*, and additionally provide simulations in *SI Appendix* showing that AdaPT appears to maintain FDR control in relevant positive dependence settings. However, one practical limitation we encounter with the FDP estimate in Equation 1 is observing p-values *exactly* equal to one. While this can understandably occur with publicly available GWAS summary statistics, p-values equal to one will *always* contribute to the estimated number of false discoveries A_t. This nuance can lead to a failure of obtaining discoveries at a desired target α, such as the reported AdaPT results by Korthauer et al. (2019) for multiple case-studies. However, we demonstrate in *SI Appendix* an adjustment to the p-values for T2D and BMI GWAS applications that alleviates this problem, although future work should explore modifications to the FDP estimator itself.

The modeling step of AdaPT estimates conditional local fdr with an EM algorithm. In this context, we use gradient boosted trees, which constructs a flexible predictive function as a weighted sum of many simple trees, fit using a gradient descent procedure that minimizes a specified objective function. The two objective functions considered correspond to estimating the probability of a test being non-null and the distribution of the effect size for non-null tests. The advantage of this approach to function fitting is that it is invariant to monotonic variable transformations, automatically incorporates important variable interactions, and is able to handle a large number of covariates without degrading significantly in performance due to the high dimensionality. In contrast, less effective methods might fail to capture useful information because the covariates are poorly transformed for a linear model, because the important information is only revealed through

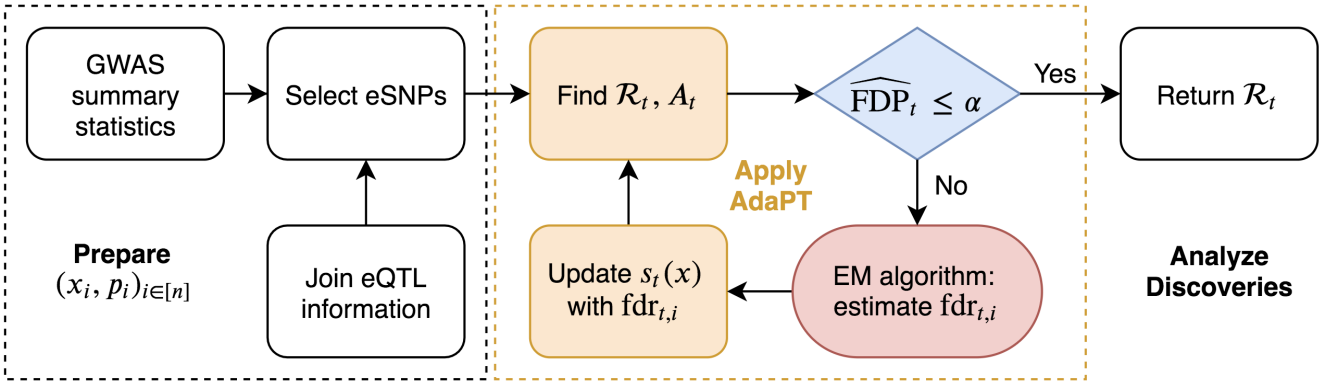

Figure 1: Summary of AdaPT GWAS implementation for selected set of SNPs. See Figure S1 for a summary of the AdaPT EM algorithm.

a combination of covariates, or because the important signal is simply swamped by the number of possible predictors to search through. Our choice of method gives the flexibility to include many potentially useful covariates without being overly concerned about the functional form with which they enter or their marginal utility. In our implementation, we employ the XGBoost library (Chen & Guestrin 2016) to capitalize on its computational advantages. Figure 1 displays the full pipeline of our implementation of AdaPT to GWAS summary statistics for SNPs using expression quantitative trait loci (eQTL) to select the SNPs under investigation.

Data

Our investigation includes AdaPT analyses of published GWAS p-values, $\{p_i, i = 1, \ldots n\}$, for body mass index (Locke et al. 2015, BMI), type 2 diabetes (Mahajan et al. 2018, T2D), and schizophrenia (Ruderfer et al. 2014, SCZ), but we focus our presentation on SCZ results. SCZ is a highly heritable, severe neuropsychiatric disorder. It is most strongly correlated, genetically, with another severe disorder, bipolar disorder (BD) (Lichtenstein et al. 2009, Cross-Disorder Group of the Psychiatric Genomics Consortium 2013). Because of this genetic correlation, reported z-statistics from BD GWAS, z_i^{BD}, can be used as informative covariates for determining the SCZ rejection threshold. As an application of our AdaPT implementation, we use the GWAS summary statistics reported by Ruderfer et al. (2014), specifically 19,779 subjects diagnosed with either SCZ or BD with 19,423 control subjects (data are available from the Psychiatric Genomics Consortium, PGC). SCZ and BD subjects were completely independent and independent controls were bulk

matched to the sample sizes of the two case samples. Results from more recent studies in Ruderfer et al. (2018) are used for replication analysis of our results (combined 53,555 SCZ and BD cases with 54,065 controls). However, the *2014-only* studies from Ruderfer et al. (2014) are a subset of the *all-2018* studies from Ruderfer et al. (2018). Although we do not have access to the raw genotype data, we use the fact that both papers report inverse variance-weighted fixed effects meta-analysis results (Willer et al. 2010). We then separate the summary statistics for the *2018-only* studies exclusive to Ruderfer et al. (2018), thus independent of the *2014-only* studies, and create an appropriate hold-out for replication analysis.

After matching alleles from both *2014-only* and *all 2018* studies and limiting SNPs to those with imputation score $INFO > 0.6$ for both BD and SCZ in *2014-only* (Ruderfer et al. 2014), we obtained 1,109,226 SNPs. Rather than test all SNPs, we chose to investigate a selected subset of SNPs, eSNPs, whose genotypes are correlated with gene expression; this additional filtering step captures a set of SNPs that are more likely to be functional and not highly correlated (Nicolae et al. 2010). These eSNPs were identified from two sources. First, we evaluated the BrainVar study of dorsolateral prefrontal cortex samples across a developmental span (Werling et al. 2019). BrainVar included cortical tissue from 176 individuals falling into two developmental periods: pre-natal, 112 individuals; and post-natal, 60 individuals. We identified $n_{\text{SCZ}} = 25{,}076$ eSNPs as any eQTL SNP-gene pairs provided by Werling et al. (2019) meeting Benjamini-Hochberg $\alpha \leq 0.05$ for at least one of the three sample sets (pre-natal, post-, and complete = all). These eSNPs were used for the SCZ analysis, which is a neurodevelopmental disorder and thus a developmental cohort seemed most appropriate for our analyses.

The second source was the Genotype-Tissue Expression (GTEx) V7 project dataset (GTEx Consortium 2015) with adult samples from fifty-three tissues. As the first winnowing step, we identified the set of GTEx eQTLs for *any* of the available tissues at target FDR level $\alpha = 0.05$. Rather than use all GTEx eQTLs, however, we selected eQTL SNP-gene whose genotypes are most predictive of expression for each gene. The GTEx eSNPs were used for analysis of T2D and BMI, both of which typically onset in adults (for details see *SI Appendix*).

For each eSNP i, we created a vector of covariates x_i to incorporate auxiliary information collected independently of p_i, including p-values from GWAS studies of related phenotypes, and relationships inferred from gene expression studies. First, we utilize the mapping of eSNPs to genes

derived from eQTLs assessed in a relevant tissue type r. Although the majority of observed eSNPs have one unique cis-eQTL gene pairing, 14% of SNPs in BrainVar were eQTL for multiple genes. Let $\mathcal{G}_i^r$ denote the set of genes whose expression is associated with eSNP i and summarize the level of expression as the average absolute eQTL slope for variants in $\mathcal{G}_i^r$ to obtain $\bar{\beta}_i^r$. Additionally, we account for gene co-expression networks as covariates using the $J = 20$ modules reported in the BrainVar study, which were generated using weighted gene co-expression network analysis (Zhang & Horvath 2005, WGCNA). For each of the $j = 1, \ldots, J$ WGCNA modules, we create an indicator variable $\ell_{i,j}^r$ denoting whether or not eSNP i has *any* associated cis-eQTL genes in module j.

For the n_{SCZ} eSNPs, we calculate $\bar{\beta}_i^{type}$ where $type \in \{\text{pre, post, complete}\}$ to capture the eSNP's overall expression association across different epochs of the developmental span. Additionally, we use the 20 WGCNA modules (including unassigned *gray*) reported by Werling et al. (2019) to create indicator variables $\ell_{i,j}^{\text{SCZ}}$ for $j = 1, \ldots, 20$. This culminates in a vector of twenty-four covariates $x_i^{\text{SCZ}} = (z_i^{\text{BD}}, \bar{\beta}_i^{\text{pre}}, \bar{\beta}_i^{\text{post}}, \bar{\beta}_i^{\text{complete}}, \ell_{i,1}^{\text{SCZ}}, \ldots, \ell_{i,20}^{\text{SCZ}})$. Although we use WGCNA modules to make use of the results from the BrainVar study, future applications could explore other approaches to account for gene set and pathway analysis (Zhu & Stephens 2018).

AdaPT discoveries

As noted elsewhere (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014), eSNPs are more likely to be associated with a GWAS phenotype than are randomly chosen SNPs. This is true for the eSNP from BrainVar too, when evaluated in light of the SCZ GWAS p-values (Figure 2A). To evaluate the performance of the AdaPT search algorithm using the eSNP data, we compare the fitted full covariate model to results from its *intercept-only* version (Figure 2B versus 2C). As expected, the intercept-only analysis performs better than BH, with all 269 BH discoveries contained within the intercept-only discoveries, because it incorporates an estimate for the proportion of non-null tests. The full model rejects $R_{\text{SCZ}} = 843$ of the $n_{\text{SCZ}} = 25{,}076$ BrainVar eSNPs versus 361 discoveries for the intercept-only model. For insight into AdaPT's performance, we sequentially include (1) only the BD z-statistics, then (2) include eQTL slope summaries, and then (3) the WGCNA indicators (Figure 2D-E). The largest number of discoveries occurs when all twenty-four covariates are fitted (Figure 2 D), highlighting that all three types of information *together* are required. Notably, only 540 associations are discovered using all covariates without

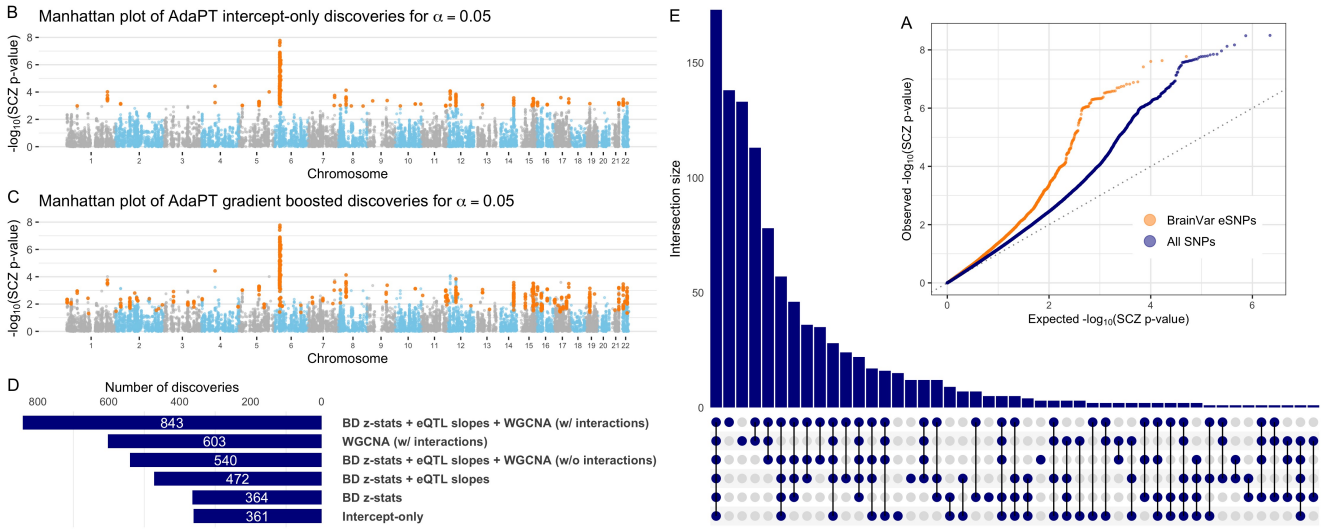

Figure 2: AdaPT results from analysis of schizophrenia (SCZ) p-values. *(A)* Comparison of qq-plots revealing SCZ enrichment for both BrainVar eSNPs compared to the full set of SNPs from *2014* studies. (B-C) Manhattan plots of SCZ AdaPT discoveries (in orange) using *(B)* intercept-only model compared to *(C)* covariate-informed model at target $\alpha = 0.05$. *(D-E)* Comparison of the number of discoveries at target $\alpha = 0.05$ for AdaPT with varying levels of covariates (D) and *(E)* their resulting discovery set intersections.

interactions, fewer discoveries than only using module-based covariates with interactions. This highlights the improvement in AdaPT's performance from modeling the interactions between covariates via gradient boosted trees. As might be expected from their counts of discoveries (Figure 2D), the greatest overlap with the full model occurs by fitting all covariates, but without interactions, or by fitting the module-based covariates (Figure 2E).

Additional discoveries are of little interest if they consist primarily of SNPs in LD with SNPs already discovered using a simpler model, such as the logit model typically used for SCZ GWAS. For context, however, of the initial 25,076 eSNPs we analyzed, only four have p-values $< 5 \times 10^{-8}$, the standard GWAS threshold, and all four occur in the discovery sets for the AdaPT full and intercept-only models. To investigate how the Adapt procedure performs using completely independent eSNPs, we identified the "lead" SNP in each LD block using the approach delineated in Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) and compared model performance for this set of approximately independent SNPs (*SI Appendix*). This thinning results in roughly 3,960 eSNPs to be analyzed by the different models (Figure 2). (Ties in q-values add or subtract a few SNPs to this 3,960 count, depending on the model analyzed.) When AdaPT is fit to these independent SNPs, we obtain analogous improvements in performance compared to the larger set of SNPs (Figures S2 and S3): the full AdaPT model discovers 95 independent loci, while the intercept-only model discovers only 42 loci. Likewise, the full model is the best model and interactions remain important. Finally, no location in the genome exerts unusual influence on the results, which is also the case for the analyses of 25,076 eSNPs.

As described previously, we performed similar analyses of T2D and BMI GWAS p-values. All results for these analyses, as well as more details regarding analyses of SCZ, are available in Dataset S1 and *SI Appendix*.

Variable importance and relationships

We examine the variable importance and partial dependence plots from the gradient boosted models to provide insight into the relationships between each of the covariates and SCZ associations. Figure 3(A) displays the change in variable importance for the probability of being non-null (π_1) at each model fitting iteration, with the top variables in the final model highlighted. We see that the BD z-statistics are estimated as the most important for each π_1 model, but they decrease in

importance in the final steps. In contrast, the unassigned *gray* module increases in important throughout the AdaPT search. This change in variable importance across the AdaPT search highlights that the difference in the discriminatory power of covariates depends on the remaining masked p-values.

Figure 3(B) displays the partial-dependence plot (Friedman 2001) at each AdaPT model fitting iteration for the estimated marginal relationship between the BD z-statistics and the probability of being non-null, evaluated at the $0, 2.5\%, 5\%, \ldots, 100\%$ percentiles. Because the goal of the AdaPT two-groups model (detailed in *Methods*) is to order the remaining masked p-values, the π_1 model predicts values relative to the remaining masked p-values: as the rejection threshold $s_t(x_i)$ becomes more stringent, the masked p-values are more likely non-null (assuming there is signal). However, for each model iteration, Figure 3(B) reveals an increasing likelihood for non-null results as the BD z-statistics grow in magnitude from zero, as well as a diminished impact of BD z-statistics on the estimated π_1 for later model iterations. Figure 3(C) displays the clear enrichment for eSNPs with cis-eQTL genes that are members of the *salmon* WGCNA module reported by Werling et al. (2019), which was the most important WGCNA module indicator in the first model fitting step. This differs from the unassigned *gray* module variable: it is predictive of SNPs that are classified as null, rather than associated with the phenotype. Taken together, Figures 3(A-C) emphasize the use of all covariates across different steps of the AdaPT search. See *SI Appendix* for more analyses highlighting the advantages of accounting for interactions between covariates.

Replication in independent studies

Next, we examine the replicability of the *2014-only* SCZ AdaPT results using independent *2018-only* studies. We find (Figure 4) an increasing smoothing spline relationship between these sets of values, with noticeably increasing evidence indicated by the *2018-only* p-values for the set of AdaPT discoveries at $\alpha = 0.05$. Additionally, of the 843 discoveries from the *2014-only* studies at target FDR level $\alpha = 0.05$, approximately 55.2% (465 eSNPs) were nominal replications for *2018-only* (p-values < 0.05), comparable to the replication fraction expected on the basis of power (see *SI Appendix* for supporting simulations).

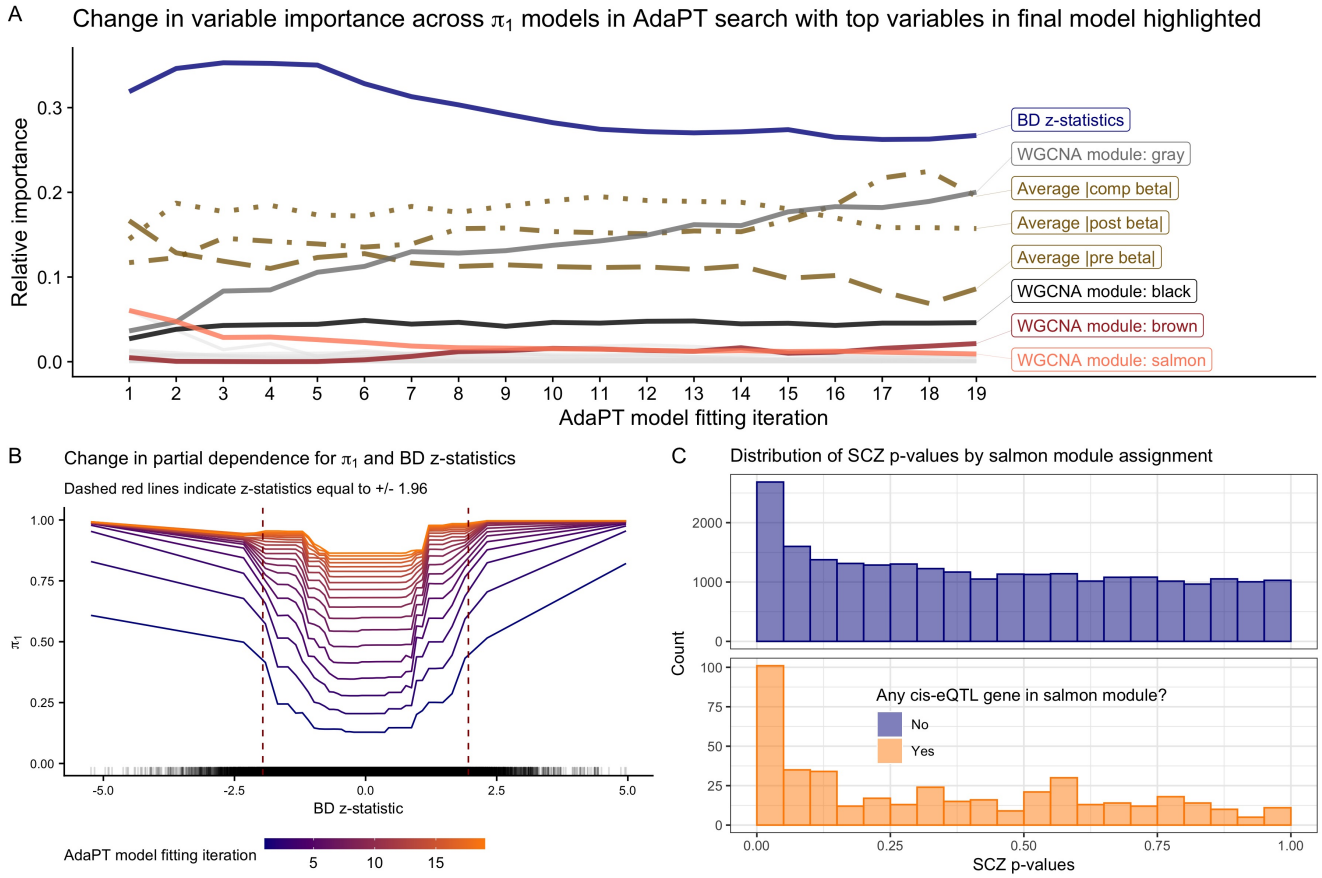

Figure 3: Variable importance and relationships. *(A)* Change in variable importance for AdaPT estimated probability of non-null π_1 model across the search, with top variables in final model highlighted. *(B)* Change in partial dependence for estimated probability of being non-null π_1 and BD z-statistics across π_1 models in AdaPT search. *(C)* SCZ enrichment of eSNPs based on *salmon* WGCNA module membership, the most important WGCNA module indicator in the first model fitting step.

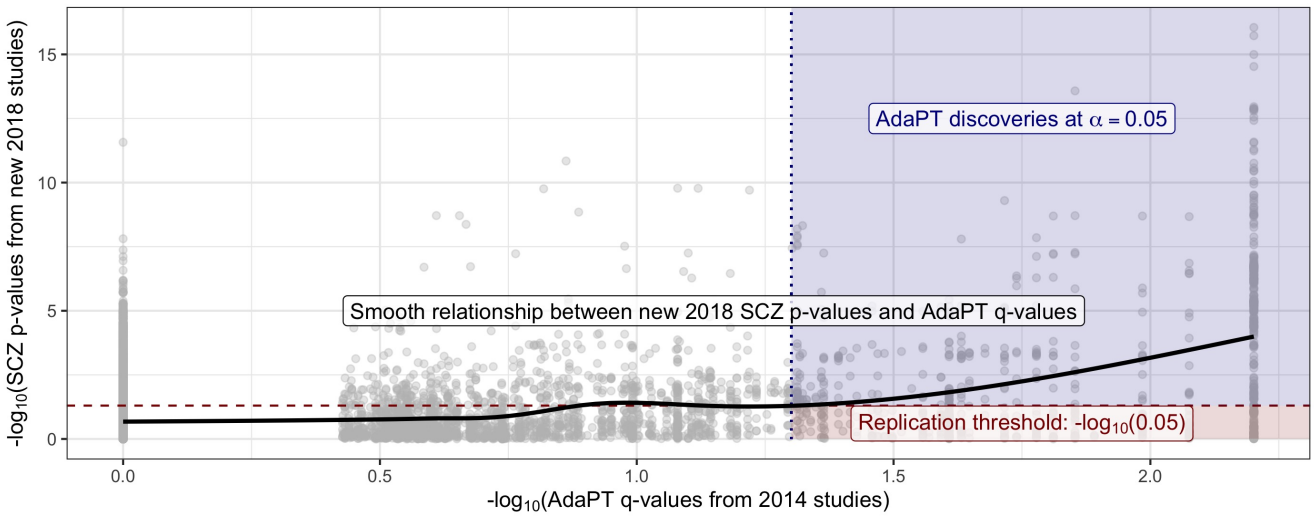

Figure 4: Relationship between the *2018-only* p-values and the resulting *2014-only* q-values from the AdaPT search. Black line displays smoothed relationship between SCZ p-values from *2018-only* studies and AdaPT q-values from *2014-only* studies. Blue region indicates AdaPT discoveries at $\alpha = 0.05$ that are nominal replications, p-values from *2018-only* studies $< .05$, while red region denotes discoveries that failed to replicate.

Gene ontology comparison

Using the SNP discoveries, which span the genome, we next sought biological insights. We applied gene ontology enrichment analysis (Ashburner et al. 2000, The Gene Ontology Consortium 2018) to the 136 genes obtained from the eQTL variant-gene pairs associated with the 843 discoveries. This analysis produced no clear signal, yielding only a minor enrichment for biological processes related to peptide antigen assembly. Several explanations are plausible, we explore two: either AdaPT is discovering SNPs of such small effect that the discoveries are not meaningful or SCZ is a highly complex disorder with a large number of biological processes involved. For comparison we applied our full pipeline to GWAS summary statistics for T2D (Mahajan et al. 2018). This comparison is of interest because T2D is a disease with a well understood functional basis and this is a well powered study (74,124 T2D cases and 824,006 controls). We restricted our analysis to 176,246 eSNPs based on eQTLs obtained using GTEx data. Next, we created eQTL-based covariates using pancreas, liver, and adipose tissue samples (see *SI Appendix* for more details). After creating a vector of covariates from GTEx, AdaPT returned 14,920 eSNPs at $\alpha = 0.05$,

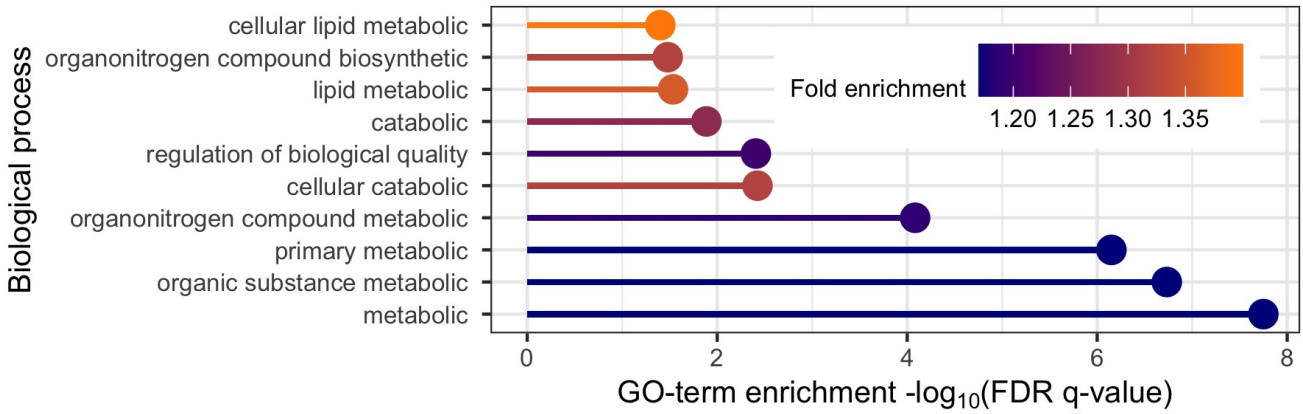

Figure 5: T2D gene ontology enrichment analysis results for top ten biological processes based on positive fold enrichment.

resulting in 5,970 associated genes. Applying gene ontology enrichment analysis to this gene list, we discovered enrichment for biological processes related to lipid metabolic process (Figure 5), consistent with previous literature (Cirillo et al. 2018). These results provide some reassurance that the lack of specificity in the SCZ results can be attributed to the complex etiology of SCZ. For comparison to the well powered BMI GWAS (339,224 subjects), we found a lack of gene ontology enrichment in our gene discoveries (*SI Appendix*).

Pipeline results for all 2018 studies

In addition to applying the pipeline to SCZ p-values from the *2014-only* studies in Ruderfer et al. (2014), we also modeled p-values from *all 2018* studies. The latter yields far more discoveries due to smaller standard errors from increased study sizes, even though the covariates were the same: for x_i^{SCZ}, we find 2,228 discoveries at target FDR level $\alpha = 0.05$ when the pipeline was applied to the p-values for most up-to-date set of studies versus 843 for the *2014-only* studies. Notably, the intercept-only version of AdaPT returned 1,865 discoveries at $\alpha = 0.05$, meaning the covariates contributed to $\approx 19\%$ increase in discovery rate for *all 2018* studies versus the $\approx 134\%$ increase (361 to 843 eSNPs) from using the covariates for the *2014-only* studies. This reinforces the value of using auxiliary information in studies with lower power. Complementary to this observation, AdaPT applied to BMI GWAS the covariate informed models did not yield more discoveries than the intercept-only version (details presented in *SI Appendix*). Simply accounting

for more auxiliary information does not guarantee an improvement in power and the advantages thereof diminishes as power increases, as witnessed by results for *all 2018* studies for SCZ and the large-scale BMI GWAS. Additionally, the larger number of discoveries for the SCZ *all 2018* studies, 2,228, maps onto 382 genes. Despite this increase, these genes did not reveal any clear signal from the Gene Ontology enrichment analysis, comporting with results from the *2014-only* results.

Discussion

Our goals in this study were to explore the use of AdaPT for high-dimensional multi-omics settings and investigate the neurobiology of SCZ in the process. AdaPT was used to analyze a selected set of GWAS summary statistics for SNPs, together with numerous covariates. Specifically, SNPs were selected if they were documented to affect gene expression; these SNP-gene pairs were dubbed eSNPs. Covariates for these eSNPs included GWAS test statistics from a genetically correlated phenotype, BD, which were mapped to eSNPs through SNP identity; as well as features of gene expression and co-expression networks, which were mapped to eSNPs through genes. By coupling flexible gradient boosted trees with the AdaPT procedure, relationships among eSNP GWAS test statistics and covariates were uncovered and more SNPs were found to be associated with SCZ, while maintaining guaranteed finite-sample FDR control. The tree-based handling of covariates addresses a perceived weakness of AdaPT, namely the unintuitive modeling framework for incorporating covariates (Korthauer et al. 2019). Moreover, it is worth noting that the original approach implemented by Lei & Fithian (2018), a generalized linear model with spline bases, yields similar results (361 discoveries at target $\alpha = 0.05$) when applied to the univariate case of only using BD z-statistics. This is an even more straightforward implementation for handling covariates without interactions. The pipeline we built should be simple to mimic for a wide variety of omics and other analyses.

The results shed light on the level of complexity underlying the neurobiology of SCZ. If the origins of SCZ arose by perturbations of one or a few pathways, we would expect to converge on those pathways as we accrue more and more genetic associations. On the other hand, if the ways to generate vulnerability to SCZ were myriad — even if there is an single ultimate cause shared across all cases — then we might expect no such convergence, at least with regards to

the common variation assessed through GWAS. Gene ontology analysis of associated discovery genes from either the *2014-only* or *all 2018* studies reveals no enrichment for biological processes for SCZ. There are many possible explanations for these null findings, one of which is simply a lack of power or specificity of our results. However, the result stands in stark contrast to the results for T2D, for which the gene ontology analysis converges nicely on accepted pathways to T2D risk; yet they comport with those for BMI, which is known to have myriad genetic and environmental origins. Therefore our results are consistent with myriad pathways to vulnerability for SCZ, although it is impossible to rule out other explanations: for example, the possibility that we understand so little about brain functions that gene ontology analyses lack specificity. In any case, our results are consistent with two recent theories underlying the genetics of SCZ, namely extreme polygenicity (O'Connor et al. 2019) and "omnigenic" origins (Boyle et al. 2017).

Although the examples considered in this manuscript pertain to omics data, this process can be adapted for a large variety of settings. We demonstrate in *SI Appendix* simulations showing that AdaPT appears to maintain FDR control in positive dependence settings emulating linkage disequilibrium (LD) block structure underlying GWAS results. There is a clear need, however, for future work to explore AdaPT's properties and computational challenges under various dependence regimes. The growing abundance of contextual information available in "omics" settings provides ample opportunity to improve power for detecting associations, using a flexible approach such as AdaPT, when addressing the multiple testing challenge.

Methods

Two-groups model

The most critical step in the AdaPT algorithm (Lei & Fithian 2018) involves updating the rejection threshold $s_t(x_i)$. Following Lei & Fithian (2018), we use a conditional version of the classical two-groups model (Efron et al. 2001, Scott et al. 2015) where the null p-values are modeled as uniform ($f_0(p|x) \equiv 1$) and we model the non-null p-value density with a beta distribution density parametrized by $\mu_i = \mathbb{E}[-\log(p_i)]$, resulting in a conditional density for a beta mixture model, $f(p|x_i) = \pi_1(x_i)\frac{1}{\mu_i}p^{1/\mu_i - 1} + 1 - \pi_1(x_i)$. In this form, we can model the non-null probability $\pi_1(x_i) = \mathbb{E}[H_i|x_i]$ and the effect size for non-null hypotheses $\mu(x_i) = \mathbb{E}[-\log(p_i)|x_i, H_i = 1]$

with two separate gradient boosted tree-based models. The XGBoost library (Chen & Guestrin 2016) provides logistic and Gamma regression implementations which we use for $\pi_1(x_i)$ and $\mu(x_i)$ respectively.

There are two categories of missing values in these regression problems: H_i is never observed, and at each step t of the search, the p-values for tests $\{i : p_i \leq s_t(x_i) \text{ or } p_i \geq 1 - s_t(x_i)\}$ are masked as $\tilde{p}_{t,i}$. An expectation-maximization (EM) algorithm can be used to estimate both $\hat{\pi}_1(x_i)$ and $\hat{\mu}(x_i)$ by maximizing the partially observed likelihood. We briefly restate the EM algorithm from (1), and provide details in our supplementary materials that reflect the approach taken in the `R adaptMT` package by the same authors, which differs slightly from (1).

During the E-step of the $d = 0, 1, \ldots$ iteration of the EM algorithm, conditional on the partially observed data fixed at step t, $(x_i, \tilde{p}_{t,i})_{i \in [n]}$, we compute both, $\hat{H}_i^{(d)}$ and $\hat{b}_i^{(d)}$, where $\hat{b}_i^{(d)}$ indicates how likely $p'_{t,i} = \min(\tilde{p}_{t,i})$ equals p_i for non-null hypotheses. The explicit calculations of $\hat{H}_i^{(d)}$ and $\hat{b}_i^{(d)}$ are available in the supplementary materials of Lei & Fithian (2018).

The M-step consists of estimating $\hat{\pi}_1^{(d)}$ and $\hat{\mu}^{(d)}$ with separate gradient boosted trees, using *pseudo*-datasets to handle the partially masked data. In order to fit the model for $\pi_1(x_i)$, we construct the response vector $y_\pi^{(d)} = (1, \ldots, 1, 0, \ldots, 0) \in \mathbb{R}^{2n}$ and use weights $w_\pi^{(d)} = (\hat{H}_1^{(d)}, \ldots, \hat{H}_n^{(d)}, 1 - \hat{H}_1^{(d)}, \ldots, 1 - \hat{H}_n^{(d)}) \in \mathbb{R}^{2n}$. Then we estimate $\hat{\pi}_1^{(d)}(x_i)$ using the first n predictions from a classification model using $y_\pi^{(d)}$ as the response variable with the covariate matrix $(x_i)_{i \in [n]}$ replicated twice and weights $w_\pi^{(d)}$. Similarly, for estimating $\hat{\mu}^{(d)}(x_i)$ we construct a response vector $y_\mu^{(d)} = (-\log(p_1), \ldots, -\log(p_n), -\log(1 - p_1), \ldots, -\log(1 - p_n)) \in \mathbb{R}^{2n}$ with weights $w_\mu^{(d)} = (\hat{b}_1^{(d)}, \ldots, \hat{b}_n^{(d)}, 1 - \hat{b}_1^{(d)}, \ldots, 1 - \hat{b}_n^{(d)}) \in \mathbb{R}^{2n}$, and again take the first n predicted values using the duplicated covariate matrix.

We follow the procedure detailed in Section 4.3 of Lei & Fithian (2018) to estimate the conditional local fdr for each $p'_{t,i}$, and then update the rejection threshold to $s_{t+1}(x_i)$ by removing test $i^* = \arg\max_{i \in \mathcal{R}_t} \text{fdr}_{t,i}$ from $\mathcal{R}_t$.

AdaPT gradient boosted trees with CV steps

As a flexible approach for modeling the conditional local fdr, we use gradient boosted trees (Friedman 2001) via the open-source XGBoost implementation (Chen & Guestrin 2016). Gradient boosted trees are an ensemble of many small tree models that jointly contribute to predictions.

Let $f_p \in \mathcal{F}$ be an individual regression tree, then the sum-of-trees model can be written as, $\hat{y}_i = \sum_{p=1}^{P} f_p(x_i)$ to minimize $\sum_i^n L(y_i, \hat{y}_i) + \sum_{p=1}^{P} \Omega(f_p)$ where L is the loss function and Ω measures the complexity of each tree such as the maximum depth, regularization, etc. Chen & Guestrin (2016) detail the algorithms for fitting the model in an additive manner as well as determining the splits for each tree.

To tune the variety of parameters for gradient boosted trees within AdaPT, such as the number of trees P and maximum depth of each tree, we use the cross-validation (CV) approach recommended in Lei & Fithian (2018). If we are considering M different options of boosting parameters, then we evaluate each of the M choices during the modeling phase of the AdaPT search. At step t, we divide the data into K folds preserving the relative proportions of masked and unmasked hypotheses. Then for each set of boosting parameters $m = 1, ..., M$, and for each fold $k = 1, ...K$: (1) apply EM-algorithm to estimate $\hat{\pi}_1^{(m)}(x_i)$ and $\hat{\mu}^{(m)}(x_i)$ using parameters m with data from folds $\{1, \ldots, K\} \backslash \{k\}$, and (2) compute expected-loglikelihood $\tilde{l}_k^{(m)}$ on hold-out set k using two-groups model parameters from m following convergence, and compute total across folds as $\tilde{l}_m = \sum_{k=1}^{K} \tilde{l}_k^{(m)}$. Finally we use the set of parameters $m^* = \arg\max_m \tilde{l}^{(m)}$ in another instance of the EM algorithm to estimate $\hat{\pi}_1^{(m^*)}(x_i)$ and $\hat{\mu}^{(m^*)}(x_i)$ on all data.

Computational aspects of AdaPT

Practical decisions are necessary to implement the AdaPT search. In addition to the covariates and p-values $(x_i, p_{t,i})_{i \in [n]}$, an initial rejection threshold $s_0(x_i)$ is required to begin the search. Rather begin the search with a high starting threshold, such as $s_0^* = 0.45$ recommended by Lei & Fithian (2018), we instead begin the AdaPT search with $s_0^* = 0.05$. Our decision to lower the starting threshold is advantageous for multiple reasons. First, intuitively, this starts our search in the regime of interest for target level $\alpha = 0.05$, whereas we would not expect to detect discoveries with larger p-values using this flexible multiple testing correction. Additionally, by lowering the starting threshold, more true information is available to the gradient boosted trees at the start of the AdaPT search. For instance, with the set of BrainVar eSNPs, 21,248 true p-values are immediately revealed with $s_0^* = 0.05$ as compared to only 2,290 when $s_0^* = 0.45$. Simulations detailed in *SI Appendix* show that on average our choice for using a lower threshold results in higher power.

The most computationally intensive part of the procedure is updating the rejection threshold via the EM algorithm. Instead of updating the model for estimating $\text{fdr}_{t,i}$ at each step of the search, we re-estimate every $[n/20]$ steps as recommended by Lei & Fithian (2018). However, the inclusion of the previously described K-fold CV procedure (we use $K = 5$) for tuning the gradient boosted trees obviously adds computational complexity to the AdaPT search, and would be expensive to apply every time the model fitting takes place. Rather, we apply the CV step once at the beginning, and then another time half-way through the search based on the similarity of simulation performance with varying number of CV steps in *SI Appendix*. Additionally, one needs to choose the potential M model parameter choices. Technically, unique combinations can be used for both models, π_1 and μ, but for simplicity we only consider matching settings for both models, i.e. both models have the same number of trees and maximum depth. As a reminder, AdaPT guarantees finite-sample FDR control **regardless** of potentially over-fitting to the data when using the CV procedure. Simulations are provided in *SI Appendix* showing how extensively increasing the number of trees P leads to decreasing power, but maintains valid FDR control.

We provide a modified version of the `adaptMT R` package to implement the AdaPT-CV tuning steps with XGBoost models at `https://github.com/ryurko/adaptMT`, and provide all code used to generate the manuscript's results at:

`https://github.com/ryurko/AdaPT-GWAS-manuscript-code`.

Acknowledgements

This work was supported by National Institute of Mental Health Grant R37MH057881, the Simons Foundation Grant SFARI SF575097, and the National Science Foundation DMS 1613202 Grant. We are grateful for help from Lambertus Klei regarding data sets and their interpretation.

References

1000 Genomes Project Consortium and others (2012), 'An integrated map of genetic variation from 1,092 human genomes', *Nature* **491**(7422), 56.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P.,

Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. & Sherlock, G. (2000), 'Gene ontology: tool for the unification of biology', *Nature Genetics* **25**(1), 25–29.
URL: *https://doi.org/10.1038/75556*

Benjamini, Y. & Hochberg, Y. (1995), 'Controlling the false discovery rate: A practical and powerful approach to multiple testing', *Journal of the Royal Statistical Society. Series B (Methodological)* **57**(1), 289–300.
URL: *http://www.jstor.org/stable/2346101*

Boca, S. M. & Leek, J. T. (2018), 'A direct approach to estimating false discovery rates conditional on covariates', *PeerJ* **6**, e6035.
URL: *https://doi.org/10.7717/peerj.6035*

Boyle, E. A., Li, Y. I. & Pritchard, J. K. (2017), 'An expanded view of complex traits: From polygenic to omnigenic', *Cell* **169**(7), 1177–1186.
URL: *https://doi.org/10.1016/j.cell.2017.05.038*

Chen, T. & Guestrin, C. (2016), Xgboost: A scalable tree boosting system, *in* 'Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', KDD '16, ACM, New York, NY, USA, pp. 785–794.
URL: *http://doi.acm.org/10.1145/2939672.2939785*

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y. & Li, Y. (2019), *xgboost: Extreme Gradient Boosting*. R package version 0.81.0.1.
URL: *https://CRAN.R-project.org/package=xgboost*

Cirillo, E., Kutmon, M., Gonzalez Hernandez, M., Hooimeijer, T., Adriaens, M. E., Eijssen, L. M. T., Parnell, L. D., Coort, S. L. & Evelo, C. T. (2018), 'From snps to pathways: Biological interpretation of type 2 diabetes (t2dm) genome wide association study (gwas) results', *PLOS ONE* **13**(4), 1–19.
URL: *https://doi.org/10.1371/journal.pone.0193515*

Cross-Disorder Group of the Psychiatric Genomics Consortium (2013), 'Genetic relationship between five psychiatric disorders estimated from genome-wide snps', *Nature Genetics* **45**, 984 EP –.
URL: *https://doi.org/10.1038/ng.2711*

Efron, B., Tibshirani, R., Storey, J. D. & Tusher, V. (2001), 'Empirical bayes analysis of a microarray experiment', *Journal of the American Statistical Association* **96**(456), 1151–1160.
URL: *https://doi.org/10.1198/016214501753382129*

Friedman, J. H. (2001), 'Greedy function approximation: A gradient boosting machine.', *The Annals of Statistics* **29**(5), 1189–1232.
URL: *https://doi.org/10.1214/aos/1013203451*

Genovese, C. R., Roeder, K. & Wasserman, L. (2006), 'False discovery control with p-value weighting', *Biometrika* **93**(3), 509–524.
URL: *https://doi.org/10.1093/biomet/93.3.509*

GTEx Consortium (2015), 'The genotype-tissue expression (gtex) pilot analysis: Multitissue gene regulation in humans', *Science* **348**(6235), 648–660.
URL: *https://science.sciencemag.org/content/348/6235/648*

Ignatiadis, N., Klaus, B., Zaugg, J. B. & Huber, W. (2016), 'Data-driven hypothesis weighting increases detection power in genome-scale multiple testing', *Nature methods* **13**(7), 577–580.
URL: *https://www.ncbi.nlm.nih.gov/pmc/PMC4930141/*

Korthauer, K., Kimes, P. K., Duvallet, C., Reyes, A., Subramanian, A., Teng, M., Shukla, C., Alm, E. J. & Hicks, S. C. (2019), 'A practical guide to methods controlling false discoveries in computational biology', *Genome Biology* **20**(1), 118.
URL: *https://doi.org/10.1186/s13059-019-1716-1*

Lei, L. & Fithian, W. (2018), 'Adapt: an interactive procedure for multiple testing with side information', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**(4), 649–679.
URL: *https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssb.12274*

Li, A. & Barber, R. F. (2019), 'Multiple testing with the structure-adaptive benjaminihochberg algorithm', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **81**(1), 45–74.
URL: *https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssb.12298*

Lichtenstein, P., Yip, B. H., Björk, C., Pawitan, Y., Cannon, T. D., Sullivan, P. F. & Hultman, C. M. (2009), 'Common genetic determinants of schizophrenia and bipolar disorder in swedish families: a population-based study', *The Lancet* **373**(9659), 234–239.
URL: *https://doi.org/10.1016/S0140-6736(09)60072-6*

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., Powell, C., Vedantam, S., Buchkovich, M. L., Yang, J., Croteau-Chonka, D. C., Esko, T., Fall, T., Ferreira, T., Gustafsson, S., Kutalik, Z., Luan, J., Mägi, R., Randall, J. C., Winkler, T. W., Wood, A. R., Workalemahu, T., Faul, J. D., Smith, J. A., Hua Zhao, J., Zhao, W., Chen, J., Fehrmann, R., Hedman, Å. K., Karjalainen, J., Schmidt, E. M., Absher, D., Amin, N., Anderson, D., Beekman, M., Bolton, J. L., Bragg-Gresham, J. L., Buyske, S., Demirkan, A., Deng, G., Ehret, G. B., Feenstra, B., Feitosa, M. F., Fischer, K., Goel, A., Gong, J., Jackson, A. U., Kanoni, S., Kleber, M. E., Kristiansson, K., Lim, U., Lotay, V., Mangino, M., Mateo Leach, I., Medina-Gomez, C., Medland, S. E., Nalls, M. A., Palmer, C. D., Pasko, D., Pechlivanis, S., Peters, M. J., Prokopenko, I., Shungin, D., Stančáková, A., Strawbridge, R. J., Ju Sung, Y., Tanaka, T., Teumer, A., Trompet, S., van der Laan, S. W., van Setten, J., Van Vliet-Ostaptchouk, J. V., Wang, Z., Yengo, L., Zhang, W., Isaacs, A., Albrecht, E., Ärnlöv, J., Arscott, G. M., Attwood, A. P., Bandinelli, S., Barrett, A., Bas, I. N., Bellis, C., Bennett, A. J., Berne, C., Blagieva, R., Blüher, M., Böhringer, S., Bonnycastle, L. L., Böttcher, Y., Boyd, H. A., Bruinenberg, M., Caspersen, I. H., Ida Chen, Y.-D., Clarke, R., Warwick Daw, E., de Craen, A. J. M., Delgado, G., Dimitriou, M., Doney, A. S. F., Eklund, N., Estrada, K., Eury, E., Folkersen, L., Fraser, R. M., Garcia, M. E., Geller, F., Giedraitis, V., Gigante, B., Go, A. S., Golay, A., Goodall, A. H., Gordon, S. D., Gorski, M., Grabe, H.-J., Grallert, H., Grammer, T. B., Gräßler, J., Grönberg, H., Groves, C. J., Gusto, G., Haessler, J., Hall, P., Haller, T., Hallmans, G., Hartman, C. A., Hassinen, M., Hayward, C., Heard-Costa, N. L., Helmer, Q., Hengstenberg, C., Holmen, O., Hottenga, J.-J., James, A. L., Jeff, J. M., Johansson, Å., Jolley, J., Juliusdottir, T., Kinnunen, L., Koenig, W., Koskenvuo, M., Kratzer, W., Laitinen, J., Lamina, C., Leander,

K., Lee, N. R., Lichtner, P., Lind, L., Lindström, J., Sin Lo, K., Lobbens, S., Lorbeer, R., Lu, Y., Mach, F., Magnusson, P. K. E., Mahajan, A., McArdle, W. L., McLachlan, S., Menni, C., Merger, S., Mihailov, E., Milani, L., Moayyeri, A., Monda, K. L., Morken, M. A., Mulas, A., Müller, G., Müller-Nurasyid, M., Musk, A. W., Nagaraja, R., Nöthen, M. M., Nolte, I. M., Pilz, S., Rayner, N. W., Renstrom, F., Rettig, R., Ried, J. S., Ripke, S., Robertson, N. R., Rose, L. M., Sanna, S., Scharnagl, H., Scholtens, S., Schumacher, F. R., Scott, W. R., Seufferlein, T., Shi, J., Vernon Smith, A., Smolonska, J., Stanton, A. V., Steinthorsdottir, V., Stirrups, K., Stringham, H. M., Sundström, J., Swertz, M. A., Swift, A. J., Syvänen, A.-C., Tan, S.-T., Tayo, B. O., Thorand, B., Thorleifsson, G., Tyrer, J. P., Uh, H.-W., Vandenput, L., Verhulst, F. C., Vermeulen, S. H., Verweij, N., Vonk, J. M., Waite, L. L., Warren, H. R., Waterworth, D., Weedon, M. N., Wilkens, L. R., Willenborg, C., Wilsgaard, T., Wojczynski, M. K., Wong, A., Wright, A. F., Zhang, Q., Study, T. L. C., Brennan, E. P., Choi, M., Dastani, Z., Drong, A. W., Eriksson, P., Franco-Cereceda, A., Gådin, J. R., Gharavi, A. G., Goddard, M. E., Handsaker, R. E., Huang, J., Karpe, F., Kathiresan, S., Keildson, S., Kiryluk, K., Kubo, M., Lee, J.-Y., Liang, L., Lifton, R. P., Ma, B., McCarroll, S. A., McKnight, A. J., Min, J. L., Moffatt, M. F., Montgomery, G. W., Murabito, J. M., Nicholson, G., Nyholt, D. R., Okada, Y., Perry, J. R. B., Dorajoo, R., Reinmaa, E., Salem, R. M., Sandholm, N., Scott, R. A., Stolk, L., Takahashi, A., Tanaka, T., van't Hooft, F. M., Vinkhuyzen, A. A. E., Westra, H.-J., Zheng, W., Zondervan, K. T., Consortium, T. A., Group, T. A.-B. W., Consortium, T. C., Consortium, T. C., GLGC, T., ICBP, T., Investigators, T. M., Consortium, T. M., Consortium, T. M., Consortium, T. P., Consortium, T. R., Consortium, T. G., Consortium, T. I. E., Heath, A. C., Arveiler, D., Bakker, S. J. L., Beilby, J., Bergman, R. N., Blangero, J., Bovet, P., Campbell, H., Caulfield, M. J., Cesana, G., Chakravarti, A., Chasman, D. I., Chines, P. S., Collins, F. S., Crawford, D. C., Adrienne Cupples, L., Cusi, D., Danesh, J., de Faire, U., den Ruijter, H. M., Dominiczak, A. F., Erbel, R., Erdmann, J., Eriksson, J. G., Farrall, M., Felix, S. B., Ferrannini, E., Ferrières, J., Ford, I., Forouhi, N. G., Forrester, T., Franco, O. H., Gansevoort, R. T., Gejman, P. V., Gieger, C., Gottesman, O., Gudnason, V., Gyllensten, U., Hall, A. S., Harris, T. B., Hattersley, A. T., Hicks, A. A., Hindorff, L. A., Hingorani, A. D., Hofman, A., Homuth, G., Kees Hovingh, G., Humphries, S. E., Hunt, S. C., Hyppönen, E., Illig, T., Jacobs, K. B., Jarvelin, M.-R., Jöckel, K.-H., Johansen, B., Jousilahti, P., Wouter Jukema, J.,

Jula, A. M., Kaprio, J., Kastelein, J. J. P., Keinanen-Kiukaanniemi, S. M., Kiemeney, L. A., Knekt, P., Kooner, J. S., Kooperberg, C., Kovacs, P., Kraja, A. T., Kumari, M., Kuusisto, J., Lakka, T. A., Langenberg, C., Le Marchand, L., Lehtimäki, T., Lyssenko, V., Männistö, S., Marette, A., Matise, T. C., McKenzie, C. A., McKnight, B., Moll, F. L., Morris, A. D., Morris, A. P., Murray, J. C., Nelis, M., Ohlsson, C., Oldehinkel, A. J., Ong, K. K., Madden, P. A. F., Pasterkamp, G., Peden, J. F., Peters, A., Postma, D. S., Pramstaller, P. P., Price, J. F., Qi, L., Raitakari, O. T., Rankinen, T., Rao, D. C., Rice, T. K., Ridker, P. M., Rioux, J. D., Ritchie, M. D., Rudan, I., Salomaa, V., Samani, N. J., Saramies, J., Sarzynski, M. A., Schunkert, H., Schwarz, P. E. H., Sever, P., Shuldiner, A. R., Sinisalo, J., Stolk, R. P., Strauch, K., Tönjes, A., Trégouët, D.-A., Tremblay, A., Tremoli, E., Virtamo, J., Vohl, M.-C., Völker, U., Waeber, G., Willemsen, G., Witteman, J. C., Carola Zillikens, M., Adair, L. S., Amouyel, P., Asselbergs, F. W., Assimes, T. L., Bochud, M., Boehm, B. O., Boerwinkle, E., Bornstein, S. R., Bottinger, E. P., Bouchard, C., Cauchi, S., Chambers, J. C., Chanock, S. J., Cooper, R. S., de Bakker, P. I. W., Dedoussis, G., Ferrucci, L., Franks, P. W., Froguel, P., Groop, L. C., Haiman, C. A., Hamsten, A., Hui, J., Hunter, D. J., Hveem, K., Kaplan, R. C., Kivimaki, M., Kuh, D., Laakso, M., Liu, Y., Martin, N. G., März, W., Melbye, M., Metspalu, A., Moebus, S., Munroe, P. B., Njølstad, I., Oostra, B. A., Palmer, C. N. A., Pedersen, N. L., Perola, M., Pérusse, L., Peters, U., Power, C., Quertermous, T., Rauramaa, R., Rivadeneira, F., Saaristo, T. E., Saleheen, D., Sattar, N., Schadt, E. E., Schlessinger, D., Eline Slagboom, P., Snieder, H., Spector, T. D., Thorsteinsdottir, U., Stumvoll, M., Tuomilehto, J., Uitterlinden, A., Uusitupa, M., van der Harst, P., Walker, M., Wallaschofski, H., Wareham, N. J., Watkins, H., Weir, D. R., Wichmann, H.-E., Wilson, J. F., Zanen, P., Borecki, I. B., Deloukas, P., Fox, C. S., Heid, I. M., O'Connell, J. R., Strachan, D. P., Stefansson, K., van Duijn, C. M., Abecasis, G. R., Franke, L., Frayling, T. M., McCarthy, M. I., Visscher, P. M., Scherag, A., Willer, C. J., Boehnke, M., Mohlke, K. L., Lindgren, C. M., Beckmann, J. S., Barroso, I., North, K. E., Ingelsson, E., Hirschhorn, J. N., Loos, R. J. F. & Speliotes, E. K. (2015), 'Genetic studies of body mass index yield new insights for obesity biology', *Nature* **518**, 197 EP –.

URL: *http://dx.doi.org/10.1038/nature14177*

Mahajan, A., Taliun, D., Thurner, M., Robertson, N. R., Torres, J. M., Rayner, N. W., Payne, A. J., Steinthorsdottir, V., Scott, R. A., Grarup, N., Cook, J. P., Schmidt, E. M., Wuttke,

M., Sarnowski, C., Mägi, R., Nano, J., Gieger, C., Trompet, S., Lecoeur, C., Preuss, M. H., Prins, B. P., Guo, X., Bielak, L. F., Below, J. E., Bowden, D. W., Chambers, J. C., Kim, Y. J., Ng, M. C. Y., Petty, L. E., Sim, X., Zhang, W., Bennett, A. J., Bork-Jensen, J., Brummett, C. M., Canouil, M., Ec kardt, K.-U., Fischer, K., Kardia, S. L. R., Kronenberg, F., Läll, K., Liu, C.-T., Locke, A. E., Luan, J., Ntalla, I., Nylander, V., Schönherr, S., Schurmann, C., Yengo, L., Bottinger, E. P., Brandslund, I., Christensen, C., Dedoussis, G., Florez, J. C., Ford, I., Franco, O. H., Frayling, T. M., Giedraitis, V., Hackinger, S., Hattersley, A. T., Herder, C., Ikram, M. A., Ingelsson, M., Jørgensen, M. E., Jørgensen, T., Kriebel, J., Kuusisto, J., Ligthart, S., Lindgren, C. M., Linneberg, A., Lyssenko, V., Mamakou, V., Meitinger, T., Mohlke, K. L., Morris, A. D., Nadkarni, G., Pankow, J. S., Peters, A., Sattar, N., Stančáková, A., Strauch, K., Taylor, K. D., Thorand, B., Thorleifsson, G., Thorsteinsdottir, U., Tuomilehto, J., Witte, D. R., Dupuis, J., Peyser, P. A., Zeggini, E., Loos, R. J. F., Froguel, P., Ingelsson, E., Lind, L., Groop, L., Laakso, M., Collins, F. S., Jukema, J. W., Palmer, C. N. A., Grallert, H., Metspalu, A., Dehghan, A., Köttgen, A., Abecasis, G. R., Meigs, J. B., Rotter, J. I., Marchini, J., Pedersen, O., Hansen, T., Langenberg, C., Wareham, N. J., Stefansson, K., Gloyn, A. L., Morris, A. P., Boehnke, M. & McCarthy, M. I. (2018), 'Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps', *Nature Genetics* **50**(11), 1505–1513.

URL: *https://doi.org/10.1038/s41588-018-0241-6*

Nicolae, D. L., Gamazon, E., Zhang, W., Duan, S., Dolan, M. E. & Cox, N. J. (2010), 'Trait-associated snps are more likely to be eqtls: Annotation to enhance discovery from gwas', *PLOS Genetics* **6**(4), 1–10.

URL: *https://doi.org/10.1371/journal.pgen.1000888*

O'Connor, L. J., Schoech, A. P., Hormozdiari, F., Gazal, S., Patterson, N. & Price, A. L. (2019), 'Extreme polygenicity of complex traits is explained by negative selection', *The American Journal of Human Genetics* **105**(3), 456 – 476.

URL: *http://www.sciencedirect.com/science/article/pii/S0002929719302666*

R Core Team (2018), *R: A Language and Environment for Statistical Computing*, R Foundation

for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Ren, Z. & Candès, E. (2020), 'Knockoffs with side information', *arXiv preprint arXiv:2001.07835* .

Ruderfer, D. M., Fanous, A. H., Ripke, S., McQuillin, A., Amdur, R. L., of the Psychiatric Genomics Consortium, S. W. G., of the Psychiatric Genomics Consortium, B. D. W. G., of the Psychiatric Genomics Consortium, C.-D. W. G., Gejman, P. V., O'Donovan, M. C., Andreassen, O. A., Djurovic, S., Hultman, C. M., Kelsoe, J. R., Jamain, S., Landén, M., Leboyer, M., Nimgaonkar, V., Nurnberger, J., Smoller, J. W., Craddock, N., Corvin, A., Sullivan, P. F., Holmans, P., Sklar, P. & Kendler, K. S. (2014), 'Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia', *Molecular psychiatry* **19**(9), 1017–1024.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/24280982*

Ruderfer, D. M., Ripke, S., McQuillin, A., Boocock, J., Stahl, E. A., Pavlides, J. M. W., Mullins, N., Charney, A. W., Ori, A. P., Loohuis, L. M. O., Domenici, E., Florio, A. D., Papiol, S., Kalman, J. L., Trubetskoy, V., Adolfsson, R., Agartz, I., Agerbo, E., Akil, H., Albani, D., Albus, M., Alda, M., Alexander, M., Alliey-Rodriguez, N., Als, T. D., Amin, F., Anjorin, A., Arranz, M. J., Awasthi, S., Bacanu, S. A., Badner, J. A., Baekvad-Hansen, M., Bakker, S., Band, G., Barchas, J. D., Barroso, I., Bass, N., Bauer, M., Baune, B. T., Begemann, M., Bellenguez, C., Belliveau, R. A., Bellivier, F., Bender, S., Bene, J., Bergen, S. E., Berrettini, W. H., Bevilacqua, E., Biernacka, J. M., Bigdeli, T. B., Black, D. W., Blackburn, H., Blackwell, J. M., Blackwood, D. H., Pedersen, C. B., Boehnke, M., Boks, M., Borglum, A. D., Bramon, E., Breen, G., Brown, M. A., Bruggeman, R., Buccola, N. G., Buckner, R. L., Budde, M., Bulik-Sullivan, B., Bumpstead, S. J., Bunney, W., Burmeister, M., Buxbaum, J. D., Bybjerg-Grauholm, J., Byerley, W., Cahn, W., Cai, G., Cairns, M. J., Campion, D., Cantor, R. M., Carr, V. J., Carrera, N., Casas, J. P., Casas, M., Catts, S. V., Cervantes, P., Chambert, K. D., Chan, R. C., Chen, E. Y., Chen, R. Y., Cheng, W., Cheung, E. F., Chong, S. A., Clarke, T.-K., Cloninger, C. R., Cohen, D., Cohen, N., Coleman, J. R., Collier, D. A., Cormican, P., Coryell, W., Craddock, N., Craig, D. W., Crespo-Facorro, B., Crowley, J. J., Cruceanu, C., Curtis, D., Czerski, P. M., Dale, A. M., Daly, M. J., Dannlowski, U., Darvasi, A., Davidson, M., Davis,

K. L., de Leeuw, C. A., Degenhardt, F., Favero, J. D., DeLisi, L. E., Deloukas, P., Demontis, D., DePaulo, J. R., di Forti, M., Dikeos, D., Dinan, T., Djurovic, S., Dobbyn, A. L., Donnelly, P., Donohoe, G., Drapeau, E., Dronov, S., Duan, J., Dudbridge, F., Duncanson, A., Edenberg, H., Edkins, S., Ehrenreich, H., Eichhammer, P., Elvsashagen, T., Eriksson, J., Escott-Price, V., Esko, T., Essioux, L., Etain, B., Fan, C. C., Farh, K.-H., Farrell, M. S., Flickinger, M., Foroud, T. M., Forty, L., Frank, J., Franke, L., Fraser, C., Freedman, R., Freeman, C., Freimer, N. B., Friedman, J. I., Fromer, M., Frye, M. A., Fullerton, J. M., Gade, K., Garnham, J., Gaspar, H. A., Gejman, P. V., Genovese, G., Georgieva, L., Giambartolomei, C., Giannoulatou, E., Giegling, I., Gill, M., Gillman, M., Pedersen, M. G., Giusti-Rodriguez, P., Godard, S., Goes, F., Goldstein, J. I., Gopal, S., Gordon, S. D., Gordon-Smith, K., Gratten, J., Gray, E., Green, E. K., Green, M. J., Greenwood, T. A., Grigoroiu-Serbanescu, M., Grove, J., Guan, W., Gurling, H., Parra, J. G., Gwilliam, R., de Haan, L., Hall, J., Hall, M.-H., Hammer, C., Hammond, N., Hamshere, M. L., Hansen, M., Hansen, T., Haroutunian, V., Hartmann, A. M., Hauser, J., Hautzinger, M., Heilbronner, U., Hellenthal, G., Henskens, F. A., Herms, S., Hipolito, M., Hirschhorn, J. N., Hoffmann, P., Hollegaard, M. V., Hougaard, D. M., Huang, H., Huckins, L., Hultman, C. M., Hunt, S. E., Ikeda, M., Iwata, N., Iyegbe, C., Jablensky, A. V., Jamain, S., Jankowski, J., Jayakumar, A., Joa, I., Jones, I., Jones, L. A., Jonsson, E. G., Julia, A., Jureus, A., Kahler, A. K., Kahn, R. S., Kalaydjieva, L., Kandaswamy, R., Karachanak-Yankova, S., Karjalainen, J., Karlsson, R., Kavanagh, D., Keller, M. C., Kelly, B. J., Kelsoe, J., Kennedy, J. L., Khrunin, A., Kim, Y., Kirov, G., Kittel-Schneider, S., Klovins, J., Knight, J., Knott, S. V., Knowles, J. A., Kogevinas, M., Konte, B., Kravariti, E., Kucinskas, V., Kucinskiene, Z. A., Kupka, R., Kuzelova-Ptackova, H., Landen, M., Langford, C., Laurent, C., Lawrence, J., Lawrie, S., Lawson, W. B., Leber, M., Leboyer, M., Lee, P. H., Keong, J. L. C., Legge, S. E., Lencz, T., Lerer, B., Levinson, D. F., Levy, S. E., Lewis, C. M., Li, J. Z., Li, M., Li, Q. S., Li, T., Liang, K.-Y., Liddle, J., Lieberman, J., Limborska, S., Lin, K., Linszen, D. H., Lissowska, J., Liu, C., Liu, J., Lonnqvist, J., Loughland, C. M., Lubinski, J., Lucae, S., Macek, M., MacIntyre, D. J., Magnusson, P. K., Maher, B. S., Mahon, P. B., Maier, W., Malhotra, A. K., Mallet, J., Malt, U. F., Markus, H. S., Marsal, S., Martin, N. G., Mata, I., Mathew, C. G., Mattheisen, M., Mattingsdal, M., Mayoral, F., McCann, O. T., McCarley, R. W., McCarroll, S. A., McCarthy, M. I., McDonald, C., McElroy, S. L., McGuffin, P., McInnis, M. G., McIntosh,

A. M., McKay, J. D., McMahon, F. J., Medeiros, H., Medland, S. E., Meier, S., Meijer, C. J., Melegh, B., Melle, I., Meng, F., Mesholam-Gately, R. I., Metspalu, A., Michie, P. T., Milani, L., Milanova, V., Mitchell, P. B., Mokrab, Y., Montgomery, G. W., Moran, J. L., Morken, G., Morris, D. W., Mors, O., Mortensen, P. B., Mowry, B. J., Mhleisen, T. W., Mller-Myhsok, B., Murphy, K. C., Murray, R. M., Myers, R. M., Myin-Germeys, I., Neale, B. M., Nelis, M., Nenadic, I., Nertney, D. A., Nestadt, G., Nicodemus, K. K., Nievergelt, C. M., Nikitina-Zake, L., Nimgaonkar, V., Nisenbaum, L., Nordentoft, M., Nordin, A., Nthen, M. M., Nwulia, E. A., OCallaghan, E., ODonovan, C., ODushlaine, C., ONeill, F. A., Oedegaard, K. J., Oh, S.-Y., Olincy, A., Olsen, L., Oruc, L., Os, J. V., Owen, M. J., Paciga, S. A., Palmer, C. N., Palotie, A., Pantelis, C., Papadimitriou, G. N., Parkhomenko, E., Pato, C., Pato, M. T., Paunio, T., Pearson, R., Perkins, D. O., Perlis, R. H., Perry, A., Pers, T. H., Petryshen, T. L., Pfennig, A., Picchioni, M., Pietilainen, O., Pimm, J., Pirinen, M., Plomin, R., Pocklington, A. J., Posthuma, D., Potash, J. B., Potter, S. C., Powell, J., Price, A., Pulver, A. E., Purcell, S. M., Quested, D., Ramos-Quiroga, J. A., Rasmussen, H. B., Rautanen, A., Ravindrarajah, R., Regeer, E. J., Reichenberg, A., Reif, A., Reimers, M. A., Ribases, M., Rice, J. P., Richards, A. L., Ricketts, M., Riley, B. P., Rivas, F., Rivera, M., Roffman, J. L., Rouleau, G. A., Roussos, P., Rujescu, D., Salomaa, V., Sanchez-Mora, C., Sanders, A. R., Sawcer, S. J., Schall, U., Schatzberg, A. F., Scheftner, W. A., Schofield, P. R., Schork, N. J., Schwab, S. G., Scolnick, E. M., Scott, L. J., Scott, R. J., Seidman, L. J., Serretti, A., Sham, P. C., Weickert, C. S., Shehktman, T., Shi, J., Shilling, P. D., Sigurdsson, E., Silverman, J. M., Sim, K., Slaney, C., Slominsky, P., Smeland, O. B., Smoller, J. W., So, H.-C., Sobell, J. L., Soderman, E., Hansen, C. S., Spencer, C. C., Spijker, A. T., Clair, D. S., Stefansson, H., Stefansson, K., Steinberg, S., Stogmann, E., Stordal, E., Strange, A., Straub, R. E., Strauss, J. S., Streit, F., Strengman, E., Strohmaier, J., Stroup, T. S., Su, Z., Subramaniam, M., Suvisaari, J., Svrakic, D. M., Szatkiewicz, J. P., Szelinger, S., Tashakkori-Ghanbaria, A., Thirumalai, S., Thompson, R. C., Thorgeirsson, T. E., Toncheva, D., Tooney, P. A., Tosato, S., Toulopoulou, T., Trembath, R. C., Treutlein, J., Trubetskoy, V., Turecki, G., Vaaler, A. E., Vedder, H., Vieta, E., Vincent, J., Visscher, P. M., Viswanathan, A. C., Vukcevic, D., Waddington, J., Waller, M., Walsh, D., Walshe, M., Walters, J. T., Wang, D., Wang, Q., Wang, W., Wang, Y., Watson, S. J., Webb, B. T., Weickert, T. W., Weinberger, D. R., Weisbrod, M., Weiser, M., Werge, T., Weston, P., Whittaker, P., Widaa, S., Wiersma,

D., Wildenauer, D. B., Williams, N. M., Williams, S., Witt, S. H., Wolen, A. R., Wong, E. H., Wood, N. W., Wormley, B. K., Wu, J. Q., Xi, S., Xu, W., Young, A. H., Zai, C. C., Zandi, P., Zhang, P., Zheng, X., Zimprich, F., Zollner, S., Corvin, A., Fanous, A. H., Cichon, S., Rietschel, M., Gershon, E. S., Schulze, T. G., Cuellar-Barboza, A. B., Forstner, A. J., Holmans, P. A., Nurnberger, J. I., Andreassen, O. A., Lee, S. H., ODonovan, M. C., Sullivan, P. F., Ophoff, R. A., Wray, N. R., Sklar, P. & Kendler, K. S. (2018), 'Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes', *Cell* **173**(7), 1705 – 1715.e16.
URL: *http://www.sciencedirect.com/science/article/pii/S0092867418306585*

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014), 'Biological insights from 108 schizophrenia-associated genetic loci', *Nature* **511**(7510), 421–427.
URL: *https://www.ncbi.nlm.nih.gov/pubmed/25056061*

Scott, J. G., Kelly, R. C., Smith, M. A., Zhou, P. & Kass, R. E. (2015), 'False discovery rate regression: An application to neural synchrony detection in primary visual cortex', *Journal of the American Statistical Association* **110**(510), 459–471.
URL: *https://doi.org/10.1080/01621459.2014.990973*

Shungin, D., Winkler, T. W., Croteau-Chonka, D. C., Ferreira, T., Locke, A. E., Mägi, R., Strawbridge, R. J., Pers, T. H., Fischer, K., Justice, A. E., Workalemahu, T., Wu, J. M. W., Buchkovich, M. L., Heard-Costa, N. L., Roman, T. S., Drong, A. W., Song, C., Gustafsson, S., Day, F. R., Esko, T., Fall, T., Kutalik, Z., Luan, J., Randall, J. C., Scherag, A., Vedantam, S., Wood, A. R., Chen, J., Fehrmann, R., Karjalainen, J., Kahali, B., Liu, C.-T., Schmidt, E. M., Absher, D., Amin, N., Anderson, D., Beekman, M., Bragg-Gresham, J. L., Buyske, S., Demirkan, A., Ehret, G. B., Feitosa, M. F., Goel, A., Jackson, A. U., Johnson, T., Kleber, M. E., Kristiansson, K., Mangino, M., Mateo Leach, I., Medina-Gomez, C., Palmer, C. D., Pasko, D., Pechlivanis, S., Peters, M. J., Prokopenko, I., Stančáková, A., Ju Sung, Y., Tanaka, T., Teumer, A., Van Vliet-Ostaptchouk, J. V., Yengo, L., Zhang, W., Albrecht, E., Ärnlöv, J., Arscott, G. M., Bandinelli, S., Barrett, A., Bellis, C., Bennett, A. J., Berne, C., Blüher, M., Böhringer, S., Bonnet, F., Böttcher, Y., Bruinenberg, M., Carba, D. B., Caspersen, I. H., Clarke, R., Warwick Daw, E., Deelen, J., Deelman, E., Delgado, G., Doney, A. S. F., Eklund, N., Erdos, M. R., Estrada, K., Eury, E., Friedrich, N., Garcia, M. E., Giedraitis, V., Gigante,

B., Go, A. S., Golay, A., Grallert, H., Grammer, T. B., Gräßler, J., Grewal, J., Groves, C. J., Haller, T., Hallmans, G., Hartman, C. A., Hassinen, M., Hayward, C., Heikkilä, K., Herzig, K.-H., Helmer, Q., Hillege, H. L., Holmen, O., Hunt, S. C., Isaacs, A., Ittermann, T., James, A. L., Johansson, I., Juliusdottir, T., Kalafati, I.-P., Kinnunen, L., Koenig, W., Kooner, I. K., Kratzer, W., Lamina, C., Leander, K., Lee, N. R., Lichtner, P., Lind, L., Lindström, J., Lobbens, S., Lorentzon, M., Mach, F., Magnusson, P. K. E., Mahajan, A., McArdle, W. L., Menni, C., Merger, S., Mihailov, E., Milani, L., Mills, R., Moayyeri, A., Monda, K. L., Mooijaart, S. P., Mühleisen, T. W., Mulas, A., Müller, G., Müller-Nurasyid, M., Nagaraja, R., Nalls, M. A., Narisu, N., Glorioso, N., Nolte, I. M., Olden, M., Rayner, N. W., Renstrom, F., Ried, J. S., Robertson, N. R., Rose, L. M., Sanna, S., Scharnagl, H., Scholtens, S., Sennblad, B., Seufferlein, T., Sitlani, C. M., Vernon Smith, A., Stirrups, K., Stringham, H. M., Sundström, J., Swertz, M. A., Swift, A. J., Syvänen, A.-C., Tayo, B. O., Thorand, B., Thorleifsson, G., Tomaschitz, A., Troffa, C., van Oort, F. V. A., Verweij, N., Vonk, J. M., Waite, L. L., Wennauer, R., Wilsgaard, T., Wojczynski, M. K., Wong, A., Zhang, Q., Hua Zhao, J., Brennan, E. P., Choi, M., Eriksson, P., Folkersen, L., Franco-Cereceda, A., Gharavi, A. G., Hedman, Å. K., Hivert, M.-F., Huang, J., Kanoni, S., Karpe, F., Keildson, S., Kiryluk, K., Liang, L., Lifton, R. P., Ma, B., McKnight, A. J., McPherson, R., Metspalu, A., Min, J. L., Moffatt, M. F., Montgomery, G. W., Murabito, J. M., Nicholson, G., Nyholt, D. R., Olsson, C., Perry, J. R. B., Reinmaa, E., Salem, R. M., Sandholm, N., Schadt, E. E., Scott, R. A., Stolk, L., Vallejo, E. E., Westra, H.-J., Zondervan, K. T., Consortium, T. A., Consortium, T. C., Consortium, T. C., Consortium, T. G., Consortium, T. G., GLGC, T., ICBP, T., Consortium, T. I. E., Study, T. L. C., Investigators, T. M., Consortium, T. M., Consortium, T. P., Consortium, T. R., Amouyel, P., Arveiler, D., Bakker, S. J. L., Beilby, J., Bergman, R. N., Blangero, J., Brown, M. J., Burnier, M., Campbell, H., Chakravarti, A., Chines, P. S., Claudi-Boehm, S., Collins, F. S., Crawford, D. C., Danesh, J., de Faire, U., de Geus, E. J. C., Dörr, M., Erbel, R., Eriksson, J. G., Farrall, M., Ferrannini, E., Ferrières, J., Forouhi, N. G., Forrester, T., Franco, O. H., Gansevoort, R. T., Gieger, C., Gudnason, V., Haiman, C. A., Harris, T. B., Hattersley, A. T., Heliövaara, M., Hicks, A. A., Hingorani, A. D., Hoffmann, W., Hofman, A., Homuth, G., Humphries, S. E., Hyppönen, E., Illig, T., Jarvelin, M.-R., Johansen, B., Jousilahti, P., Jula, A. M., Kaprio, J., Kee, F., Keinanen-Kiukaanniemi, S. M., Kooner, J. S., Kooperberg, C., Kovacs, P., Kraja, A. T., Kumari, M.,

Kuulasmaa, K., Kuusisto, J., Lakka, T. A., Langenberg, C., Le Marchand, L., Lehtimäki, T., Lyssenko, V., Männistö, S., Marette, A., Matise, T. C., McKenzie, C. A., McKnight, B., Musk, A. W., Möhlenkamp, S., Morris, A. D., Nelis, M., Ohlsson, C., Oldehinkel, A. J., Ong, K. K., Palmer, L. J., Penninx, B. W., Peters, A., Pramstaller, P. P., Raitakari, O. T., Rankinen, T., Rao, D. C., Rice, T. K., Ridker, P. M., Ritchie, M. D., Rudan, I., Salomaa, V., Samani, N. J., Saramies, J., Sarzynski, M. A., Schwarz, P. E. H., Shuldiner, A. R., Staessen, J. A., Steinthorsdottir, V., Stolk, R. P., Strauch, K., Tönjes, A., Tremblay, A., Tremoli, E., Vohl, M.-C., Völker, U., Vollenweider, P., Wilson, J. F., Witteman, J. C., Adair, L. S., Bochud, M., Boehm, B. O., Bornstein, S. R., Bouchard, C., Cauchi, S., Caulfield, M. J., Chambers, J. C., Chasman, D. I., Cooper, R. S., Dedoussis, G., Ferrucci, L., Froguel, P., Grabe, H.-J., Hamsten, A., Hui, J., Hveem, K., Jöckel, K.-H., Kivimaki, M., Kuh, D., Laakso, M., Liu, Y., März, W., Munroe, P. B., Njølstad, I., Oostra, B. A., Palmer, C. N. A., Pedersen, N. L., Perola, M., Pérusse, L., Peters, U., Power, C., Quertermous, T., Rauramaa, R., Rivadeneira, F., Saaristo, T. E., Saleheen, D., Sinisalo, J., Eline Slagboom, P., Snieder, H., Spector, T. D., Thorsteinsdottir, U., Stumvoll, M., Tuomilehto, J., Uitterlinden, A., Uusitupa, M., van der Harst, P., Veronesi, G., Walker, M., Wareham, N. J., Watkins, H., Wichmann, H.-E., Abecasis, G. R., Assimes, T. L., Berndt, S. I., Boehnke, M., Borecki, I. B., Deloukas, P., Franke, L., Frayling, T. M., Groop, L. C., Hunter, D. J., Kaplan, R. C., O'Connell, J. R., Qi, L., Schlessinger, D., Strachan, D. P., Stefansson, K., van Duijn, C. M., Willer, C. J., Visscher, P. M., Yang, J., Hirschhorn, J. N., Carola Zillikens, M., McCarthy, M. I., Speliotes, E. K., North, K. E., Fox, C. S., Barroso, I., Franks, P. W., Ingelsson, E., Heid, I. M., Loos, R. J. F., Cupples, L. A., Morris, A. P., Lindgren, C. M. & Mohlke, K. L. (2015), 'New genetic loci link adipose and insulin biology to body fat distribution', *Nature* **518**.

URL: *https://doi.org/10.1038/nature14132*

The Gene Ontology Consortium (2018), 'The gene ontology resource: 20 years and still going strong', *Nucleic Acids Research* **47**(D1), D330–D338.

URL: *https://doi.org/10.1093/nar/gky1055*

Werling, D. M., Pochareddy, S., Choi, J., An, J.-Y., Sheppard, B., Peng, M., Li, Z., Dastmalchi, C., Santpere, G., Sousa, A. M. M., Tebbenkamp, A. T. N., Kaur, N., Gulden, F. O., Breen, M. S., Liang, L., Gilson, M. C., Zhao, X., Dong, S., Klei, L., Cicek, A. E., Buxbaum, J. D.,

Adle-Biassette, H., Thomas, J.-L., Aldinger, K. A., O'Day, D. R., Glass, I. A., Zaitlen, N. A., Talkowski, M. E., Roeder, K., State, M. W., Devlin, B., Sanders, S. J. & Sestan, N. (2019), 'Whole-genome and rna sequencing reveal variation and transcriptomic coordination in the developing human prefrontal cortex', *bioRxiv* .
URL: *https://www.biorxiv.org/content/early/2019/03/22/585430*

Willer, C. J., Li, Y. & Abecasis, G. R. (2010), 'Metal: fast and efficient meta-analysis of genomewide association scans', *Bioinformatics* **26**(17), 2190–2191.
URL: *http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2922887/*

Xiao, N., Wang, G. & Sun, L. (2018), *grex: Gene ID Mapping for Genotype-Tissue Expression (GTEx) Data*. R package version 1.8.
URL: *https://CRAN.R-project.org/package=grex*

Yengo, L., Sidorenko, J., Kemper, K. E., Zheng, Z., Wood, A. R., Weedon, M. N., Frayling, T. M., Hirschhorn, J., Yang, J., Visscher, P. M. & the GIANT Consortium (2018), 'Meta-analysis of genome-wide association studies for height and body mass index in 700000 individuals of european ancestry', *Human Molecular Genetics* **27**(20), 3641–3649.
URL: *https://doi.org/10.1093/hmg/ddy271*

Zhang, B. & Horvath, S. (2005), 'A general framework for weighted gene co-expression network analysis a general framework for weighted gene co-expression network analysis', *Statistical Applications in Genetics and Molecular Biology* **4**(1).

Zhang, M. J., Xia, F. & Zou, J. (2019), 'Fast and covariate-adaptive method amplifies detection power in large-scale multiple hypothesis testing', *Nature Communications* **10**(1), 3433.
URL: *https://doi.org/10.1038/s41467-019-11247-0*

Zhu, X. & Stephens, M. (2018), 'Large-scale genome-wide enrichment analyses identify new trait-associated genes and pathways across 31 human phenotypes', *Nature communications* **9**(1), 1–14.

Supporting Information Text

AdaPT conditional two-groups model

This section provides a more detailed explanation of updating the rejection threshold $s_t(x_i)$ in the AdaPT procedure, expanding on the description from *Methods* in the main manuscript. As in the main text, this is essentially an explanation of the EM approach of Lei & Fithian (2018). Note that for coherence some text is repeated from the main manuscript. Lei & Fithian (2018) use a conditional version of the classical two-groups model (Efron et al. 2001) yielding the conditional mixture density,

$$f(p|x) = \pi_1(x) f_1(p|x) + 1 - \pi_1(x), \tag{S1}$$

where the null p-values are modeled as uniform ($f_0(p|x) \equiv 1$). They proceed to use a *conservative* estimate for the conditional local false discovery rate, $\text{fdr}(p|x) = \hat{f}(1|x)/\hat{f}(p|x)$, by setting $1 - \pi_1(x) = f(1|x)$.

We model the non-null p-value density with a beta distribution density parametrized by μ_i,

$$f_1(p|x_i) = h(p; \mu_i) = \frac{1}{\mu_i} p^{1/\mu_i - 1}, \tag{S2}$$

where $\mu_i = \mathbb{E}[-\log(p_i)]$, resulting in a conditional density for a beta mixture model,

$$f(p|x_i) = \pi_1(x_i) \frac{1}{\mu_i} p^{1/\mu_i - 1} + 1 - \pi_1(x_i). \tag{S3}$$

In this form, we can model the non-null probability $\pi_1(x_i) = \mathbb{E}[H_i|x_i]$ and the effect size for non-null hypotheses $\mu(x_i) = \mathbb{E}[-\log(p_i)|x_i, H_i = 1]$ with two separate gradient boosted tree-based models. The XGBoost library (Chen & Guestrin 2016) provides logistic and Gamma regression implementations which we use for $\pi_1(x_i)$ and $\mu(x_i)$ respectively.

There are two categories of missing values in these regression problems: H_i is never observed, and at each step t of the search, the p-values for tests $\{i : p_i \leq s_t(x_i) \text{ or } p_i \geq 1 - s_t(x_i)\}$ are masked as $\tilde{p}_{t,i}$. An expectation-maximization (EM) algorithm can be used to estimate both $\hat{\pi}_1(x_i)$ and $\hat{\mu}(x_i)$ by maximizing the partially observed likelihood. The complete log-likelihood for the conditional two-groups model is,

$$l(\pi_1, \mu; p, H, x) = \sum_{i=1}^{n} \{H_i \log(\pi_1(x_i) + (1 - H_i)\log(1 - \pi_1(x_i))\} + \sum_{i=1}^{n} H_i \log\{h(p_i; \mu(x_i))\}. \tag{S4}$$

During the E-step of the $d = 0, 1, \ldots$ iteration of the EM algorithm, conditional on the partially observed data fixed at step t, $(x_i, \tilde{p}_{t,i})_{i\in[n]}$, we compute both,

$$\hat{H}_i^{(d)} = \mathbb{E}_{\hat{\pi}_1^{(d-1)},\hat{\mu}^{(d-1)}}[H_i \big| (x_i, \tilde{p}_{t,i})_{i\in[n]}] \tag{S5}$$

$$\hat{b}_i^{(d)} = \mathbb{E}_{\hat{\pi}_1^{(d-1)},\hat{\mu}^{(d-1)}}[\mathbf{1}(p'_{t,i} = p_i) \big| (x_i, \tilde{p}_{t,i})_{i\in[n]}, H_i = 1], \tag{S6}$$

where $\hat{b}_i^{(d)}$ indicates how likely $p'_{t,i} = \min(\tilde{p}_{t,i})$ equals p_i for non-null hypotheses. The explicit calculations of $\hat{H}_i^{(d)}$ and $\hat{b}_i^{(d)}$ for both the revealed, $\tilde{p}_{t,i} = p'_{t,i}$, and masked p-values, $\tilde{p}_{t,i} = \{p_i, 1-p_i\}$, are available in the supplementary materials of Lei & Fithian (2018).

The M-step consists of estimating $\hat{\pi}_1^{(d)}$ and $\hat{\mu}^{(d)}$ with separate gradient boosted trees, using *pseudo*-datasets to handle the partially masked data. In order to fit the model for $\pi_1(x_i)$, we construct the response vector $y_\pi^{(d)} = (1, \ldots, 1, 0, \ldots, 0) \in \mathbb{R}^{2n}$ and use weights $w_\pi^{(d)} = (\hat{H}_1^{(d)}, \ldots, \hat{H}_n^{(d)}, 1-\hat{H}_1^{(d)}, \ldots, 1-\hat{H}_n^{(d)}) \in \mathbb{R}^{2n}$. Then we estimate $\hat{\pi}_1^{(d)}(x_i)$ using the first n predictions from a classification model using $y_\pi^{(d)}$ as the response variable with the covariate matrix $(x_i)_{i\in[n]}$ replicated twice and weights $w_\pi^{(d)}$. Similarly, for estimating $\hat{\mu}^{(d)}(x_i)$ we construct a response vector $y_\mu^{(d)} = (-\log(p_1), \ldots, -\log(p_n), -\log(1-p_1), \ldots, -\log(1-p_n)) \in \mathbb{R}^{2n}$ with weights $w_\mu^{(d)} = (\hat{b}_1^{(d)}, \ldots, \hat{b}_n^{(d)}, 1-\hat{b}_1^{(d)}, \ldots, 1-\hat{b}_n^{(d)}) \in \mathbb{R}^{2n}$, and again take the first n predicted values using the duplicated covariate matrix.

The conditional local fdr is estimated for each $p'_{t,i}$,

$$\text{fdr}_{t,i} = \frac{\hat{\pi}_1(x_i)h(1; \hat{\mu}(x_i) + 1 - \hat{\pi}_1(x_i)}{\hat{\pi}_1(x_i)h(p'_{t,i}; \hat{\mu}(x_i) + 1 - \hat{\pi}_1(x_i)}, \tag{S7}$$

and we follow the procedure detailed in Section 4.3 of Lei & Fithian (2018) to update the rejection threshold to $s_{t+1}(x_i)$ by removing test $i^* = \arg\max_{i\in\mathcal{R}_t} \text{fdr}_{t,i}$ from $\mathcal{R}_t$. A summary diagram of the EM algorithm is displayed in Figure S1.

SCZ results with independent loci

One potential concern regarding the assessment of performance of AdaPT is the impact of linkage disequilibrium (LD). In the Manhattan plots of Figures 2(B-C), the discoveries visually appear to be located close to one another. However, the visual appearance of genomic positions is somewhat misleading because our initial selection of eSNPs greatly reduces the number of SNPs commonly

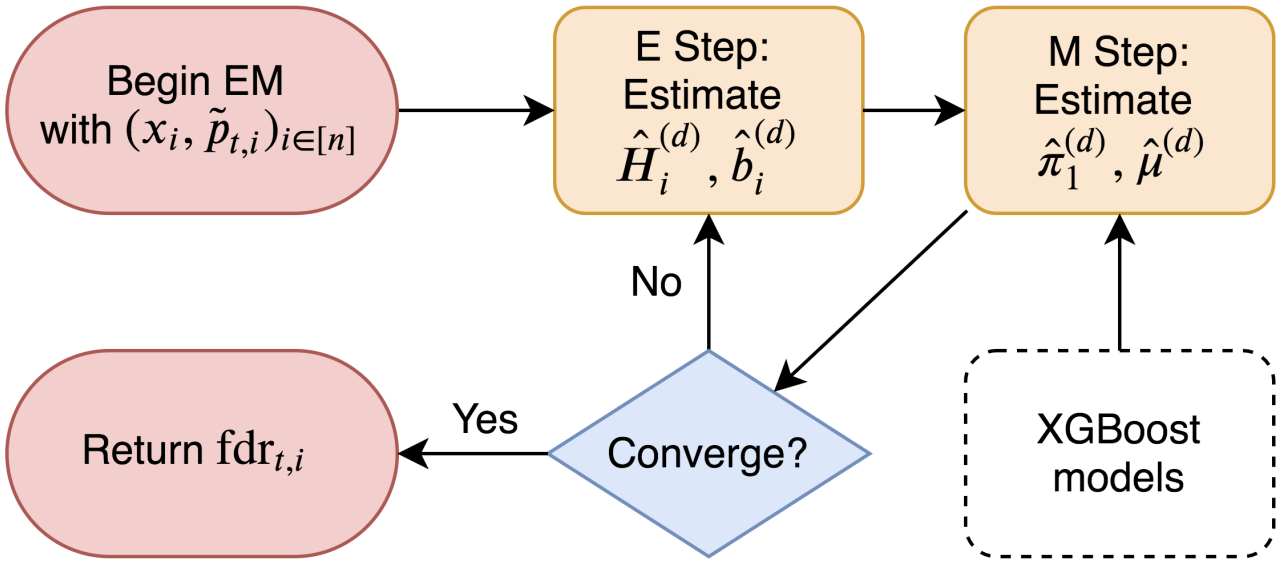

Figure S1: Summary of AdaPT EM algorithm.

portrayed in Manhattan plots many of these SNPs are not very close to each other in the genome and not in high LD, although the format of the Manhattan plot makes this feature hard to see. To take this analysis further we follow common practice for GWAS results by identifying the "best" or "lead" SNPs in a LD block/cluster, using a similar approach as Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014), for each of the set of discoveries presented in Figure 2:

1. order the SNPs by the AdaPT -$\log_{10}$(q-value) in descending order,
2. starting with the SNP with the largest value for the AdaPT -$\log_{10}$(q-value),
   - remove all SNPs with $r^2 \geq 0.1$ within a 500kb window,
   - move on to next SNP that is still remaining,
3. return the retained SNPs as the LD-independent SNPs in low LD ($r^2 < 0.1$). (Remark: this approach excludes SNPs whose contribution to the GWAS signal is partially independent of the lead SNP, but it has the advantage of simplicity.)

We use the reference European sample genotype data from the 1000 Genomes project (1000 Genomes Project Consortium and others 2012) to compute the r^2 values between SNPs. In the GWAS setting this LD clumping procedure is typically applied to the reported SNP p-values, but

because the ordering of SNPs varies between the different sets of discoveries (intercept-only versus use of covariates) we perform the operation separately with their respective q-values. For each of the different set of covariates considered, this results in reducing the 25,076 selected eSNPs down to the following number of "independent loci":

- Intercept-only: 3,958
- BD z-stats: 3,966
- BD z-stats + eQTL slopes: 3,962
- BD z-stats + eQTL slopes + WGCNA (w/ interactions): 3,963
- BD z-stats + eQTL slopes + WGCNA (w/o interactions): 3,959
- WGCNA: 3,954

The differences in counts are due to the different number of ties that take place between the resulting q-values for each considered set of covariates. Next, for the identified set of "lead" SNPs we observe how many have q-values less than the target FDR level $\alpha = 0.05$ (i.e. associations detected at $\alpha = 0.05$). The results are displayed in the Figure S2, including Manhattan plots Figures S2(A-B) of the q-values for the AdaPT intercept-only and BD z-stats + eQTL slopes + WGCNA (w/ interactions) results, rather than using the actual p-values. The lead SNPs in each of the Manhattan plots are denoted by an X shape. In conjunction with Figures S2(C-D), the relative improvement in the set of independent loci within the discovery sets from AdaPT is analogous to the results presented in Figure 2, emphasizing the advantage of accounting for covariates and their interactions via gradient boosted trees. Additionally, Figure S3 further emphasizes that the improvement in power is not restricted to a particular section of the genome. As seen in Figure S4, we observe a similar improvement in the number of independent loci when ordering the SNPs with the observed 2014-only studies SCZ p-values.

While we maintain FDR control on the original set of discoveries (see Figure 3 in *Results*), we do not retain any guarantees regarding the detected independent loci presented in Figure S2. In order to maintain FDR control on the set of discovered independent loci, an alternative approach or adjustment to the AdaPT algorithm is required. A simple alternative is to first apply LD pruning/clumping as initial step prior to applying AdaPT to a reduced set of lead SNPs. However,

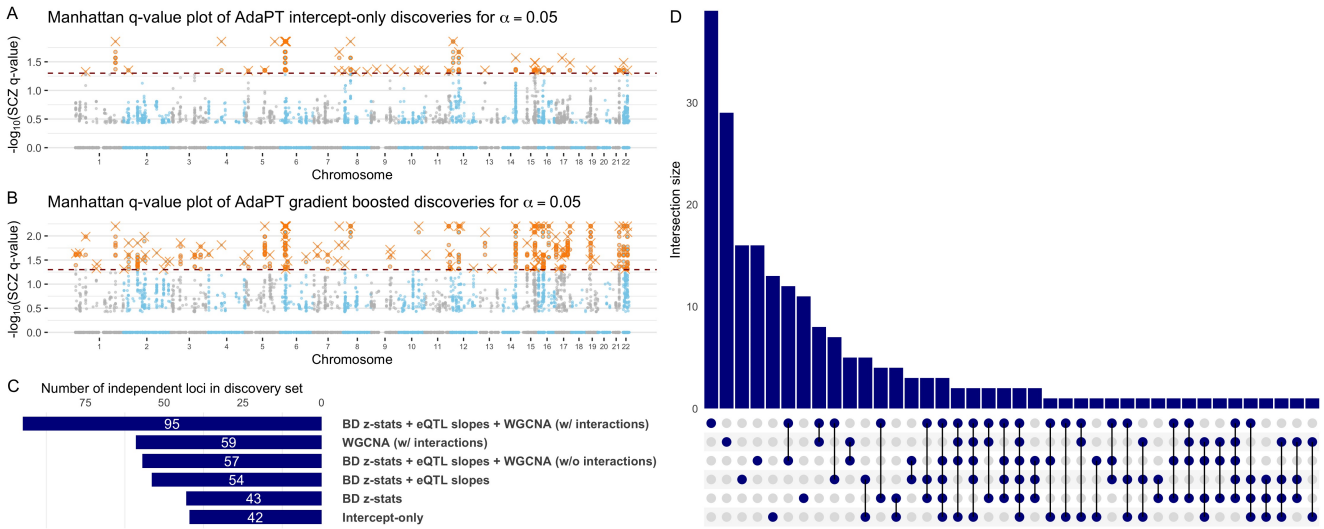

Figure S2: Manhattan q-value plots of SCZ AdaPT discoveries (orange) using *(A)* intercept-only model compared to *(B)* covariate informed model at target $\alpha = 0.05$, with lead SNPs for independent loci denoted by Xs. *(C)* Comparison of the number of independent loci for each discovery set at target $\alpha = 0.05$ based on LD pruning with the respective AdaPT q-values and *(D)* their resulting discovery set intersections.

this encounters the challenge of defining lead SNPs without data "snooping" based on using the observed p-values. Future work will explore modifications for AdaPT, potentially exploring recent developments (Ren & Candès 2020), to maintain FDR control on an independent subset of SNPs.

SCZ variable importance and partial dependence

We explore further the variable relationships from the gradient boosted trees. First, Figure S5 displays the change in variable importance for the non-null effect size (μ) at each model fitting iteration, with the top variables in the final model highlighted. The variable importance measures are relatively stable across all model iterations with the BD z-statistics and eQTL slope measures maintaining the highest level of importance. Figure S6 displays the partial-dependence plot at each AdaPT model fitting iteration for the estimated marginal relationship between the BD z-statistics and the non-null effect size μ, evaluated at the $0, 2.5\%, 5\%, \ldots, 100\%$ percentiles. The estimates reveal an increasing effect size as the BD z-statistics grow in magnitude, which is relatively stable across the model iterations. Figures S7(A-C) display the relationships for the probability of non-

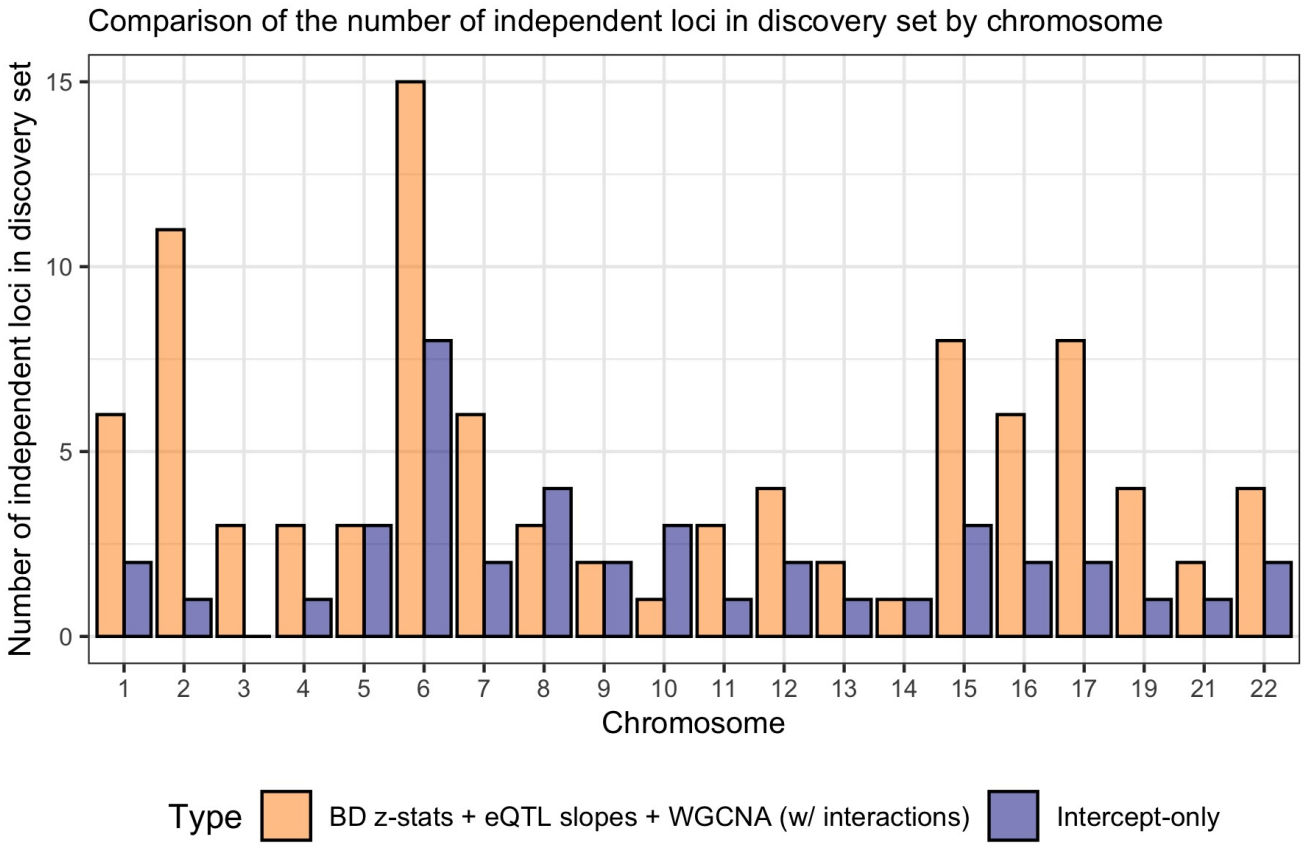

Figure S3: Comparison of the number of independent loci in the AdaPT discovery sets by type for each chromosome.

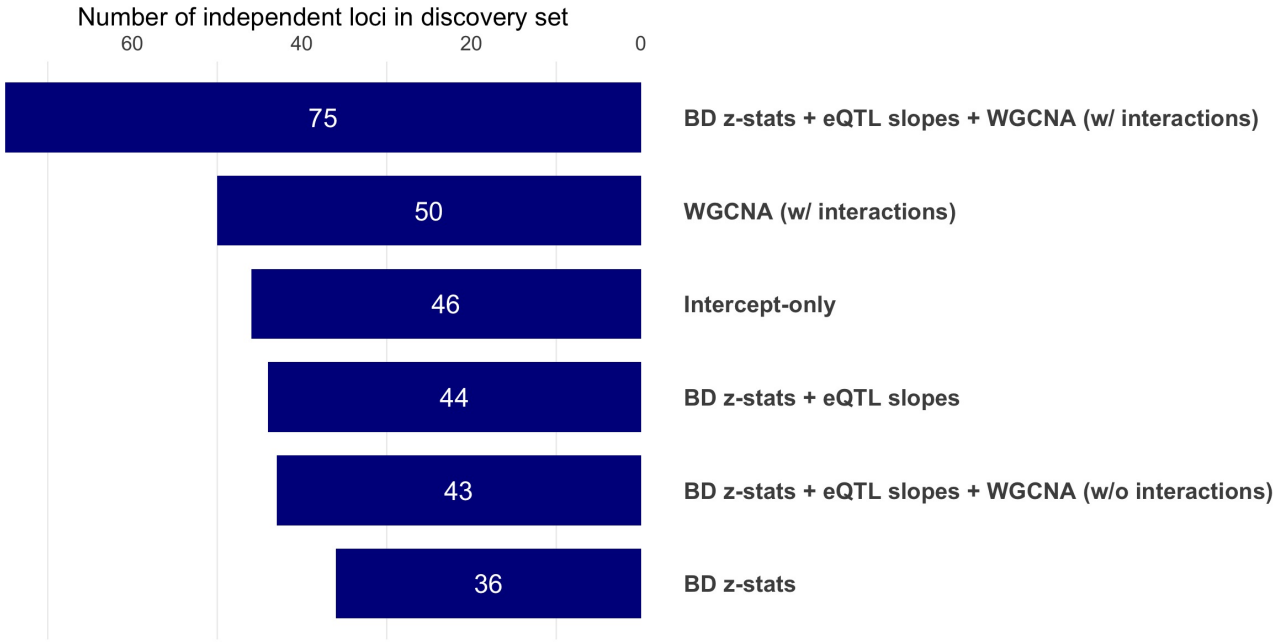

Figure S4: Comparison of the number of independent loci for each discovery set at target $\alpha = 0.05$, based on LD pruning with the with *2014-only* SCZ p-values.

null model, while (D-F) display relationships for the effect size under the alternative. Although the partial dependence plots show considerable variability due to the high dimensional of the model, we can still see general trends consistent with the variable importance plots from Figure 3(A) and Figure S5.

In Figure S8 we display the p-value distributions comparing the enrichment for membership in the different WGCNA modules reported by Werling et al. (2019). While many of the WGCNA modules lack clear evidence or contain too few eSNPs, as denoted by their respective y-axes, the *cyan* and *salmon* modules display noticeable enrichment. Additionally, as mentioned previously, membership in the *gray* module displays a lack of enrichment versus no associated cis-eQTL gene affiliated with the unassigned WGCNA module.

As additional context for the improved performance from using all covariates with interactions, Figures S9(A-B) display the change in partial dependence between the BD z-statistics and probability of being non-null π_1 across the AdaPT search for the AdaPT results using (A) BD z-statistics only and (B) all covariates without interactions. When compared to the results using all covariates with interactions in Figure Figure 3(B), we see that both versions of these results

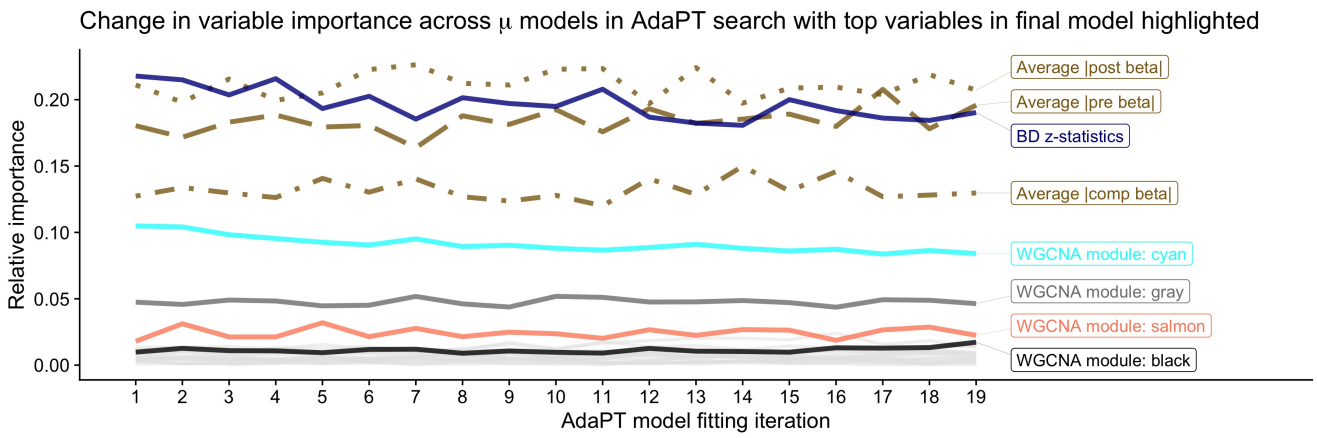

Figure S5: Change in variable importance for AdaPT non-null effect size μ model across search, with top variables in final model highlighted.

display relatively flat relationships near the end of the AdaPT search. This provides evidence of the importance of the interactions between other covariates and the BD z-statistics in retaining discriminatory power of the eSNPs near the end of the AdaPT search.

Replication simulations

We use simulations to empirically assess the observed nominal replication rate, percentage of discoveries with p-values less than 0.05 in holdout *2018-only* studies, of 55.2% for the 843 SCZ discoveries from the *2014-only* studies at target FDR level $\alpha = 0.05$. We use the final non-null effect size model returned by the AdaPT, $\hat{\mu}^*$, to generate simulated p-values $\boldsymbol{p}^{sim}$ and nominal replication rates to compare the observed rate against. For the simulations, we assume that all 843 SCZ discoveries from the *2014-only* studies are truly non-null, and we use the actual eSNPs, their observed standard errors σ_{14}, σ_{18} from the *2014-only* and *2018-only* studies respectively, as well as their actual covariates for generating $\boldsymbol{p}^{sim}$. A single iteration of the simulation proceeds as follows:

- For each of the $R_{\text{SCZ}} = 843$ discoveries $i \in \mathcal{R}_{\text{SCZ}}$:
    1. Assume test status is non-null: $H_i = 1$.

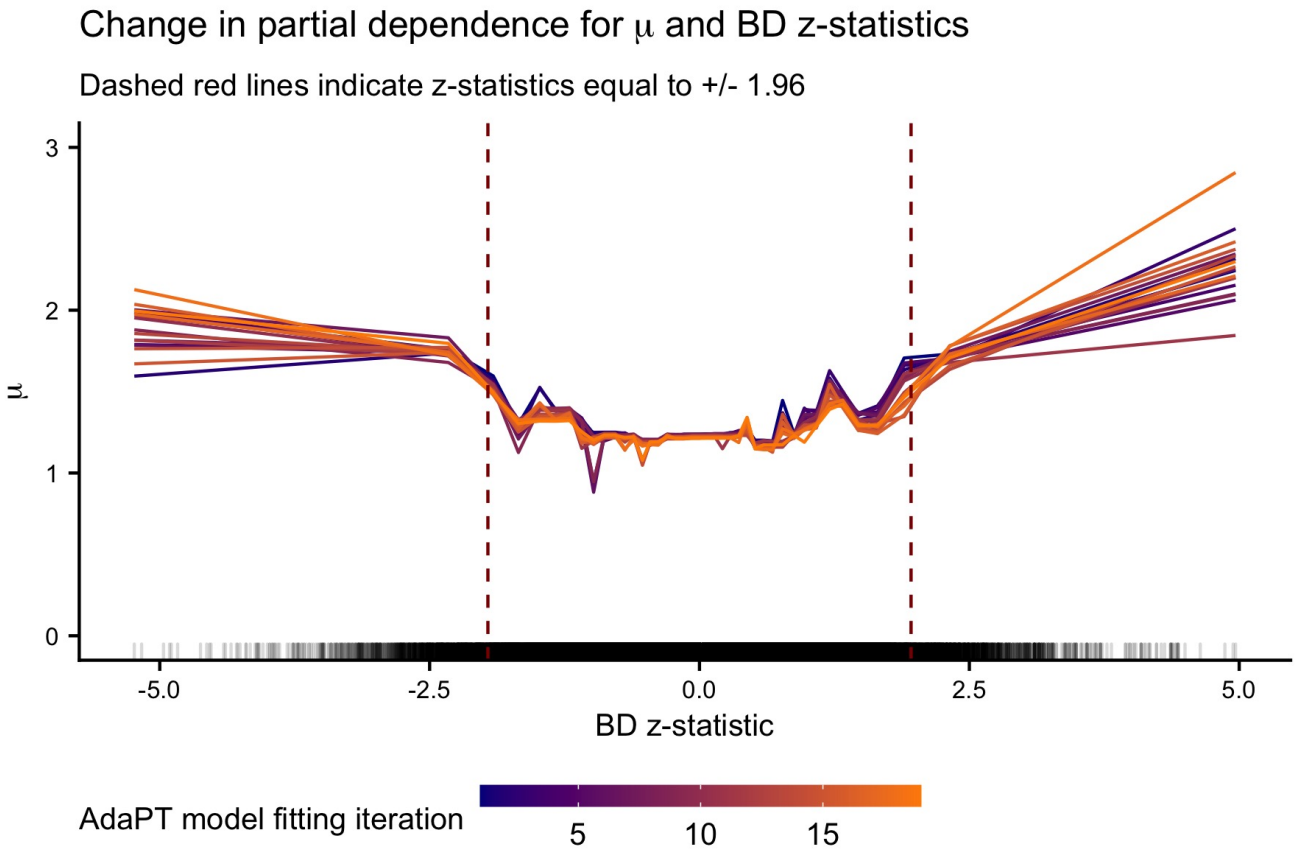

Figure S6: Change in partial dependence for non-null effect size μ and BD z-statistics across μ models in AdaPT search.

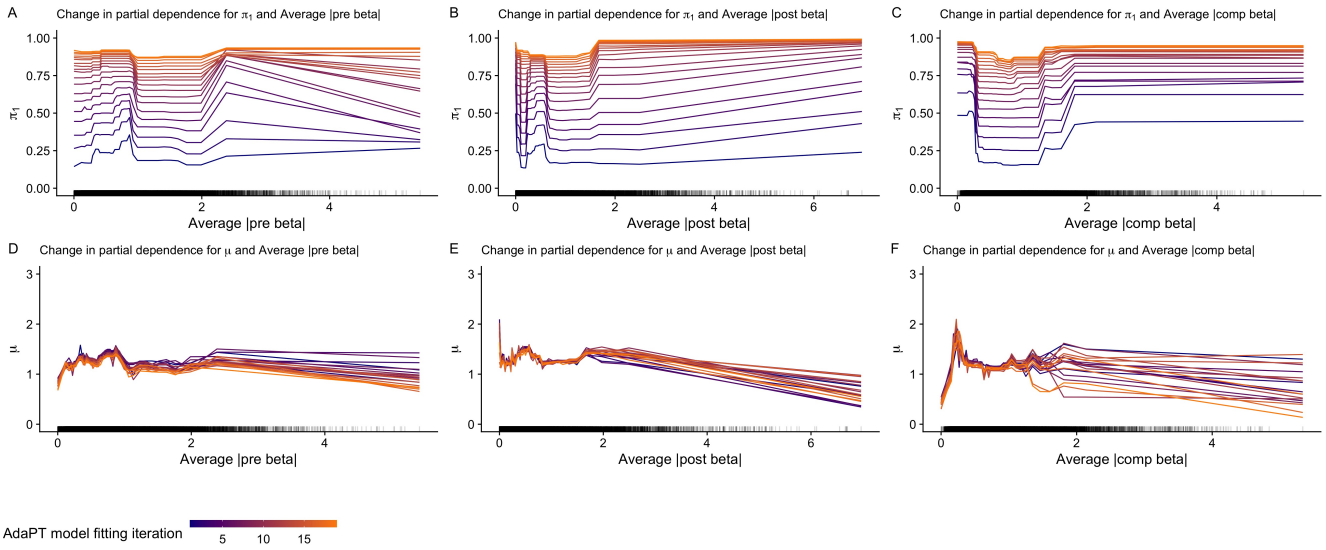

Figure S7: Change in partial dependence plots for probability of being non-null π_1 in *(A-C)*, and the effect size under alternative μ in *(D-F)*, for each type of eQTL slope. Rugs along x-axis denote distribution of values for each variable.

2. Generate effect size using final AdaPT model as truth:

$$-\log p_i^{sim} | x_i^{\text{SCZ}} \sim \text{Exp}\big(1/\hat{\mu}^*(x_i^{\text{SCZ}})\big). \tag{S8}$$

3. Transform effect sizes to p-value p_i^{sim}.
4. Convert simulated p-value to z-statistic $z_i^{sim} = \big|\Phi^{-1}(p_i^{sim}/2)\big|$.
5. Calculate updated z-statistic to reflect observed reduction in standard error for *2018-only* studies relative to *2014-only*,

$$z_i^{*,sim} = z_i^{sim} \cdot \frac{\sigma_{14}}{\sigma_{18}}. \tag{S9}$$

6. Convert updated z-statistic to p-value:

$$p_i^{*,sim} = 2 \cdot \Phi(-|z_i^{*,sim}|). \tag{S10}$$

- Calculate nominal replication rate using $\boldsymbol{p}^{sim} = (p_i^{*,sim}, \ldots, p_{R_{\text{SCZ}}}^{*,sim})$,

$$\text{Nominal replication rate } = \frac{|\{i : p_i^{*,sim} \leq .05\}|}{R_{\text{SCZ}}}. \tag{S11}$$

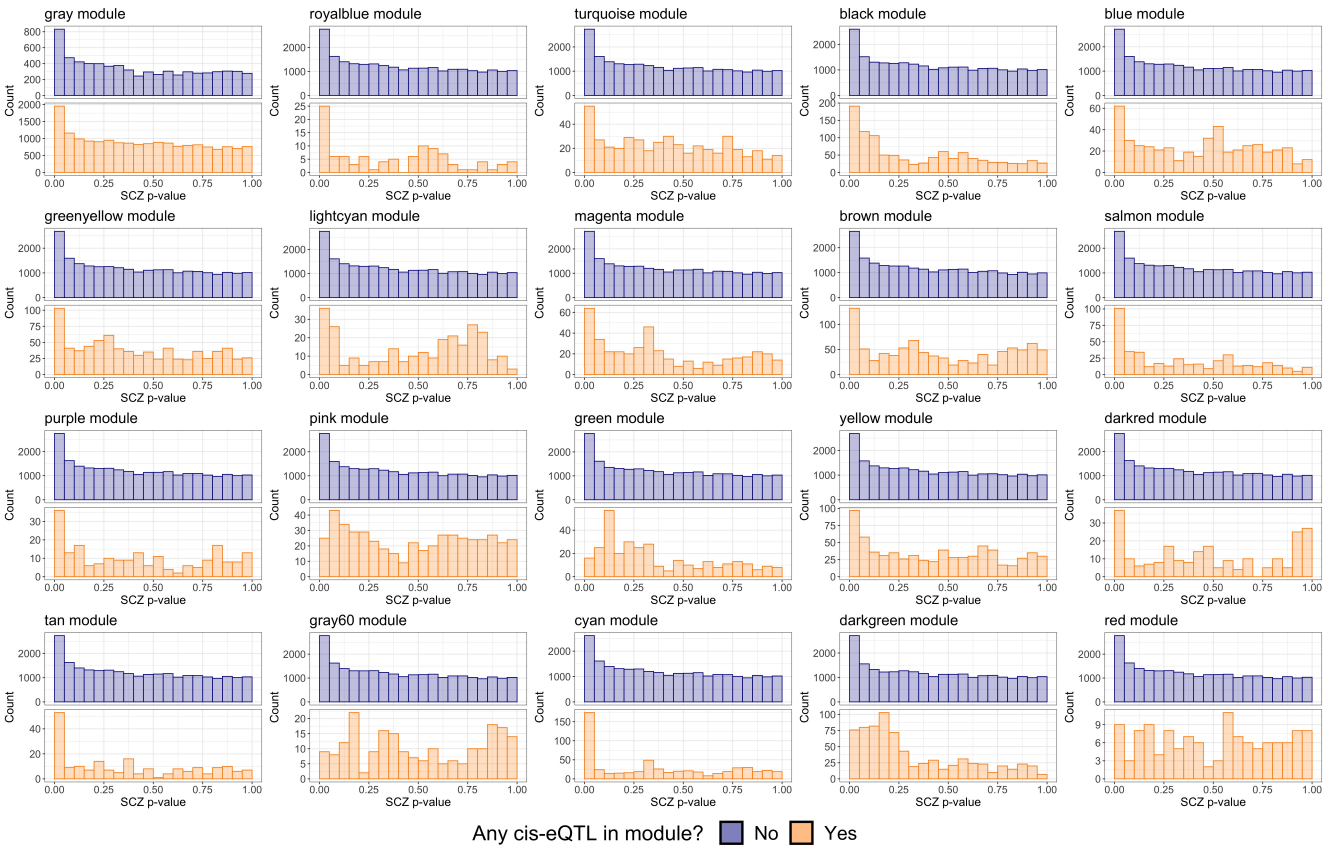

Figure S8: Comparison of SCZ p-value distributions from *2014* studies by whether or not the eSNP had an associated cis-eQTL gene in the module.

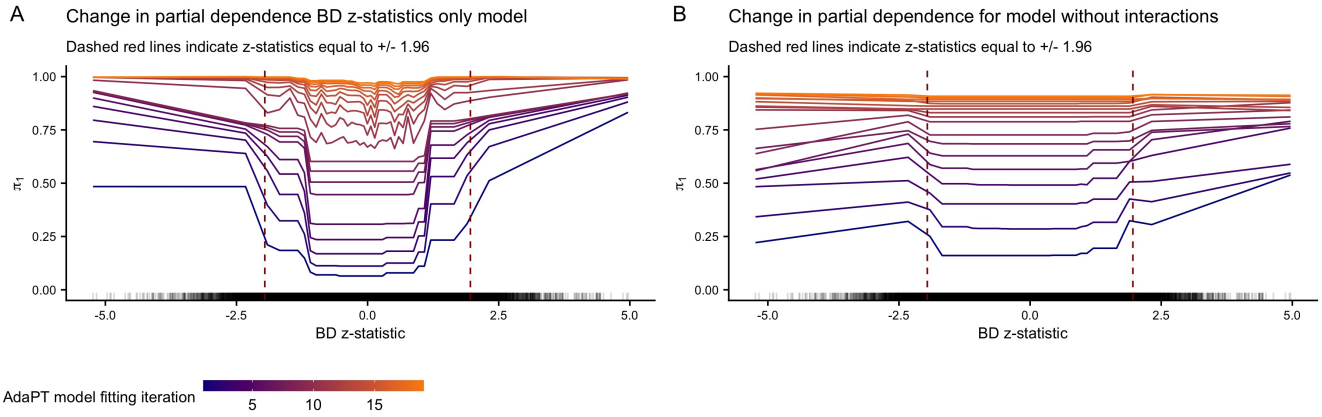

Figure S9: Change in partial dependence for BD z-statistics and probability of being non-null π_1 for the AdaPT results using *(A)* only BD z-statistics and *(B)* all covariates without any interactions.

We repeat this process to generate ten-thousand simulated values for the nominal replication rate. The distribution of the simulated values ranges from approximately 51% to 63%, with an average and median of $\approx 57\%$, close to the observed rate of 55.2%. Obviously, assuming that all of the 843 rejections are truly non-null is an overtly optimistic assumption given the use of FDR error control. Thus, the average simulated nominal replication rate of $\approx 56.6\%$ is reassuringly close to the observed rate and likely higher than what would be expected if false discoveries were accounted for among the 843 considered eSNPs.

SCZ results with *all 2018* studies

We generate the AdaPT results using the SCZ p-values from *all-2018* studies to the same set of $n_{\text{SCZ}} = 25,076$ eSNPs with the same covariates x_i^{SCZ}. As a comparison to the results displayed in Figure 2 using the *2014-only* studies, Figures S10(A-D) display the same figures but with the results from *all 2018* at target FDR level $\alpha = 0.05$. In contrast to before, we see that due to the increase in power from the study size, the use of modeling the auxiliary information provides a much smaller increase in power with just an approximately 19% increase in discoveries from the intercept-only results (1,865 discoveries) to using all twenty-four covariates with interactions (2,228 discoveries).

For comparison, we additionally examine the change in variable importance and partial dependence plots returned by AdaPT using *all 2018* studies. Similar to before, Figures S11(A-B) display the change in variable importance plots for both the probability of being non-null π_1 and effect size under alternative μ models using the SCZ p-values from *all 2018* studies respectively. The results are similar to before, but with the complete sample eQTL slopes possessing the highest importance. The BD z-statistics are again highly important for *all 2018* studies, displaying the similarly increasing relationships across the AdaPT models as seen in the partial dependence plots in Figures S12(C-D). The change in partial dependence plots for the different eQTL slopes summaries are seen in Figures S13(A-F). Figure S14 displays the levels of SCZ enrichment for *all 2018* studies, revealing modules that are consistent with the *2014-only* studies such as *cyan* and *salmon*.

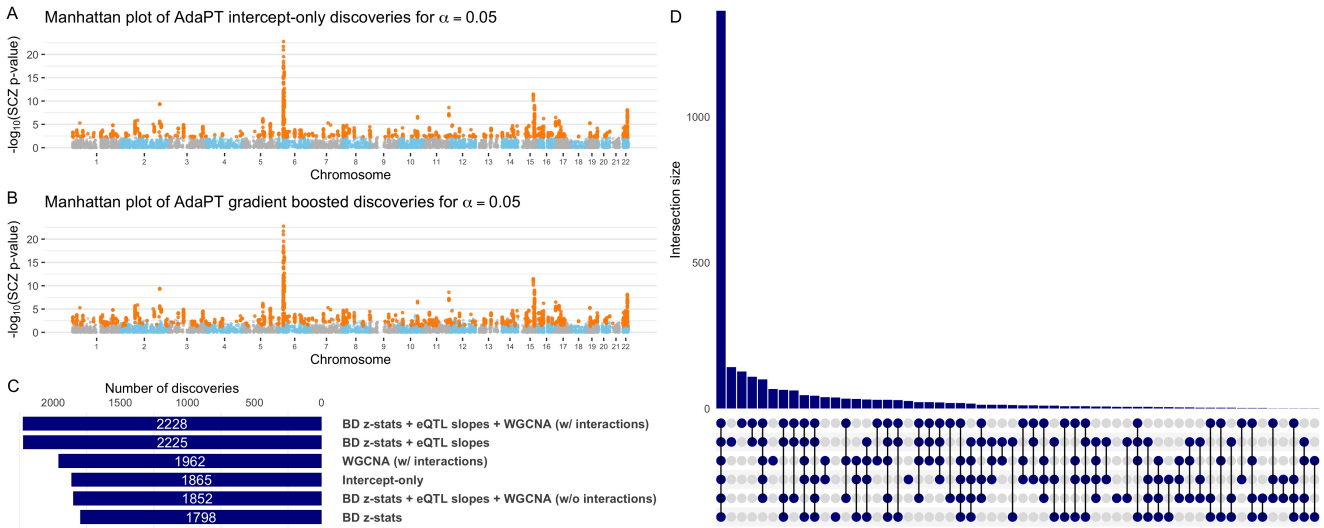

Figure S10: Manhattan plots of SCZ AdaPT discoveries (in orange) with *all 2018* studies using *(A)* intercept-only model compared to *(B)* covariate informed model at target $\alpha = 0.05$. *(C)* Comparison of the number of discoveries at target $\alpha = 0.05$ for AdaPT with varying levels of covariates and *(D)* their resulting discovery set intersections.

Type 2 diabetes results

Using GWAS summary statistics for type 2 diabetes (T2D), unadjusted for BMI, available from Diabetes Genetics Replication And Meta-analysis (DIAGRAM) consortium (Mahajan et al. 2018), we applied our full pipeline outlined in Figure 1. Of the initial set of over twenty-three million SNPs available, we identified 176,246 eSNPs from eQTL variant-gene pairs from any GTEx tissue sample using the definition of the GTEx eSNPs explained in *Data*. Figure S15 displays the enrichment for these GTEx eSNPs compared to the original set of SNPs from the T2D GWAS results.

We create a vector of covariates x_i^{T2D} summarizing expression level information from GTEx for pancreas, liver, and two adipose tissues, *subcutaneous* and *visceral (omentum)*. Specifically, we calculate $\tilde{\beta}_i^{r^{\text{T2D}}}$ for each r^{T2D} in the set of tissues: pancreas, liver, adipose - subcutaneous, adipose - visceral (omentum). Additionally, we generate WGCNA module assignments using protein coding genes for pancreas samples from GTEx. To generate the WGCNA results, we only consider protein coding genes identified using the `grex` package in `R` (Xiao et al. 2018, R Core Team 2018). Additionally, all genes with expression levels of zero for over half of the provided samples were removed. This resulted in fourteen different module, including the unassigned *gray*

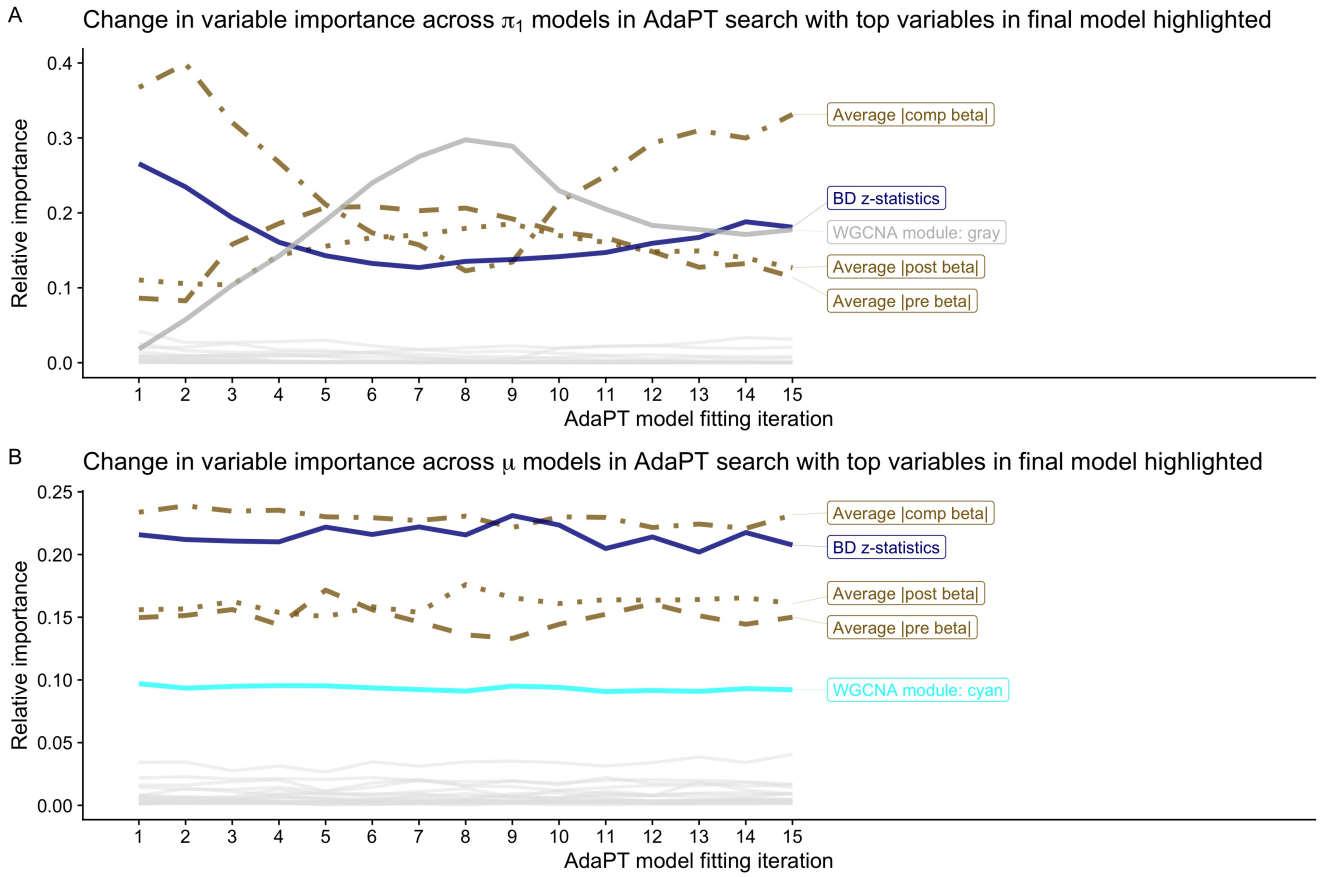

Figure S11: Using *all 2018* studies: change in variable importance for AdaPT *(A)* probability of being non-null π_1 and *(B)* effect size under alternative μ models across search, with top variables in final model highlighted.

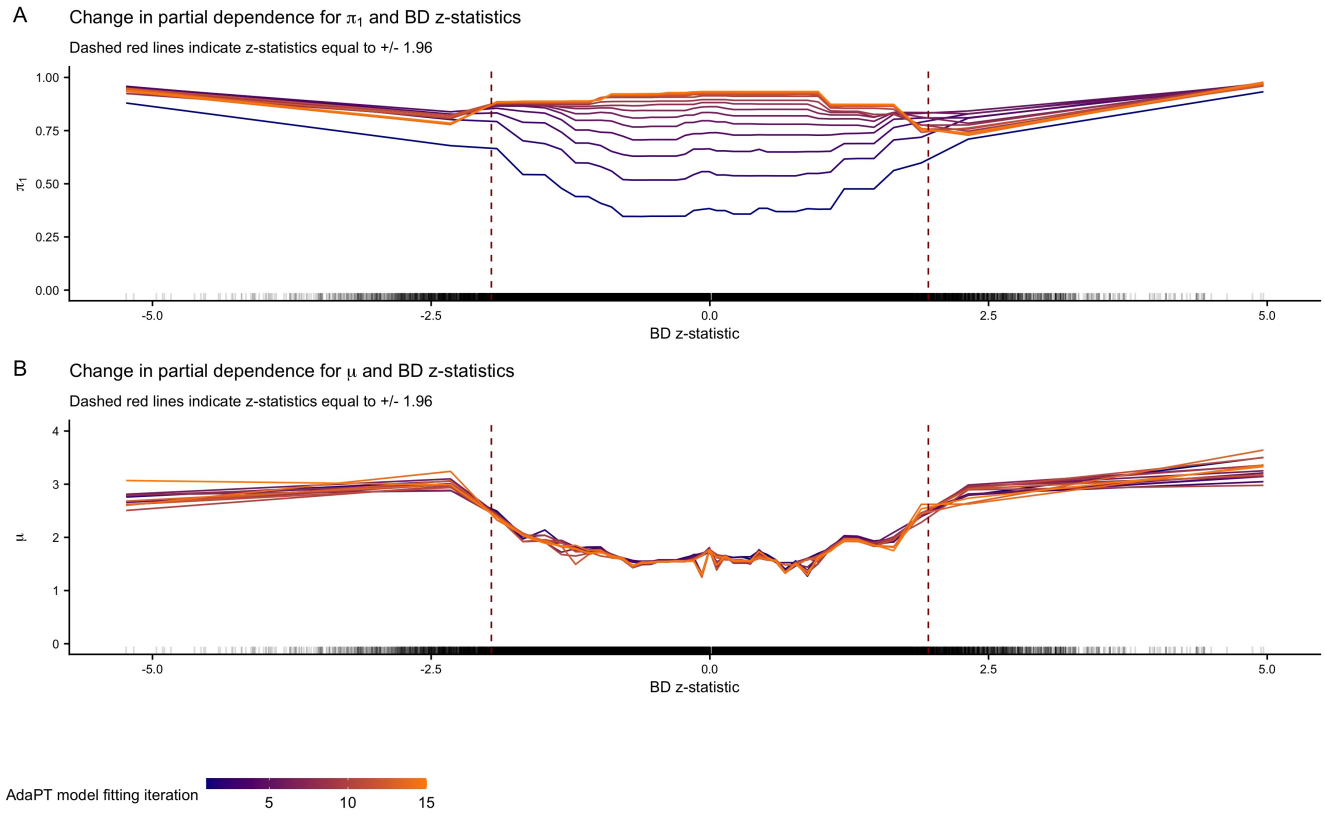

Figure S12: Using *all 2018* studies: change in partial dependence for BD z-statistics and AdaPT *(A)* probability of being non-null π_1 and *(B)* effect size under alternative μ models across search.

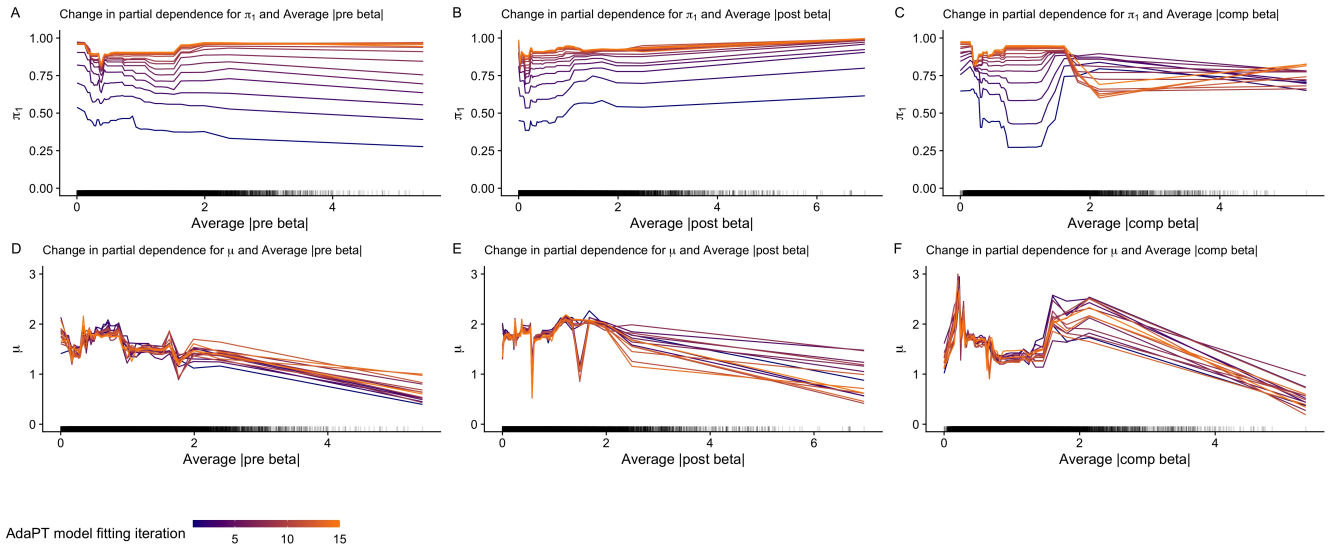

Figure S13: Using *all 2018* studies: change in partial dependence plots for probability of being non-null π_1 in *(A-C)*, and the effect size under alternative μ in *(D-F)*, for each type of BrainVar eQTL slope. Rugs along x-axis denote distribution of values for each variable.

module. Unlike the SCZ application, we do not use independent GWAS results from another phenotype.

Using x_i^{T2D} defined above, we applied AdaPT to the 176,246 GTEx eSNPs. However, we encountered an issue for this data where we were unable to discover any hypotheses at target FDR level $\alpha \leq 0.05$. This was due to the fact that 640 eSNPs had p-values *exactly* equal to one. While this can understandably occur with publicly available GWAS summary statistics, p-values equal to one will then *always* contribute to the *pseudo*-estimate for the number of false discoveries A_t during the AdaPT search (see *Methodology overview*). With a relatively high number of p-values equal to one, AdaPT is unable to search through rejection sets for lower α values. To overcome this challenge, we draw random replacement p-values for the 640 eSNPs from a uniform distribution between 0.97 and $1 - 1\text{E}^{-15}$, a value strictly less than one, to allow some leeway. We refer to this set of p-values as *adjusted*, while the original observed p-values are *unadjusted*. For comparison, Figure S16 shows the difference in the number of discoveries for the *adjusted* and *unadjusted* p-values across different target α values. Due to the similarity in performance for α values greater than 0.1, we use results for the *adjusted* p-values moving forward.

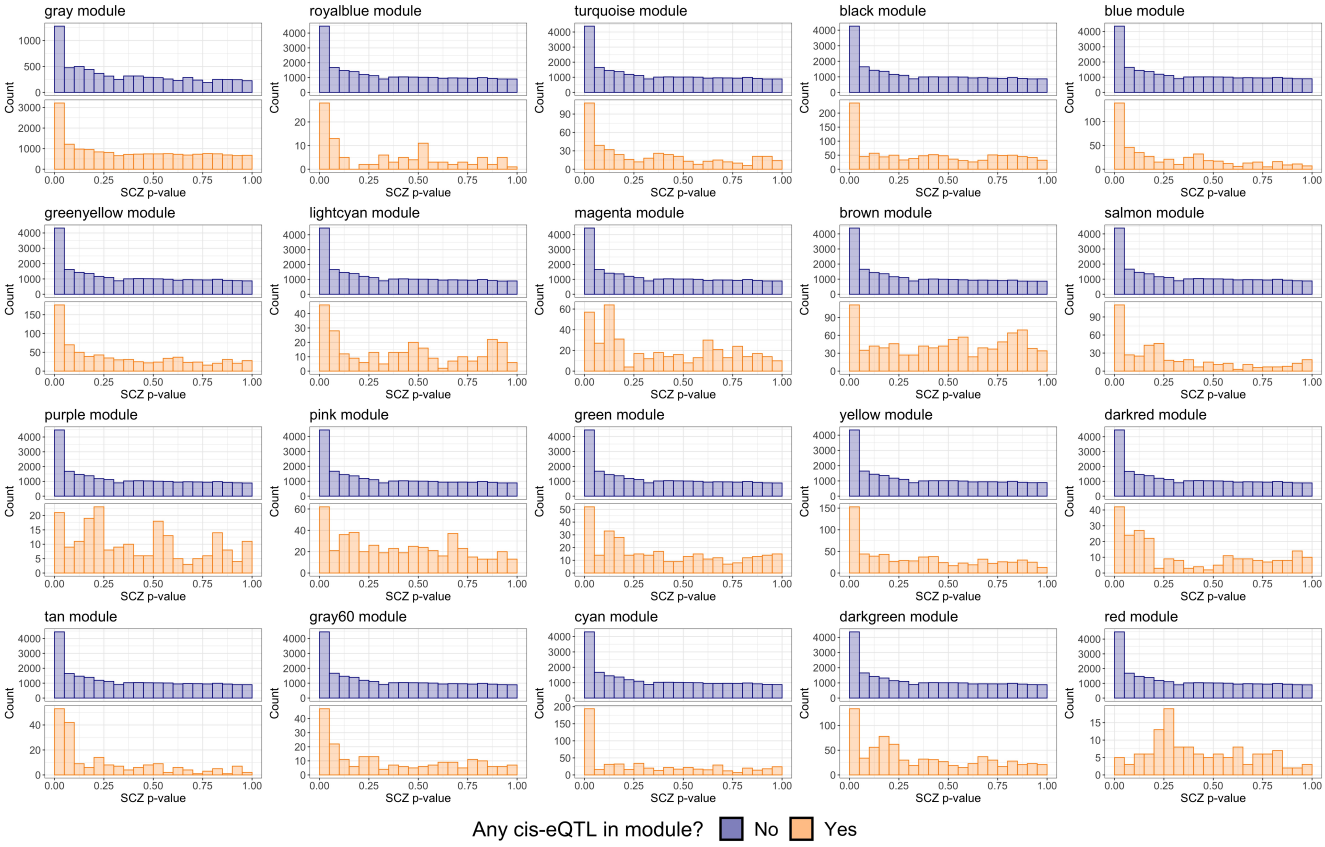

Figure S14: Using *all 2018* studies: comparison of SCZ p-value distributions from *2014* studies by whether or not the eSNP had an associated cis-eQTL gene in the module.

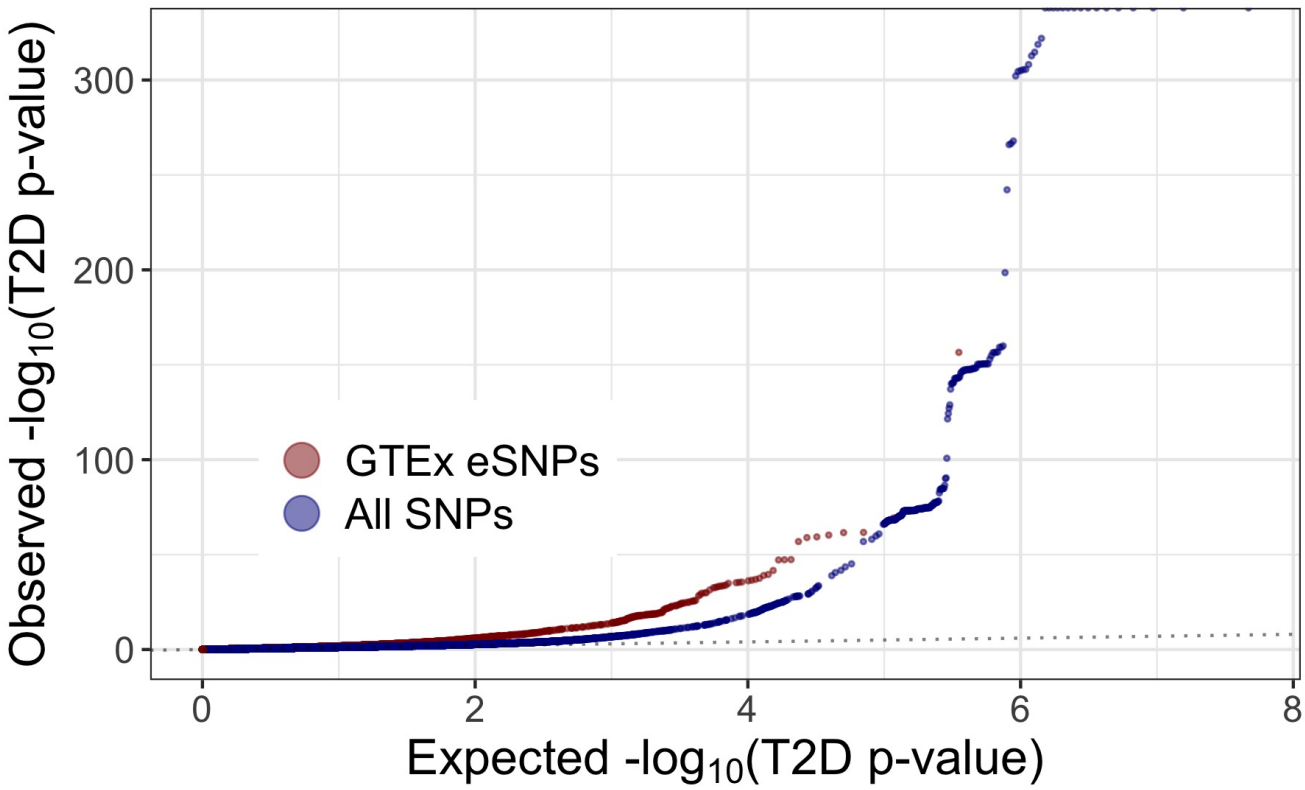

Figure S15: A comparison of qq-plots revealing T2D enrichment for GTEx eSNPs compared to full set of SNPs.

At target FDR level $\alpha = 0.05$, AdaPT yields 14,920 T2D discoveries using the *adjusted* p-values with covariates x_i^{T2D} (compared to 14,693 intercept-only discoveries). The change in variable importance for the T2D AdaPT models are displayed in Figure S17. This set of eSNPs is associated with 5,970 cis-eQTL genes for which we then applied gene ontology enrichment analysis to (Ashburner et al. 2000, The Gene Ontology Consortium 2018), identifying the gene enrichment for biological processes displayed in Figure 5.

BMI results

We also applied our pipeline of analysis to BMI, unadjusted for waist-to-hip ratio (WHR), using GWAS results for individuals of European ancestry available from the GIANT Consortium. Specifically, we approached BMI in the same manner as SCZ: apply AdaPT to GWAS results from earlier studies with a sample size of 322,154 individuals (Locke et al. 2015); then compare

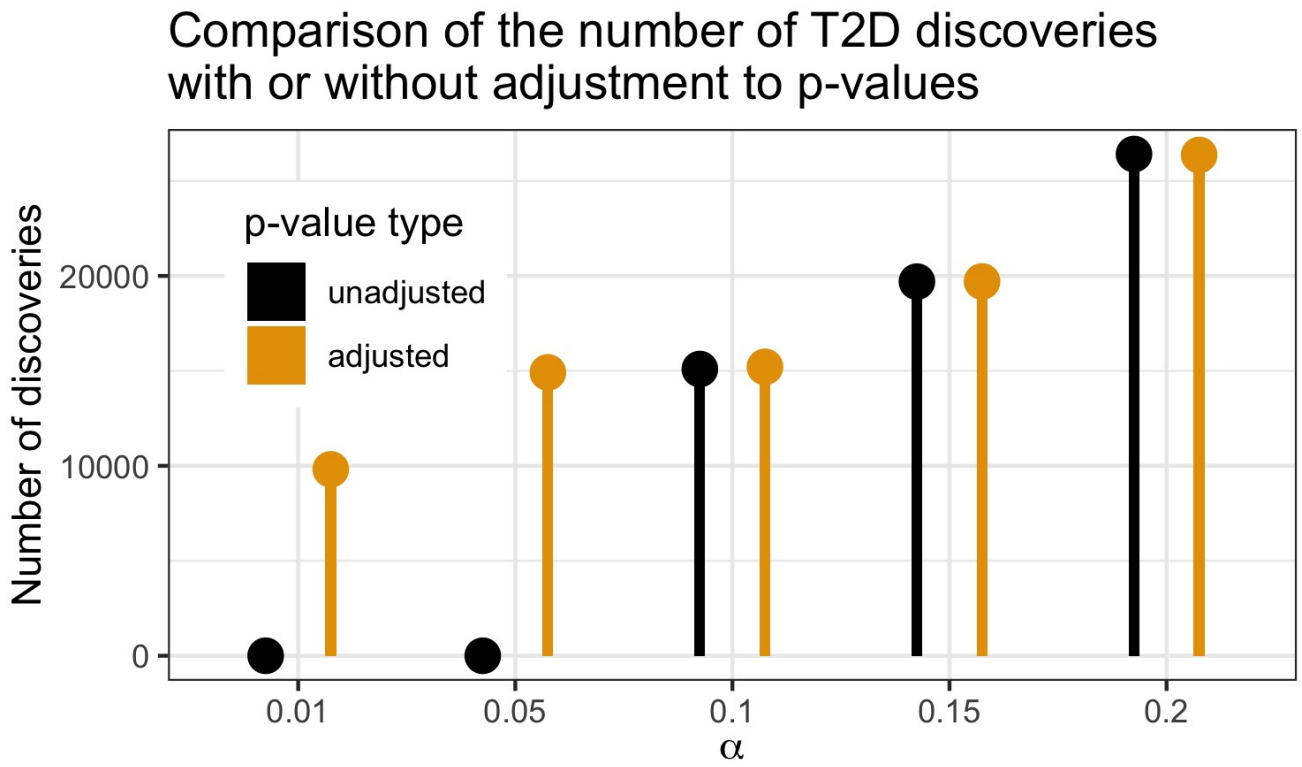

Figure S16: Comparison of the number of discoveries by AdaPT for T2D by whether or not the adjusted or unadjusted p-values were used.

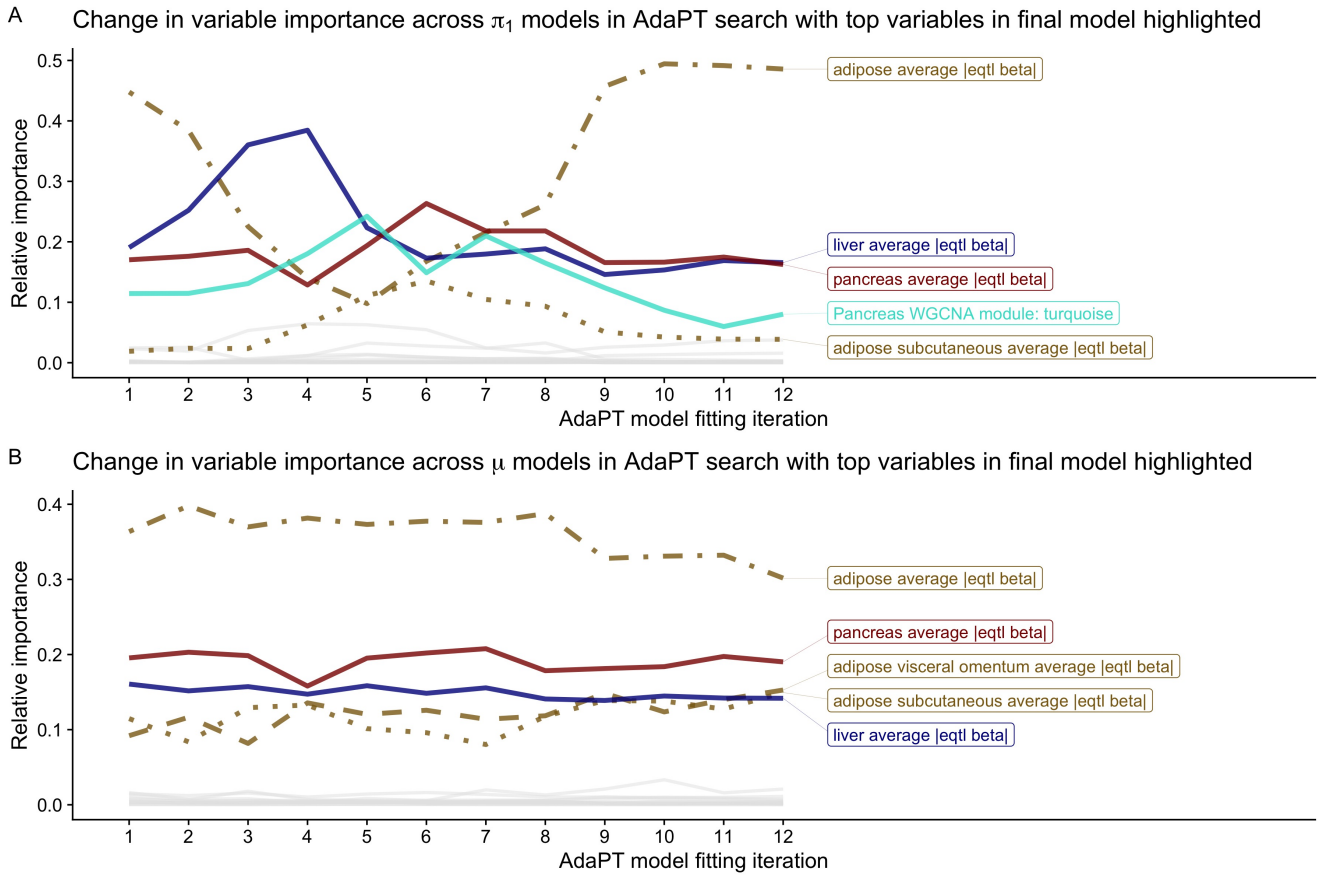

Figure S17: Change in T2D variable importance for AdaPT *(A)* probability of being non-null π_1 and *(B)* effect size under alternative μ models across search, with top variables in final model highlighted.

the nominal replication results on recently conducted studies with a sample size of approximately 700,000 individuals (Yengo et al. 2018). As before, all of the *2015-only* studies from Locke et al. (2015) were included as a subset of *all 2018* studies Yengo et al. (2018). Because both Locke et al. (2015) and Yengo et al. (2018) use the inverse variance-weighted fixed effects approach for meta-analysis, we then compute statistics for the studies exclusive to *2018-only* studies in Yengo et al. (2018). Additionally, to make this example more comparable to the SCZ use, we also use GWAS results for WHR (Shungin et al. 2015) as a covariate (analogous to BD for SCZ). Following pre-processing steps (matching SNPs across studies and effect alleles in both WHR and BMI), we identified 47,690 GTEx eSNPs from a set of nearly two million SNPs, based on the definition explained in *Data*. Figure S18 displays the enrichment for the GTEx eSNPs compared to the original set of pre-processed SNPs for the *2015-only* studies.

Based on previous knowledge of BMI tissue expression associations (Locke et al. 2015), we create a vector of covariates x_i^{BMI} summarizing expression level information from GTEx for brain and adipose tissues (both *subcutaneous* and *visceral (omentum)*). Specifically, we calculate $\tilde{\beta}_i^{r^{\text{BMI}}}$ for each $r^{\text{BMI}} \in \{$GTEx brain tissues, adipose - subcutaneous, adipose - visceral (omentum)$\}$, where we consider the following brain tissues: (1) *amygdala*, (2) *anterior cingulate cortex BA24*, (3) *caudate basal ganglia*, (4) *cerebellar hemisphere*, (5) *frontal cortex BA9*, (6) *hippocampus*, (7) *hypothalamus*, (8) *nucleus accumbens basal ganglia*, (9) *putamen basal ganglia*, (10) *spinal cord cervical c-1*, and (11) *substantia nigra*. We do not consider the available *cerebellum cortex* tissue samples from GTEx as these are duplicates of *cerebellar hemisphere* and *frontal cortex BA9* respectively. We instead only use the samples taken the same time as the other brain sub-regions at the University of Miami Brain Endowment Bank, preserved by snap freezing (see GTEx FAQs).

We also created an aggregate across $\mathcal{G}_i^{r^{\text{nc}}}$, all cis-eQTL genes associated with eSNP i for each non-cerebellar hemisphere brain tissue region r^{nc},

$$\bar{\beta}_i^{\text{nc}} = \frac{1}{|\mathcal{G}_i^{r^{\text{nc}}}|} \sum_{g \in \mathcal{G}_i^{r^{\text{nc}}}} |\beta_{i,g}^{r^{\text{nc}}}|. \tag{S12}$$

We did not include the cerebellum tissue samples in this aggregate due to the reported distinctness of the cerebellum relative to other brain tissue samples (GTEx Consortium 2015). Similarly, we computed an average across the two adipose tissues. As before, when calculating the various eQTL slopes summaries, if eSNP i was not an eQTL for a particular region then we impute a value of zero reflecting the lack of associated expression.

Furthermore, WGCNA module assignments were generated using protein coding genes for three different sets of tissues: (1) all non-cerebellar hemisphere brain tissues, (2) cerebellar hemisphere only tissue, and (3) adipose tissues (using same settings described previously in *Type 2 diabetes results*). Together with the WHR z-statistics and covariates accounting for the associations and WGCNA module indicators, x_i^{BMI} contained 110 variables.

For BMI eSPS, 376 have p-value exactly equal to one, leading to the same problem as we encountered in the T2D analysis. Again, we proceed by randomly drawing replacement p-values for these 376 eSNPs from a uniform distribution between 0.97 and $1-1\text{E}^{-15}$. Figure S19 shows how AdaPT fails to obtain any discoveries across the various α levels without making an adjustment to the p-values. With this limitation recognized, we proceed to focus on the discoveries returned by AdaPT using the adjusted p-values at $\alpha = 0.05$.

Unlike SCZ and T2D, AdaPT using all of the covariates (with the same tuning parameters as SCZ) detected fewer discoveries: 1,383 eSNPs compared to 1,624 eSNPs discovered by the intercept-only AdaPT model at target FDR level $\alpha = 0.05$. With further boosting regularization, beyond what is considered here, one could achieve the intercept-only results with gradient boosted trees. Of these 1,383 discoveries, approximately 83% (1,140 eSNPs) were nominal replications with p-values less than or equal to 0.05 in the independent *2018-only* studies. Figure S20 displays the increasing smoothing spline relationship between the *2018-only* p-values and the resulting *2015-only* q-values from the AdaPT search on the $\log_1 0$ scale. The much higher observed nominal replication rate is not surprising given the well powered size of the BMI studies, as indicated by the y-axis of Figure S20, which reflects the level of enrichment for the *2018-only* studies.

Additionally, gene ontology enrichment analysis for the 1,383 discoveries using all covariates revealed no significant biological process enrichment at target FDR level $\alpha = 0.05$. One concern is that a model with 110 variables is excessive, because the variable importance plots for the BMI AdaPT models in Figures S21(A-B), along with the partial dependence plots in Figures S22(A-B), emphasize the relative importance of the WHR z-statistics compared to other covariates. To test this conjecture, we explored two simpler models using (1) WHR z-statistics only and (2) WHR z-statistics with eQTL slope summaries. These produced 1,324 and 1,351 discoveries at the 0.05 level, respectively. We conclude that the available covariates do not provide sufficient additional information beyond the signal available with this immense sample and consequently including

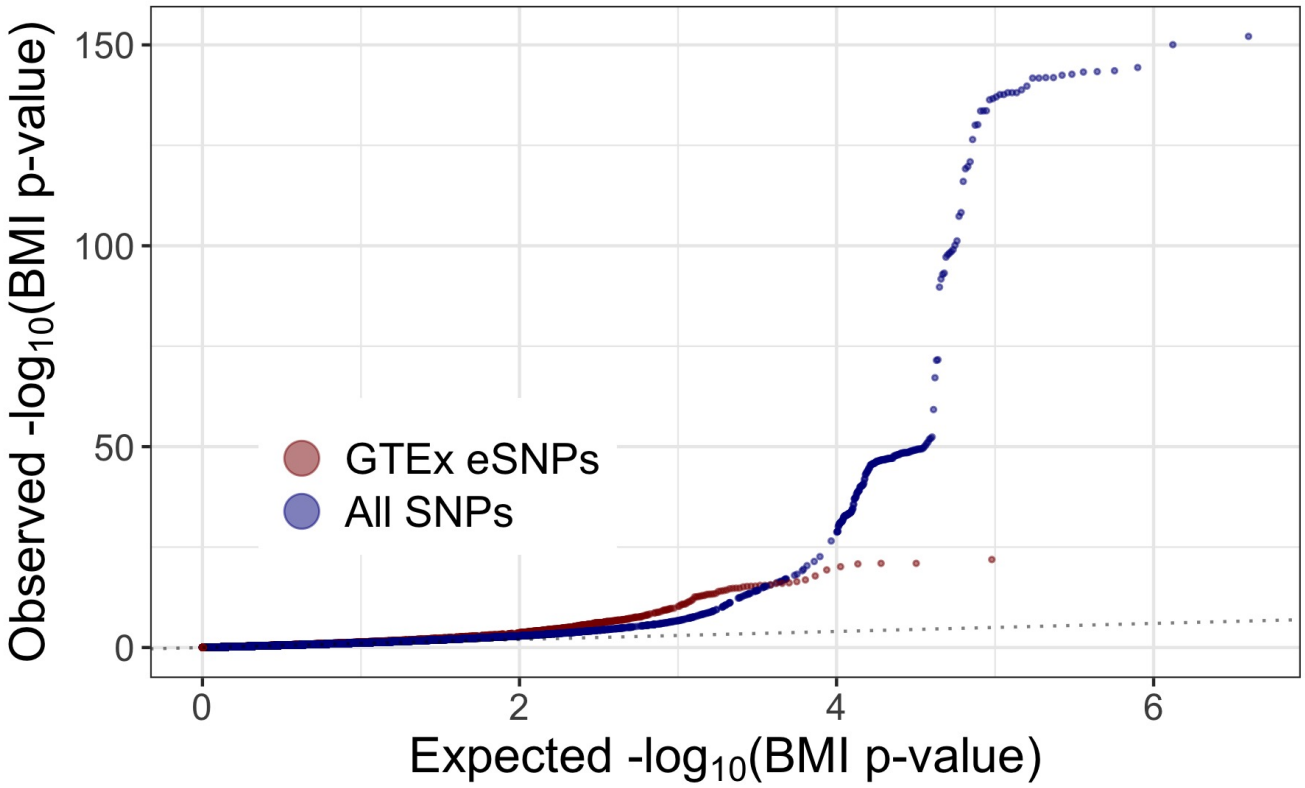

Figure S18: Comparison of qq-plots revealing BMI enrichment for GTEx eSNPs compared to full set of SNPs.

covariates in the AdaPT model does not increase the power of the procedure.

CV tuning for SCZ, T2D, and BMI results

Rather than fixing the parameter settings for the XGBoost gradient boosted trees, we use the CV algorithm (detailed in *Methods*) at two steps of the search to tune the models (see the following section for justification of using two CV steps). For our search space, we evaluate a small range of values for the number of trees P and limit the maximum tree depth D to result in reasonably shallow trees (referred to as `nrounds` and `max_depth` in the `xgboost` package (Chen et al. 2019)).

First, for SCZ analysis, when exploring the improvement in discovery rate for the eSNPs by incrementally including more information, we used the following XGBoost settings:

- BD z-stats: Combinations of $P \in \{100, 150\}$, $D \in \{1, 6\}$,

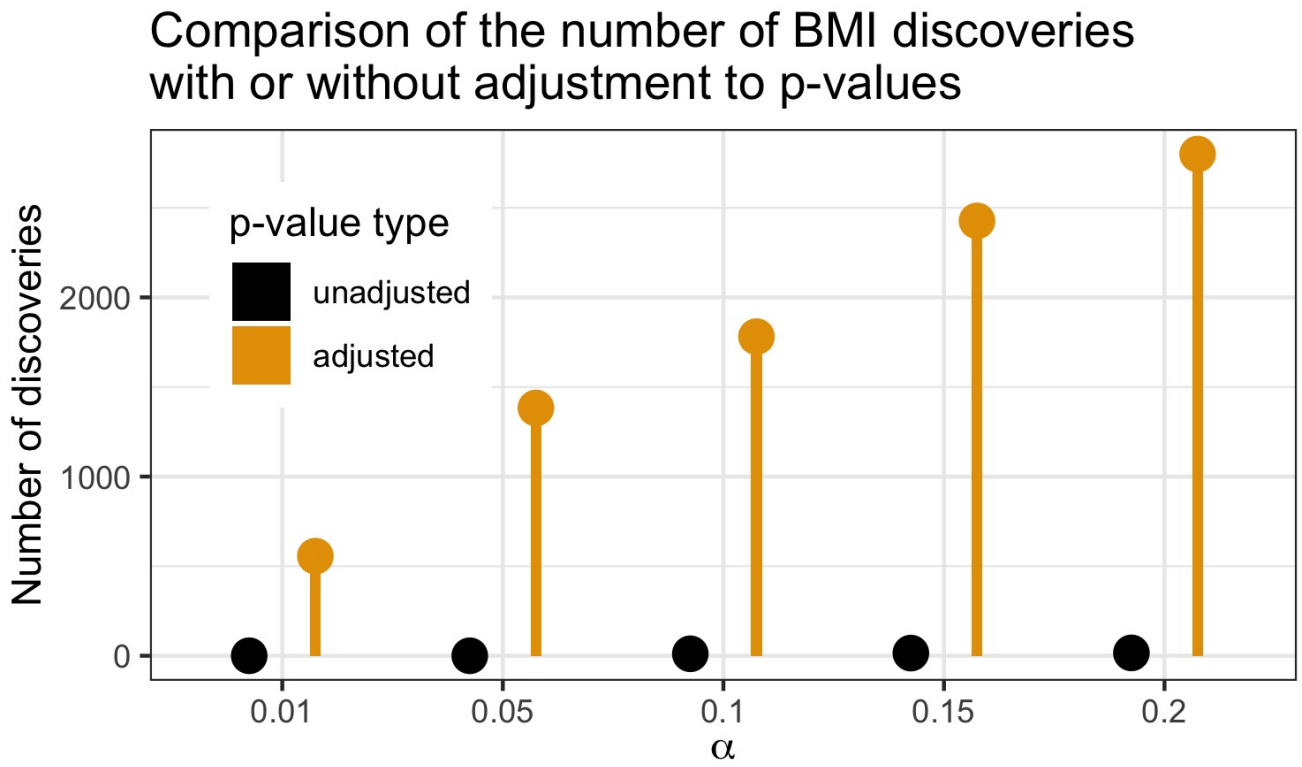

Figure S19: Comparison of the number of discoveries by AdaPT for BMI by whether or not the adjusted or unadjusted p-values were used.

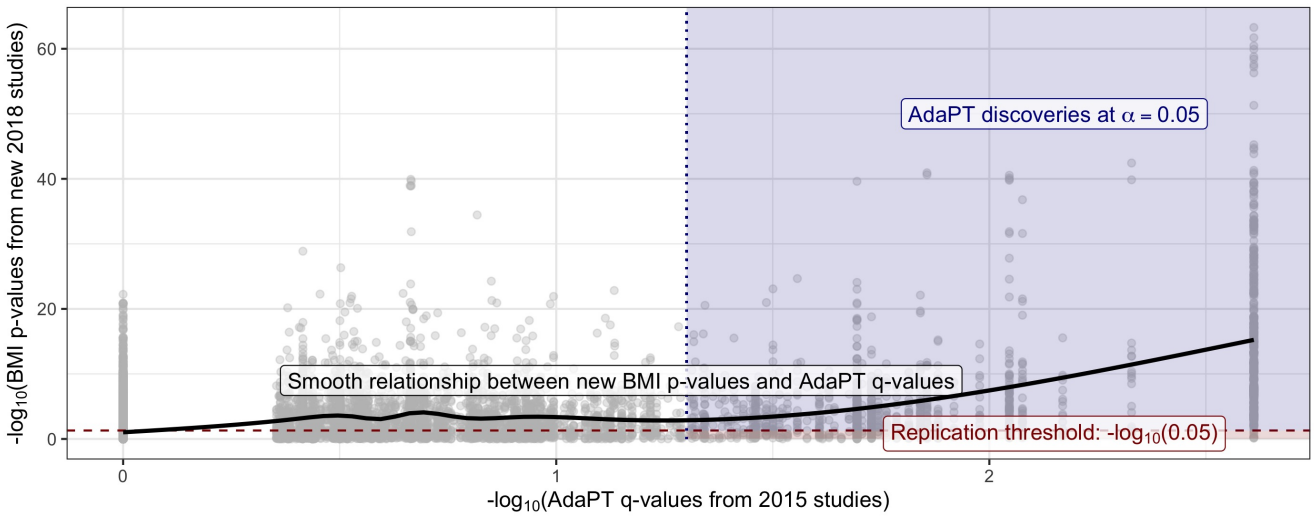

Figure S20: Black line displays smooth relationship between BMI p-values from *2018-only* studies and the AdaPT q-values from the *2015-only* studies. Blue-shaded region indicates AdaPT discoveries at $\alpha = 0.05$ that are nominal replications, p-values from the *2018-only* studies < 0.05 while red denotes discoveries which failed to replicate.

- BD z-stats + eQTL slopes: Combinations of $P \in \{100, 150\}$, $D \in \{3, 6\}$,
- BD z-stats + eQTL slopes + WGCNA: Combinations of $P \in \{100, 150\}$, $D \in \{2, 3\}$,
- WGCNA only: Combinations of $P \in \{100, 150\}$, $D \in \{1, 2, 3\}$.

We explored different settings for the different possible covariates to address the types of variables included. For instance, when using the BD z-statistics only, we considered both single-split "stumps" as well as more depth with six splits to potentially handle the variable's symmetric relationship. Once we have all three types of covariates (BD z-statistics, eQTL slope summaries, and WGCNA results), we limit the maximum depth to be at least two to ensure possible interactions can be captured.

The selected number of trees P and maximum depth D for each of these sets of covariates is displayed in Table S1. When using only the BD z-statistics, as well as only including the eQTL slopes, the single-split settings were selected in the first CV step while the higher depth was selected in the second CV step. When using all covariates, the most complex settings (largest number of trees and largest depth) are selected in both CV steps. This agreement in selection

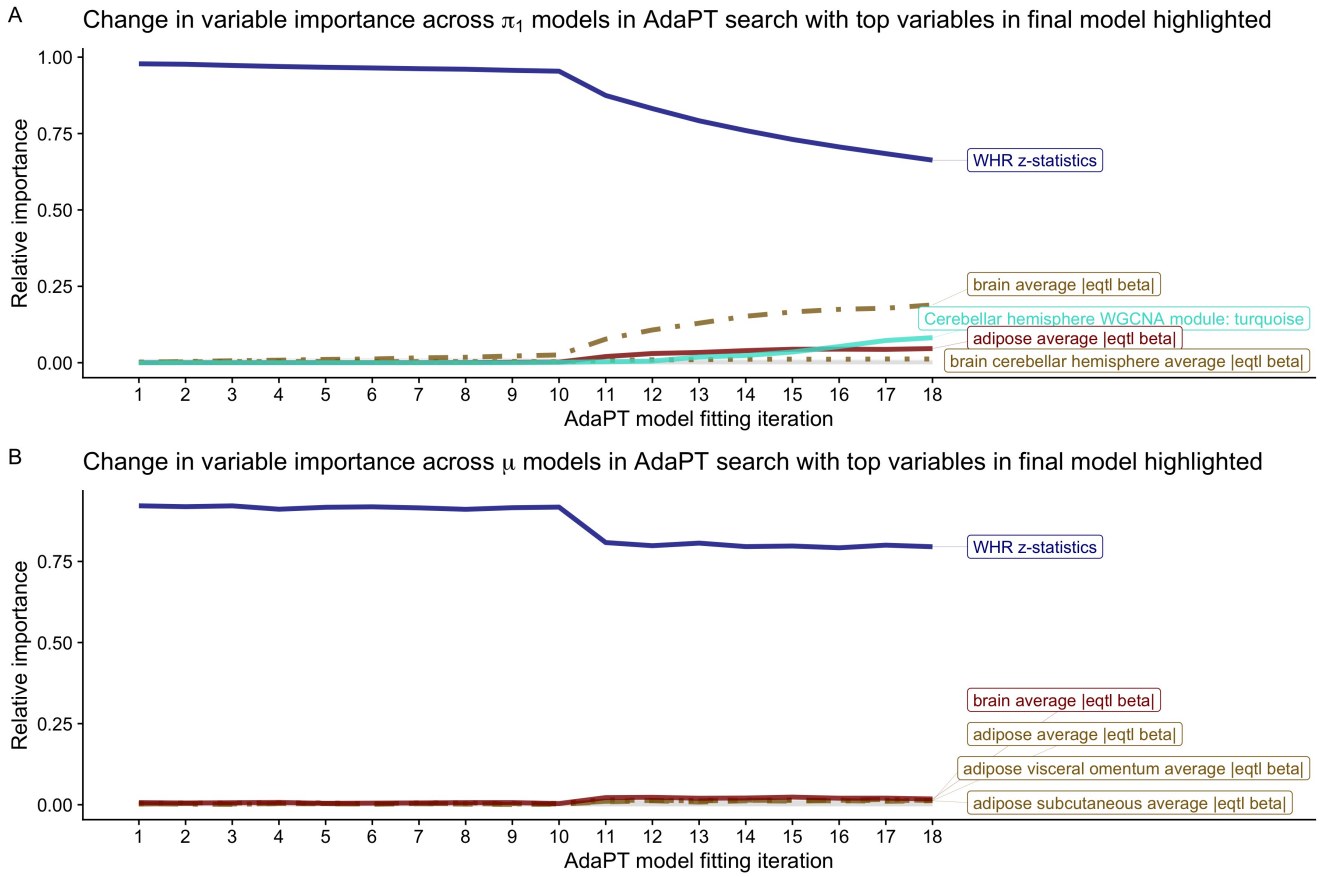

Figure S21: Change in BMI variable importance for AdaPT *(A)* probability of being non-null π_1 and *(B)* effect size under alternative μ models across search, with top variables in final model highlighted.

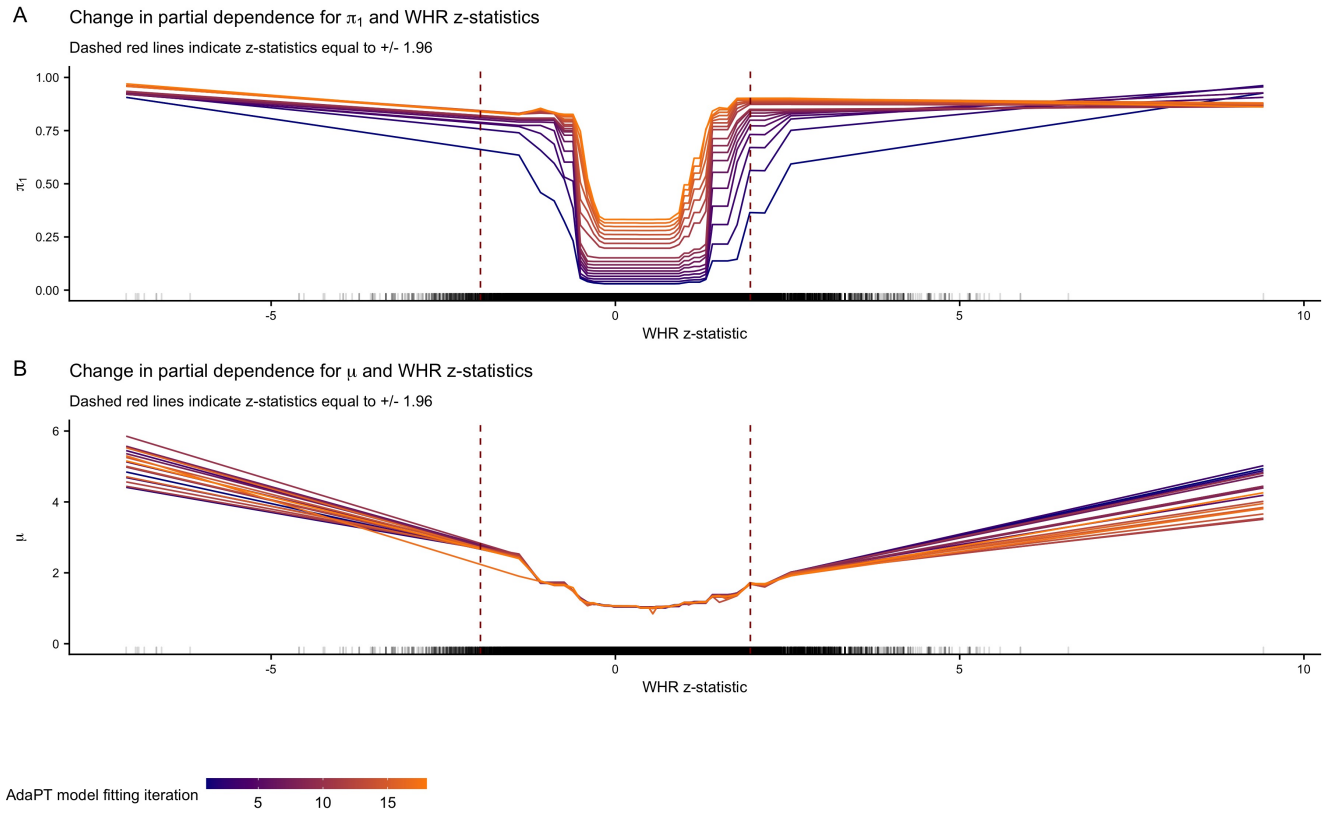

Figure S22: Change in BMI partial dependence for WHR z-statistics and AdaPT *(A)* probability of being non-null π_1 and *(B)* effect size under alternative μ models across search.

is not surprising given the choice of the low starting threshold $s_0 = 0.05$, which differs from the results displayed in Table S3 of the next section using $s_0 = 0.45$. We evaluated the same possible settings for the various *all 2018* results displayed in Figures S10(C-D): the same choices for P and D displayed in Table S1 were selected in both CV steps.

For the T2D and BMI results with their full set of covariates, we evaluated four combinations: (1) $P = 100,\ D = 2$, (2) $P = 150,\ D = 2$, (3) $P = 100,\ D = 3$, and (4) $P = 150,\ D = 3$. For the BMI results using only WHR z-statistics, we varied over $P \in \{100, 150\}$ and $D \in \{1, 6\}$; for the results using WHR z-statistics with the eQTL slopes, we used combinations of $P \in \{100, 150\}$, $D \in \{3, 6\}$. The selected number of trees P and maximum depth D for each of these sets of AdaPT results at both CV steps is displayed in Table S2.

Table S1: Selected boosting settings for number of trees P and maximum depth D with AdaPT CV algorithm by covariates for eSNPs in each CV step.

Covariates	m_1^*	m_2^*
BD z-stats	$P = 150,\ D = 1$	$P = 150,\ D = 6$
BD z-stats + eQTL slopes	$P = 150,\ D = 3$	$P = 150,\ D = 6$
BD z-stats + eQTL slopes + WGCNA	$P = 150,\ D = 3$	$P = 150,\ D = 3$
WGCNA only	$P = 150,\ D = 3$	$P = 150,\ D = 3$

Table S2: Selected boosting settings for number of trees P and maximum depth D with AdaPT CV algorithm by GWAS results in each CV step.

GWAS results	m_1^*	m_2^*
T2D	$P = 100,\ D = 2$	$P = 150,\ D = 3$
BMI (all covariates)	$P = 100,\ D = 2$	$P = 150,\ D = 3$
BMI (WHR z-stats only)	$P = 150,\ D = 1$	$P = 150,\ D = 1$
BMI (WHR z-stats + eQTL slopes)	$P = 100,\ D = 3$	$P = 150,\ D = 3$

Table S3: Selected boosting settings for number of trees P and maximum depth D with AdaPT CV algorithm by covariates for eSNPs with $s_0 = 0.45$.

Covariates	m_1^*	m_2^*
BD z-stats	$P = 50,\ D = 1$	$P = 150,\ D = 1$
BD z-stats + eQTL slopes	$P = 100,\ D = 1$	$P = 150,\ D = 2$
BD z-stats + eQTL slopes + WGCNA	$P = 100,\ D = 2$	$P = 150,\ D = 3$
WGCNA only	$P = 150,\ D = 3$	$P = 150,\ D = 3$

Selection of s_0 and number of CV steps

To justify the selection of both the starting threshold s_0 and number of CV steps for the AdaPT search, we generated simulations from the first AdaPT models returned from the SCZ *2014-only* results. While these models are based on AdaPT results with a starting threshold of $s_0 = 0.05$ following one CV step, they are only from the first model and are not explicitly parametrized by s_0 and the number of CV steps. We know, however, that these first models are the result of using $P = 150$ trees with a maximum depth of $D = 3$, as indicated in Table S1 of the previous section.

Let $\hat{\pi}_1^*$ and $\hat{\mu}^*$ be the first models for the probability of non-null and effect size under the alternative that AdaPT returns for the eSNPs using all covariates x_i^{SCZ}. We use these models as the "truth" for generating data, in which a single iteration of the simulation proceeds as follows:

- For each eSNP $i \in [n_{\text{SCZ}}^*]$
  1. Generate test status: $H_i | x_i^{\text{SCZ}} \sim \text{Bernoulli}\big(\hat{\pi}_1^*(x_i^{\text{SCZ}})\big)$.
  2. Generate simulated effect sizes:
  $$-\log p_i | H_i, x_i^{\text{SCZ}} \sim \begin{cases} \text{Exp}\big(1\big) \text{ if } H_i = 0, \\ \text{Exp}\big(1/\hat{\mu}^*(x_i^{\text{SCZ}})\big) \text{ if } H_i = 1. \end{cases} \tag{S13}$$
  3. Transform to p-values p_i.
- Apply AdaPT to simulated study p-values with specified s_0 and v CV steps with two candidate settings:
  1. number of trees $P = 100$ and maximum depth $D = 2$,

2. number of trees $P = 150$ and maximum depth $D = 3$.

- Compute observed power and FDP at range of target FDR α values.

We generate one-hundred simulations this way for each possible threshold $s_0 \in \{0.05, 0.25, 0.45\}$ and $v \in \{1, 2, 5\}$ CV steps. Figure S23 displays the average difference in power between the different starting threshold values by the number of CV steps. Although the differences are small, we see that using $s_0 = 0.05$ results in higher power, on average, than both 0.25 and the recommended 0.45 value. Using this low starting threshold of $s_0 = 0.05$, we then directly compute the difference in power between the different number of CV steps displayed in Figure S24. Unsurprisingly, while again the differences are small, only one CV step results in the lowest power, on average. Since the computational cost of AdaPT with CV tuning is reduced by only using two CV steps instead of a higher number, such as five, and the simulations demonstrate on average no difference in power at both α values of 0.05 and 0.10, we use the starting threshold of $s_0 = 0.05$ with two CV steps in our applications of AdaPT.

In the previous section, Table S1 displayed the selections in both CV steps with $s_0 = 0.05$. For comparison, Table S3 displays the selections using $s_0 = 0.45$. Instead of selecting the same settings in both steps, the higher initial threshold selects the least complex settings (smallest number of trees and minimum depth) in the first CV step before flipping to the most complex settings in the second step. Intuitively, the higher initial threshold means more information is masked from the models, so it is not surprising to see less complex settings chosen. This further reinforces the use of the lower initial threshold $s_0 = 0.05$: it starts with more revealed information and selects model settings corresponding to improved CV performance for tests with lower p-values of interest.

Dependent p-value block simulation

To demonstrate the performance of AdaPT in the presence of dependent tests, we construct simulations with a block-correlation scheme to emulate LD structure for SNPs. We consider a setting with two independent covariates,

$$x_i = (x_{i1}, x_{i2}),$$
$$\text{where } x_{i1}, x_{i2} \sim \text{Uniform}(0, 1).$$

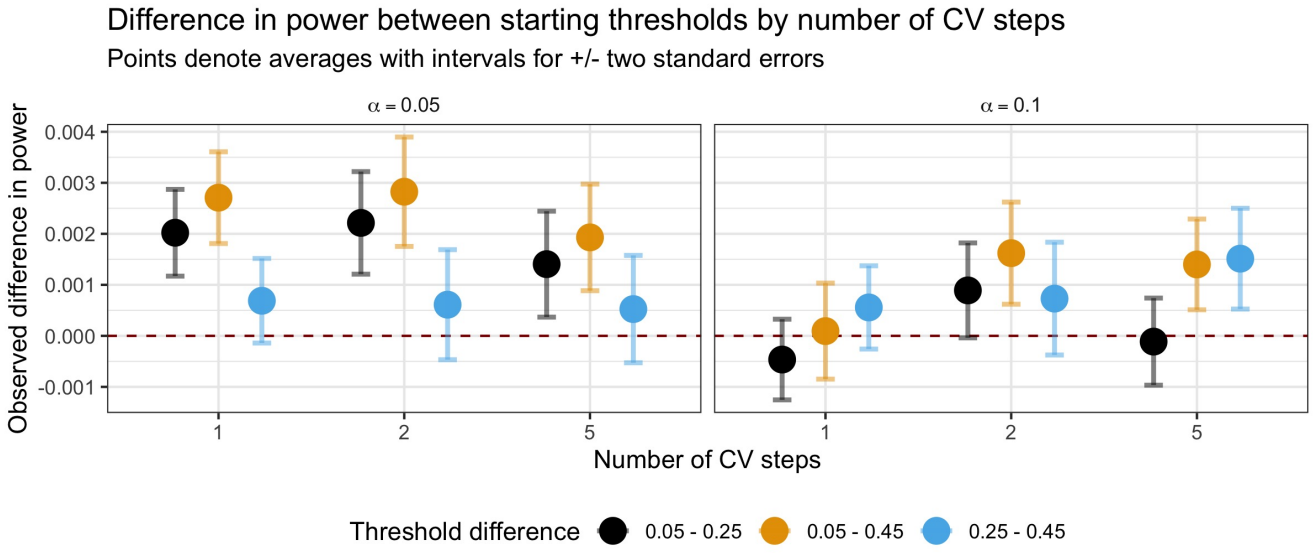

Figure S23: Difference in simulation power between different initial thresholds s_0 for AdaPT search by number of CV steps. Points denote averages with plus/minus two standard error bars.

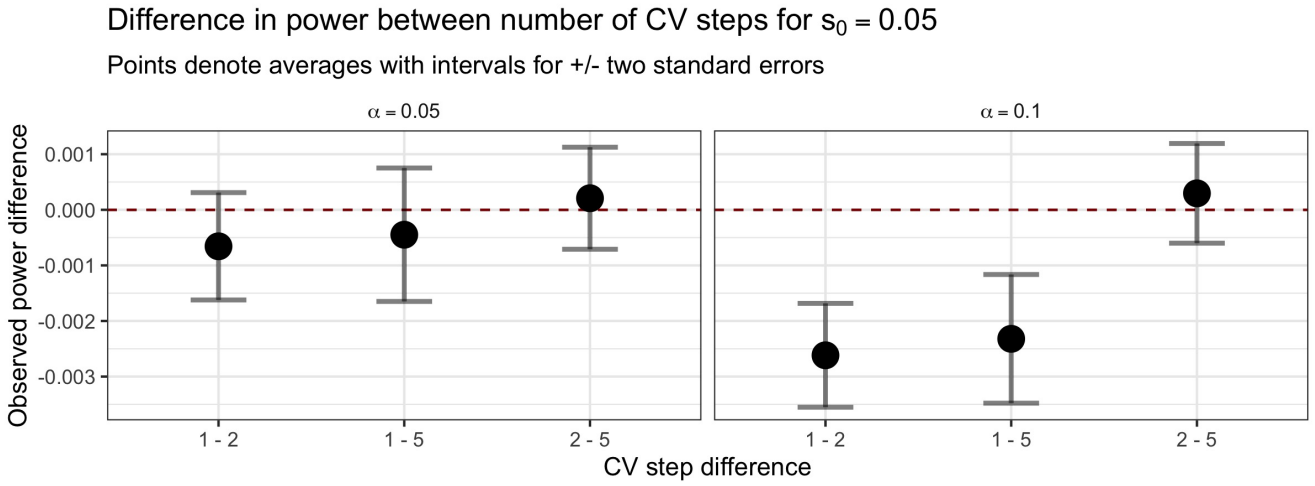

Figure S24: Difference in simulation power between the number of CV steps with $s_0 = 0.05$. Points denote averages with plus/minus two standard error bars.

For each test $i \in [n]$, we define a linear relationship for the log-odds of being non-null using these covariates,

$$\text{logit}(\pi_{1,i}(x_i)) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}.$$

Then, the resulting status of the test H_i is a Bernoulli random variable based on the probability $\pi_{1,i}(x_i)$ where $H_i = 1$ indicates the test i is non-null while $H_i = 0$ indicates a true null,

$$H_i \sim \text{Bernoulli}(\pi_{1,i}(x_i)).$$

Given this test status, a vector of true effect sizes $\boldsymbol{\mu} = c(\mu_i, \ldots, \mu_n)$ is also generated as a function of the covariates,

$$\mu_i(x_i) = \begin{cases} \max\{\mu_{floor},\ \gamma_1 x_{i1} + \gamma_2 x_{i2}\} \text{ if } H_i = 1, \\ 0 \text{ otherwise.} \end{cases}$$

To simulate observed effect sizes, we construct an $n \times n$ covariance matrix $\boldsymbol{\Sigma}$ with B blocks of equal size $\frac{n}{B}$. Each block $b \in [B]$ has constant correlation ρ between all tests *within* the block, while each block is independent of each other. This results in constructing individual block covariance matrices, Σ_b, with ones along the diagonal and ρ for the off-diagonal elements. Each of these individual matrices are placed along the diagonal of $\boldsymbol{\Sigma}$, with the remaining off-diagonal elements set to zero so blocks are independent of each other. As an example, if each block contained only two tests they would be constructed in the following manner,

$$\Sigma_b = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \quad \Rightarrow \quad \boldsymbol{\Sigma} = \begin{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{0} & \ldots & \ldots & \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \end{bmatrix}$$

Using this block-wise construction of the covariance matrix, we then proceed to generate the vector of observed effect sizes $\boldsymbol{z} = (z_i, \ldots, z_n)$ from a multivariate Gaussian distribution,

$$\boldsymbol{z} \sim \text{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

We compute the resulting two-side p-value $p_i = 2 \cdot \Phi(-|z_i|)$ for each test's observed effect size.

For each dataset generated using this process above, we compute both the observed FDP and power for the classical BH procedure and two different versions of AdaPT:

1. intercept-only,

2. gradient boosted trees with covariates: $x_i = (x_{i1}, x_{i2})$.

We fix both $n = 10{,}000$ and $B = 500$ blocks, resulting in 500 blocks of twenty tests each. Rather than force all non-nulls together in the same blocks, we first calculate the minimum number of blocks required to hold all non-null tests, $B_A^* = \lceil |\{i : H_i = 1\}|/20 \rceil$. The non-null tests are then randomly assigned to $B_A = \lceil (500 + B_A^*)/2 \rceil$ blocks, ensuring that there will be blocks containing both null and non-null tests. The $|\{i : H_i = 0\}|$ tests are randomly assigned to available spots within the B_A blocks as well as the remaining $500 - B_A$ strictly null blocks.

In our simulations, we fix $\beta_0 = -3$ and require that both $\beta_1 = \beta_2$ and $\gamma_1 = \gamma_2$. We vary the following settings in our simulations:

1. block correlation $\rho \in \{0, 0.25, 0.5, 0.75, 1\}$ where each block has the same value for ρ,

2. $\beta_1, \beta_2 \in \{1, 2, 3\}$,

3. $\mu_{floor} \in \{0.5, 1, 1.5\}$,

4. $\gamma_1, \gamma_2 \in \{0.5, .75, 1\}$.

We generate 100 simulations using the data generating process above, computing both the FDP and power for BH and the two different versions of AdaPT. For the covariate-informed version of AdaPT, we use gradient boosted trees via XGBoost with $P = 100$ trees and maximum depth $D = 1$. For both versions of AdaPT results, we start with the initial threshold of $s_0 = 0.45$ and update the model ten times throughout the search (rather than the recommended twenty for computational speed).

Figures S25, S26, and S27 display points for the average observed FDP and power across the 100 simulations with plus/minus two standard errors bars for μ_{floor} =0.5, 1, and 1.5 respectively, with target FDR level $\alpha = 0.05$. The columns in each figure correspond to the different values considered for $\gamma_1 = \gamma_2$, while the rows correspond to $\beta_1 = \beta_2$. The x-axis for the figures displays the increasing block correlation ρ. Regardless of the simulation setting, we see that the AdaPT

results when accounting for covariates (x_{i1}, x_{i2}) maintains valid FDR control at 0.05 similar to BH. This holds in the settings with greater effect sizes, as well as when the covariate information displays the best performance in terms of observed power (the bottom right panels of each figure). We can see that the intercept-only approach fails to achieve FDR control under block settings with perfect correlation, while the use of covariate information appears to inhibits such behavior. Our focus on positive correlation values is synonymous with the setting faced in genomics regarding LD structure. Further exploration of AdaPT's performance in settings with arbitrary dependence structure presents an opportunity for future work, as well as accounting for covariate information that predict observed correlated noise.

Simulations demonstrating effects of overfitting

It is possible that flexible methods like gradient boosted trees can be overfit, especially on small data sets. This could potentially lead to concerns about their incorporation in AdaPT. To assess the effects of overfitting the gradient boosted trees in AdaPT, we constructed simulated datasets using the first models returned by AdaPT on the SCZ GWAS results, $\hat{\pi}_1^*$ and $\hat{\mu}^*$, with the actual covariates x_i^{SCZ} for each of the $n^*_{\text{SCZ}} = 25{,}076$ eSNPs. We then simulated data using these models in the same manner previously explained for choosing s_0 and the number of CV steps, and computed the observed power and FDP over a range of number of trees $P \in \{100, 300, 500, 700, 900\}$.

Figure S28(A) displays the distributions for fifty simulations of the observed FDP as the number of trees in the gradient boosted model increases. Regardless of the number of trees, we still maintain valid FDR control. However, Figure S28(B) shows as the number of trees increases, the method will overfit, resulting in a reduction in power. This reinforces that, although good model tuning can be important for power, the AdaPT method continues to maintain FDR control even as the model breaks down.

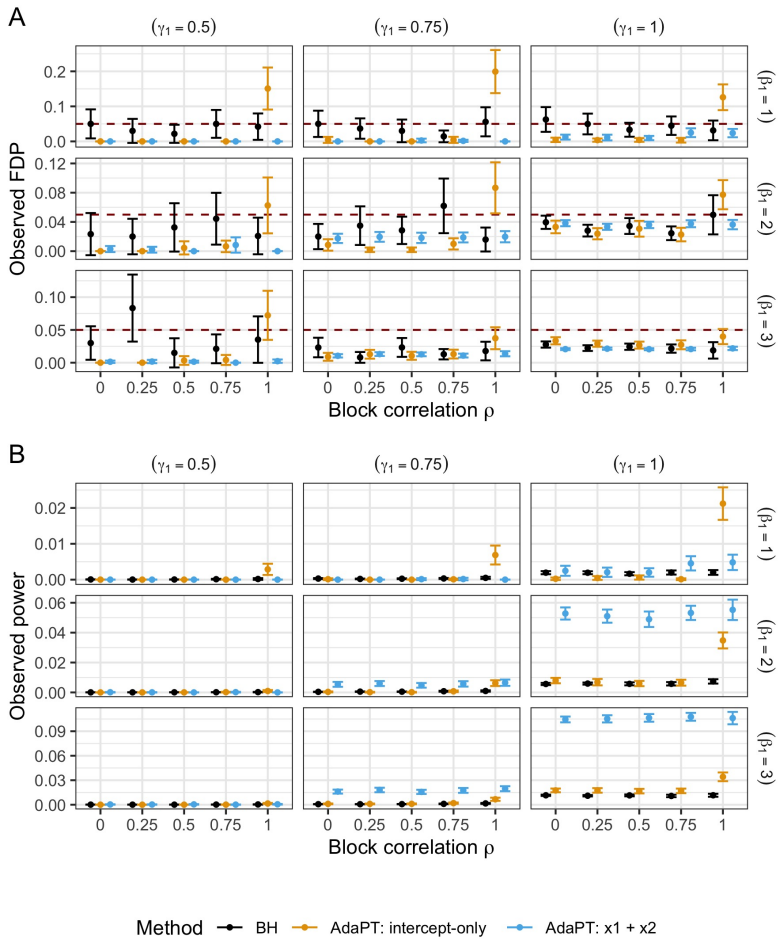

Figure S25: Comparison of average (A) FDP and (B) power with plus/minus two standard error bars for 100 simulations with $\mu_{floor} = 0.5$, and varying values for β_1 (rows) and γ_1 (columns) and block correlation ρ.

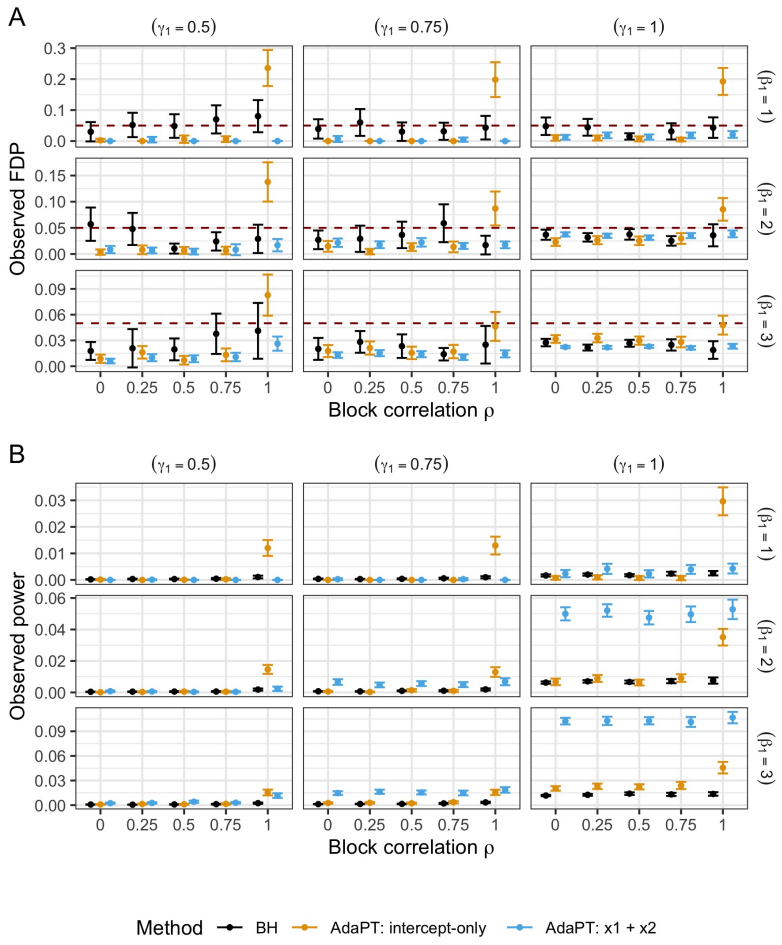

Figure S26: Comparison of average (A) FDP and (B) power with plus/minus two standard error bars for 100 simulations with $\mu_{floor} = 1$, and varying values for β_1 (rows) and γ_1 (columns) and block correlation ρ.

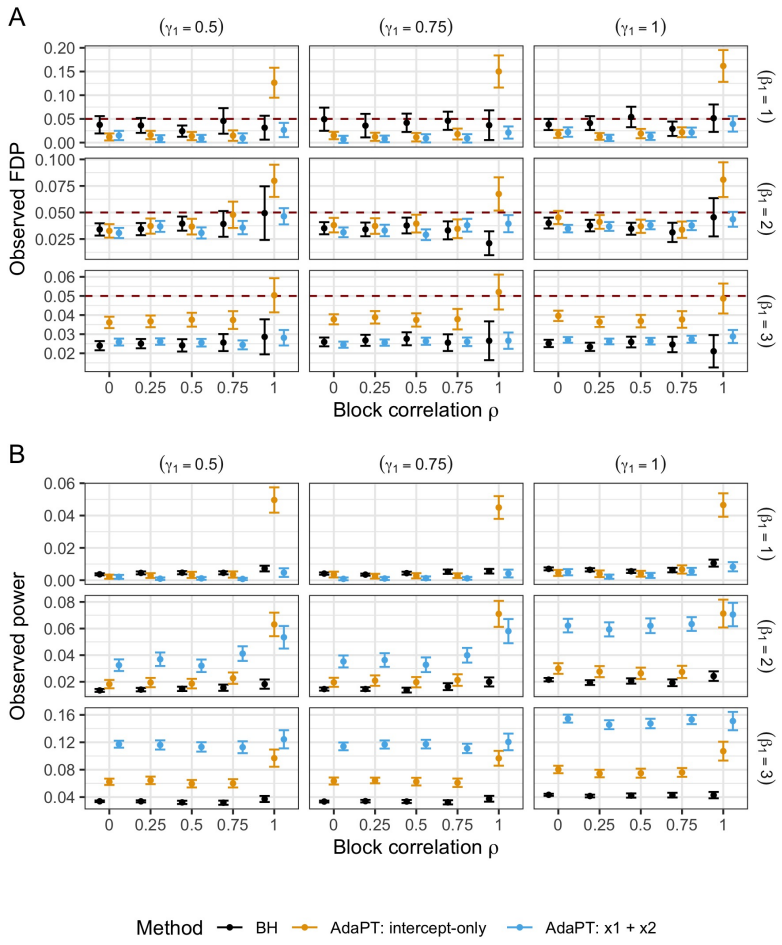

Figure S27: Comparison of average (A) FDP and (B) power with plus/minus two standard error bars for 100 simulations with $\mu_{floor} = 1.5$, and varying values for β_1 (rows) and γ_1 (columns) and block correlation ρ.

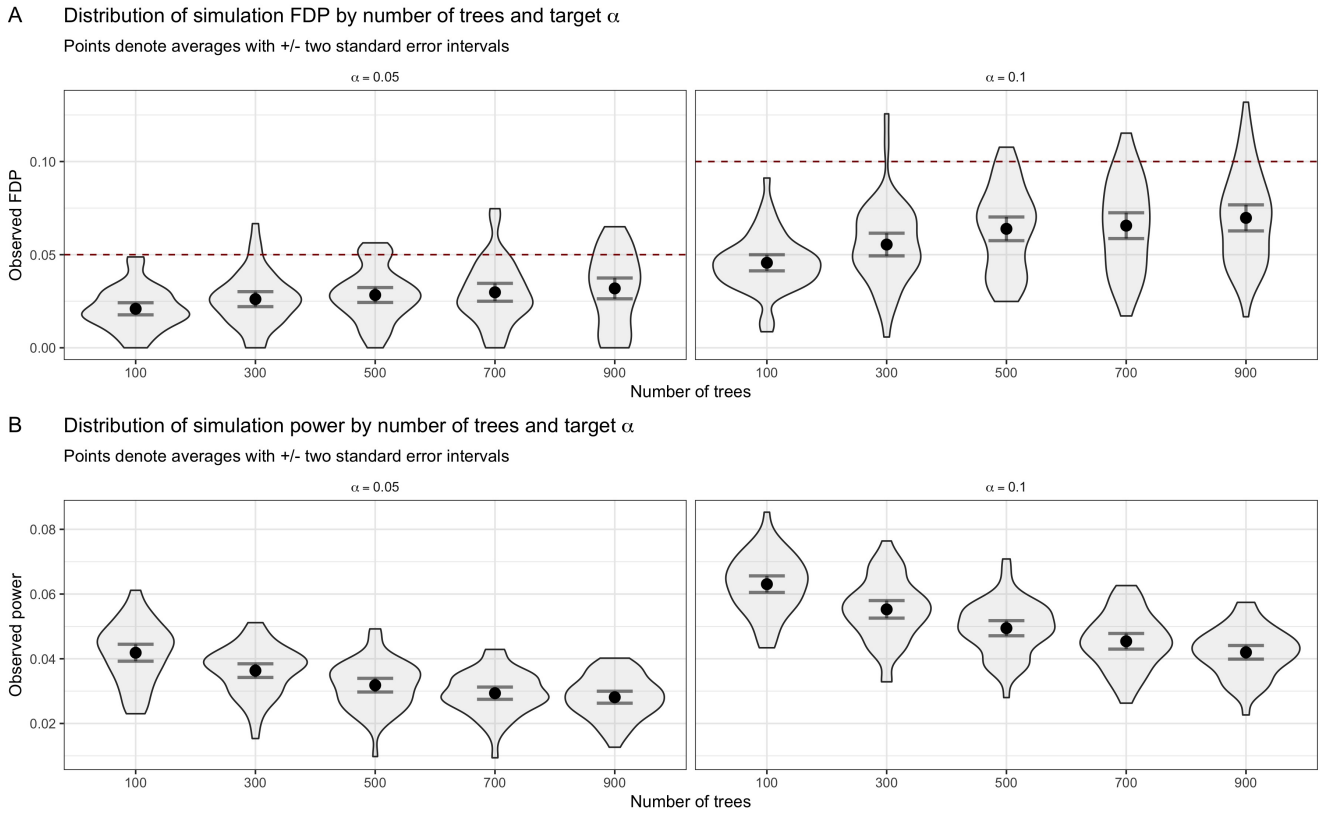

Figure S28: Distributions of observed *(A)* FDP and *(B)* power for simulations as the number of AdaPT gradient boosted trees increases by target FDR level α. Points denote averages with plus/minus two standard error intervals.