KnockoffTrio: A knockoff framework for the identification of putative causal variants in genome-wide association studies with trio design

Authors

Yi Yang, Chen Wang, Linxi Liu, Joseph Buxbaum, Zihuai He, Iuliana Ionita-Laza

Correspondence

ii2135@columbia.edu

We introduce KnockoffTrio, a statistical method to identify putative causal genetic variants for the father-mother-child trio design built upon a recently developed knockoff framework. KnockoffTrio controls the false discovery rate, protects against confounding due to population stratification, and is more powerful than conventional methods that control the family-wise error rate.

Yang et al., 2022, The American Journal of Human Genetics *109*, 1761–1776
October 6, 2022

https://doi.org/10.1016/j.ajhg.2022.08.013

KnockoffTrio: A knockoff framework for the identification of putative causal variants in genome-wide association studies with trio design

Yi Yang,[1,2,3] Chen Wang,[1] Linxi Liu,[4] Joseph Buxbaum,[5] Zihuai He,[6,7] and Iuliana Ionita-Laza[1,*]

Summary

Family-based designs can eliminate confounding due to population substructure and can distinguish direct from indirect genetic effects, but these designs are underpowered due to limited sample sizes. Here, we propose KnockoffTrio, a statistical method to identify putative causal genetic variants for father-mother-child trio design built upon a recently developed knockoff framework in statistics. KnockoffTrio controls the false discovery rate (FDR) in the presence of arbitrary correlations among tests and is less conservative and thus more powerful than the conventional methods that control the family-wise error rate via Bonferroni correction. Furthermore, KnockoffTrio is not restricted to family-based association tests and can be used in conjunction with more powerful, potentially nonlinear models to improve the power of standard family-based tests. We show, using empirical simulations, that KnockoffTrio can prioritize causal variants over associations due to linkage disequilibrium and can provide protection against confounding due to population stratification. In applications to 14,200 trios from three study cohorts for autism spectrum disorders (ASDs), including AGP, SPARK, and SSC, we show that KnockoffTrio can identify multiple significant associations that are missed by conventional tests applied to the same data. In particular, we replicate known ASD association signals with variants in several genes such as *MACROD2*, *NRXN1*, *PRKAR1B*, *CADM2*, *PCDH9*, and *DOCK4* and identify additional associations with variants in other genes including *ARHGEF10*, *SLC28A1*, *ZNF589*, and *HINT1* at FDR **10**%.

Introduction

The father-mother-child trio design is a popular family-based design, especially for early-onset diseases. One important example is autism spectrum disorders (ASDs), where several prominent studies have successfully employed such a design.[1–3] The main advantages of the family-based design are that it is robust to external confounders such as population structure[4,5] and can help distinguish between direct and indirect effects.[6] Although popular methods have been proposed to account for confounding effects of population structure in the context of population-based designs,[7–9] a more reliable approach to eliminating such confounders is to use randomized experiments, and family-based designs provide an analogy to such experiments because of the randomness in transmission of genetic material from parents to offspring.[10] However, a main limitation of genome-wide association studies (GWASs) with family-based designs is the modest sample sizes, which ultimately leads to reduced power.

Most of the existing studies have focused on controlling the family-wise error rate (FWER) to account for multiple testing in genome-wide association studies. Given the polygenic nature of many complex traits, with a large number of small-effect loci accounting for most of the trait heritability, a more meaningful and powerful strategy is to control the false-discovery rate (FDR) that quantifies the expected proportion of false discoveries. Control of FDR has been previously suggested in genome-wide association studies[11,12] and has been successfully employed in genetic association studies of ASD.[13,14] Valid control of FDR is, however, difficult to achieve using the standard Benjamini-Hochberg (BH) procedure due to possible complex correlations among genetic variants. The knockoff-based framework we employ here allows valid FDR control under arbitrary correlations.

The idea of the knockoff-based inference is to construct knockoff copies of the original features (genotypes) that preserve the correlation structure and are independent of the trait conditional on the original features.[15] These knockoff features serve as negative controls and, when compared with the original features, help identify the truly causal ones. The knockoff-based inference provides rigorous control of FDR under arbitrary correlation structure and is thus more versatile than the BH procedure that requires independence or positive dependence[16] for the FDR control. Several knockoff procedures have been proposed with applications to population-based designs, including KnockoffZoom[17] for genome-wide association studies based on hidden Markov models and KnockoffScreen[18] for whole-genome sequencing data based on the sequential conditional independent tuples (SCIT) algorithm. These

[1]Department of Biostatistics, Columbia University, New York, NY 10032, USA; [2]Department of Biostatistics, City University of Hong Kong, Hong Kong SAR, China; [3]School of Data Science, City University of Hong Kong, Hong Kong SAR, China; [4]Department of Statistics, University of Pittsburgh, Pittsburgh, PA 15260, USA; [5]Departments of Psychiatry, Neuroscience, and Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; [6]Quantitative Sciences Unit, Department of Medicine, Stanford University, Stanford, CA 94305, USA; [7]Department of Neurology and Neurological Sciences, Stanford University, Stanford, CA 94305, USA
*Correspondence: ii2135@columbia.edu
https://doi.org/10.1016/j.ajhg.2022.08.013.

methods, however, were designed for independent individuals in population-based studies, making them unsuitable for family-based studies as considered in this article. Likewise, KnockoffGWAS[19] is a population-based knockoff procedure adjusting for possible relatedness among individuals in the study and is not based on a within-family test as proposed here. A related approach to construct synthetic offspring has been proposed before in order to perform causal inference with trio designs.[10] Specifically, Bates et al. proposed a digital twin test based on the conditional randomization test,[15] a method related to the knockoff but that can produce valid empirical p values. Computational cost is a concern for this test, especially in high-dimensional genome-wide settings where a large number of random drawings are needed to get small empirical p values.

In this paper we propose KnockoffTrio, a knockoff-based framework for the analysis of trio data in genome-wide association studies. Conventional association tests for family-based designs include the family-based association test (FBAT),[5] a generalization of the transmission disequilibrium test (TDT)[20] to handle various practical complexities such as missing parental data, covariate adjustment, and different types of phenotypes. Methods based on kernel machine regression under a generalized linear mixed model framework have also been proposed for family-based designs[21,22] and for population-based designs adjusting for population structure and relatedness.[8] Compared to these conventional testing strategies, KnockoffTrio enjoys several advantages of the general knockoff-based inference, such as higher statistical power, prioritization of causal variants over associations due to linkage disequilibrium, and robustness in controlling false positives in the presence of linkage disequilibrium between causal and non-causal variants,[17,18] while providing protection against external confounders such as population stratification. Furthermore, KnockoffTrio can leverage more general machine learning models while increasing power and maintaining proper FDR control regardless of the validity of the assumed model.

Material and methods

Knockoff generation for trio design

We assume a study with n trios and p genetic variants. We denote the matrix of trio genotypes by $G \in \{0,\ 1,\ 2\}^{3n\times p}$. Our goal is to test the conditional null hypothesis

$$H_{0g} : Y \perp G_g | G_{-g},$$

where Y are the phenotypes and $g \subset [p]$ is a continuous block. That is, variant(s) in group g (e.g., a gene or a region) are null if Y is independent of G_g given variants outside g.

We describe a knockoff generation method for the trio design to capture sample relatedness and test the above hypothesis. Our method assumes knowledge of haplotype phase; most phasing algorithms are able to provide highly accurate estimates of haplotypes when applied to trio datasets.[23] We first generate knockoff haplotypes for the parents, and then, conditional on them, we generate the knockoff haplotypes for the offspring. We describe the algorithm as follows:

Algorithm 1: Generation of knockoff trios

1. Sample one haplotype from each father into a group; assign the remaining haplotypes to the second group.
2. Repeat step 1 for mothers and obtain two additional groups of haplotypes.
3. Apply the SCIP algorithm[18] to each group of haplotypes and obtain the corresponding knockoffs (see below).
4. Generate knockoff offspring haplotypes conditional on the knockoff parental haplotypes (see below).

Note that in steps 1 and 2, we assign an individual's two haplotypes to two separate groups when generating their knockoffs so that the permutation-based SCIP algorithm below does not use the residual from one haplotype to generate the other haplotype's knockoff. This is done to increase the contrast between the original and knockoff genotypes in an individual and, hence, to improve power.

SCIP algorithm to generate knockoff parental haplotypes

We adopt the residual permutation method proposed in KnockoffScreen[18] to generate knockoff haplotypes for the parents. The residual permutation method is based on the general sequential conditional independent pairs (SCIP) algorithm[15,] defined as follows:

Algorithm 2: SCIP algorithm for knockoff haplotype generation

$j = 1$
while $j \le p$ do
  Sample $\tilde{H}_j$ independently from $L(H_j | H_{k\in B_j}, \tilde{H}_{1\le k\le j-1, k\in B_j})$
  $j = j + 1$
end while

where H_j and $\tilde{H}_j$ denote the original and knockoff parental haplotypes for the jth variant, respectively, and B_j denotes the subset of variants in a neighborhood of the jth variant (± 100 kb from the variant). Algorithm 2 has been shown to generate knockoffs that preserve the exchangeability conditions between the original and the knockoff genotypes necessary for controlling the FDR.[18] In the context of genetic data, the exchangeability implies the invariance in the linkage disequilibrium structure when one swaps a subset S of genetic variants with their knockoffs, i.e., $(H,\ \tilde{H})_{\text{swap}(S)} \overset{D}{=} (H,\ \tilde{H})$, in which $(H,\ \tilde{H})_{\text{swap}(S)}$ is obtained from $(H,\ \tilde{H})$ by swapping H_j and $\tilde{H}_j$, $\forall j \in S$.

As in He et al.,[18] we consider a semiparametric model for $L(H_j | H_{k\in B_j}, \tilde{H}_{1\le k\le j-1, k\in B_j})$ in KnockoffTrio:

$$H_j = \beta_0 + \sum_{k\neq j, k\in B_j} \beta_k H_k + \sum_{k\le j-1, k\in B_j} \gamma_k \tilde{H}_k + \varepsilon_j,$$

where ε_j is a random error term with a mean of zero. We obtain $\widehat{\beta}$, $\widehat{\gamma}$, fitted values $\widehat{H}_j$, and residuals $\widehat{\varepsilon}_j = H_j - \widehat{H}_j$ by minimizing the mean squared loss. We then obtain permuted residuals $\widehat{\varepsilon}_j^*$ and define the parental knockoffs $\tilde{H}_j = \widehat{H}_j + \widehat{\varepsilon}_j^*$.

Generating knockoff offspring haplotypes

Conditional on the knockoff parental haplotypes generated as above, we then proceed to generate the knockoff offspring haplotypes. Given the phased haplotypes of the original trio for a region, we first infer which parental haplotypes were transmitted to the offspring by matching parental haplotypes with offspring

haplotypes. We assume that no recombination occurs in the transmission of haplotypes from parents to offspring in any small region. We then use the knockoff haplotypes that correspond to the transmitted haplotypes in the original trio as the offspring's knockoff haplotypes.

Missing parental data

It is possible to accommodate missing parental data, i.e., one parent in a trio is completely missing. In such cases, one can still generate knockoff versions of such incomplete trios: the haplotype transmitted by the missing parent remains the same, while the other haplotype is obtained based on the knockoff haplotypes for the available parent. Because the FBAT test in the importance score can deal with missing parental data by design, the same feature importance score described below can be calculated.

Exchangeability property

As with independent samples, we need certain exchangeability properties to hold for the trio design in order for the FDR control to hold.[15] We formally prove the exchangeability property and FDR control for the trio design in supplemental note 1.

Multiple knockoffs to improve power and stability

The knockoff generation algorithm described above generates one single knockoff haplotype for each original haplotype. However, the inference based on a single knockoff often has limited power due to the detection threshold of $\frac{1}{q}$, i.e., the number of independent signals required for making any discoveries at the target FDR q. In particular, there is no power at the target FDR q if there are fewer than $\frac{1}{q}$ discoveries to be made, which is not uncommon when q is low and the signal is sparse. Moreover, the randomness in the sampling of a single knockoff makes the results unstable particularly for weak causal effects. Therefore, to further improve the stability and power at low target FDR, we extend the above single-knockoff algorithm to generating multiple knockoffs. For M knockoffs, the detection threshold decreases from $\frac{1}{q}$ to $\frac{1}{Mq}$, making it more powerful to detect sparse signals even when the target FDR level q is low. Furthermore, multiple knockoffs help improve the stability and reproducibility of the results.

Algorithm 3: SCIT algorithm for multiple knockoffs

$j = 1$

while $j \leq p$ do

Sample $\tilde{H}_j^1, \ldots, \tilde{H}_j^M$ independently from $L(H_j | H_{k \in B_j}, \tilde{H}^1_{1 \leq k \leq j-1, k \in B_j}, \ldots, \tilde{H}^M_{1 \leq k \leq j-1, k \in B_j})$

$j = j + 1$

end while

The semiparametric model for $L(H_j | H_{k \in B_j}, \tilde{H}^1_{1 \leq k \leq j-1, k \in B_j}, \ldots, \tilde{H}^M_{1 \leq k \leq j-1, k \in B_j})$ in the multiple-knockoff setting is:

$$H_j = \beta_0 + \sum_{k \neq j, k \in B_j} \beta_k H_k + \sum_{1 \leq m \leq M} \sum_{k \leq j-1, k \in B_j} \gamma_k^m \tilde{H}_k^m + \varepsilon_j,$$

where ε_j is a random error term with a mean of zero. We obtain $\widehat{\beta}$, $\widehat{\gamma}$, fitted values $\widehat{H}_j$, and the residuals $\widehat{\varepsilon}_j$ and their permutations $\widehat{\varepsilon}_j^*$. We then define the mth knockoff $\tilde{H}_j^m = \widehat{H}_j + \widehat{\varepsilon}_j^{*m}$.

KnockoffTrio: A knockoff framework for trio design

We describe here a knockoff-based test using a FBAT to compute the importance scores. Note that the use of FBAT to calculate feature importance statistics in KnockoffTrio helps protect against external confounders such as population stratification (see also supplemental note 1).

KnockoffTrio-FBAT

Once the knockoff generation for the father-mother-child trio data is completed, KnockoffTrio-FBAT performs a genome-wide scanning procedure with a window ϕ_{kl} in both the original and the knockoff data. We consider several candidate window sizes (e.g., in our applications 1 bp and 1, 5, 10, 20, and 50 kb) for ϕ_{kl}, with half of each window overlapping with neighboring windows of the same size. We employ the weighted burden FBAT,[24] which is a generalization of the SNP-based FBAT for a set of variants. Let n denote the number of trios and p denote the number of variants in a window. When $p = 1$, the weighted burden FBAT is equivalent to the SNP-based FBAT. The weighted burden FBAT statistic W_w for trio design is computed as:

$$W_w = \sum_{j=1}^{p} w_j U_j,$$

$$U_j = \sum_{i=1}^{n} (Y_i - u) U_{ij},$$

$$U_{ij} = X_{ij} - E\left(X_{ij} | P_{ij}^1, P_{ij}^2\right),$$

in which w_j is a weight associated with the jth variant, Y_i is a dichotomous or quantitative trait for the offspring in the ith trio, u is an offset parameter, X_{ij} is the offspring genotype, P_{ij}^1 and P_{ij}^2 are the parental genotypes, and $E(X_{ij} | P_{ij}^1, P_{ij}^2)$ is the expected value of the offspring genotype conditional on parental genotypes. Typically, $u = 0$ for dichotomous traits and $u = \overline{Y}$ for quantitative traits. The choice of w_j is flexible and can reflect any prior functional information on the variant; in this study we consider $w_j = (\sqrt{np_j(1-p_j)})^{-1}$, in which n is the number of trios and p_j is the minor allele frequency (MAF) for the jth variant. We can further obtain the variance of W_w as

$$Var(W_w) = \sum_{i=1}^{n} (Y_i - \mu_i)^2 \left[\sum_{j=1}^{p} w_j^2 Var\left(X_{ij} \middle| P_{ij}^1, P_{ij}^2\right) + \sum_{j \neq k} w_j w_k Cov\left(X_{ij}, X_{ik} \middle| P_{ij}^1, P_{ij}^2, P_{ik}^1, P_{ik}^2\right) \right].$$

Therefore, the standardized test statistic $Z = W_w / \sqrt{Var(W_w)}$ approximately follows a standard normal distribution in large samples under the null hypothesis of no association between any of the p variants and the trait.

Aggregated Cauchy association test to compute importance scores

For a given window ϕ_{kl} we compute an importance score as follows:

- For a 1 bp window, KnockoffTrio-FBAT implements SNP-based FBAT for variants with a MAF ≥ 0.01 and obtain $p_{\phi_{kl}}$ and $p^m_{\phi_{kl}}$ (for the mth knockoff).
- For a 1, 5, 10, 20, or 50 kb window, KnockoffTrio-FBAT implements:
  1. Weighted burden FBAT for variants with MAF ≥ 0.01.
  2. SNP-based FBAT for variants with MAF ≥ 0.01.
  3. The aggregated Cauchy association test (ACAT)[25] to combine the p values in steps 1 and 2 and obtain $p_{\phi_{kl}}$ and $p^m_{\phi_{kl}}$.

KnockoffTrio-X

The application of KnockoffTrio is not restricted to the FBAT test. Alternatively, p values can be obtained from different, more sophisticated methods that can help increase power in complex scenarios, e.g., the error terms for quantitative traits are not normally distributed. As a proof of concept, we investigate in simulations KnockoffTrio-iQRAT, in which we replace FBAT with the integrated quantile rank test (iQRAT), a gene-level association test that integrates quantile rank score process to accommodate more complex, non-linear associations.[26] iQRAT considers a quantile model for quantitative trait Y:

$$Q_{Y_i}(\tau) = \alpha_0(\tau) + \beta(\tau)^\top X_i^{\text{adj}},$$

where $\tau \in (0,1)$ is the quantile level, $\beta(\tau)^\top = (\beta_1(\tau), \beta_2(\tau), \ldots, \beta_p(\tau))$ is the quantile coefficient functions, $\alpha_0(\tau)$ is the intercept function, and $X_i^{\text{adj}} = X_i - E(X_i|P_i^1, P_i^2)$ is the adjusted offspring genotype where we subtract the conditional expectation (conditional on parental genotypes) so that it corresponds to FBAT formulation. iQRAT tests the null hypothesis $\beta(\tau) = 0, \forall \tau \in (0,1)$. The iQRAT statistics that generalize the sequence kernel association tests (S) and burden tests (B) are computed, respectively, as:

$$Q_S^\phi = S^{\phi\top} W^2 S^\phi,$$

$$Q_B^\phi = S^{\phi\top} W 1_p 1_p^\top W S^\phi,$$

where $S^\phi = n^{-1/2}\sum_{i=1}^n X_i^{\text{adj}\top}\widehat{\phi}_i^\phi$, $\widehat{\phi}_i^\phi = \int_0^1 \widehat{a}_i(\tau)\mathrm{d}\phi(\tau)$, $\widehat{a}_i(\tau) = 1\{Y_i < \widehat{\alpha}_0(\tau)\} - \tau$, $\phi(\tau)$ is the weight function, $\widehat{\alpha}_0(\tau)$ is the estimated intercept via quantile regression under the null, and $W = \text{diag}(w_1, \ldots, w_p)$ is the weight matrix. iQRAT considers four different weight functions and combines the results using ACAT. We use Q_B^ϕ, the burden version of iQRAT, in KnockoffTrio-iQRAT so that it is comparable to the burden FBAT in KnockoffTrio-FBAT.

Knockoff filter procedure for FDR control

For each given window ϕ_{kl}, KnockoffTrio calculates a feature statistic, defined as

$$W_{\phi_{kl}} = \left(T_{\phi_{kl}} - \text{median } T_{\phi_{kl}}^m\right) I_{T_{\phi_{kl}} \geq \max T_{\phi_{kl}}^m}, \qquad \text{(Equation 1)}$$

in which $T_{\phi_{kl}} = -\log_{10} p_{\phi_{kl}}$ and $T_{\phi_{kl}}^m = -\log_{10} p_{\phi_{kl}}^m$ where $p_{\phi_{kl}}$ and $p_{\phi_{kl}}^m$ are the p values computed above for the original and the knockoff trios, respectively. KnockoffTrio then calculates a threshold τ and selects windows with $W_{\phi_{kl}} > \tau$ while controlling the FDR at a target level q. The corresponding value of τ is computed as (see also KnockoffScreen[18]):

$$\tau = \min\left\{ t > 0 : \frac{\frac{1}{M} + \frac{1}{M}\#\{\phi_{kl} : \kappa_{\phi_{kl}} \geq 1, \tau_{\phi_{kl}} \geq t\}}{\#\{\phi_{kl} : \kappa_{\phi_{kl}} = 0, \tau_{\phi_{kl}} \geq t\}} \leq q \right\}, \qquad \text{(Equation 2)}$$

where $\tau_{\phi_{kl}} = T_{\phi_{kl}}^{(0)} - \text{median} T_{\phi_{kl}}^{(m)}$ is the largest importance score minus the median of the remaining importance scores, $\kappa_{\phi_{kl}} = 0$ when $T_{\phi_{kl}}$ is the largest importance score, and $\kappa_{\phi_{kl}} = m$ when $T_{\phi_{kl}}^m$ for the mth knockoff is the largest importance score.

We show a schematic flowchart for KnockoffTrio in Figure 1.

Calculation of q values

We also calculate a q value $q_{\phi_{kl}}$ for ϕ_{kl}, which is the p value analogue in the FDR setting and unifies $W_{\phi_{kl}}$ and τ for declaring significance. Specifically, the q value is the minimum FDR when all tests that show evidence against the null hypothesis at least as strong as the current test are declared as significant. Under the knockoff framework, we follow KnockoffScreen and define the q value for window ϕ as

$$q_\phi = \min_{t \leq \tau_\phi} \frac{\frac{1}{M} + \frac{1}{M}\#\{\phi_{kl} : \kappa_{\phi_{kl}} \geq 1, \tau_{\phi_{kl}} \geq t\}}{\#\{\phi_{kl} : \kappa_{\phi_{kl}} = 0, \tau_{\phi_{kl}} \geq t\}},$$

where $\left(\frac{1}{M} + \frac{1}{M}\#\{\phi_{kl} : \kappa_{\phi_{kl}} \geq 1, \tau_{\phi_{kl}} \geq t\}\right) / \#\{\phi_{kl} : \kappa_{\phi_{kl}} = 0, \tau_{\phi_{kl}} \geq t\}$ is the estimated FDR if we declare significant windows with feature statistics $\kappa_{\phi_{kl}} = 0, \tau_{\phi_{kl}} \geq t$. We define $q_\phi = 1$ for windows with $\kappa_{\phi_{kl}} > 0$ so that they will not be selected. By definition, the windows selected by $W_\phi > \tau$ are equivalent to those selected by $q_\phi < q$, where q is the target FDR.

Meta-analysis for KnockoffTrio

For a variant or set of variants, meta-analysis can be performed by integrating summary statistics from individual studies into a combined summary statistic. KnockoffTrio can be naturally extended to the meta-analysis setting because KnockoffTrio's feature statistics are defined based on summary statistics for the original and the knockoff cohorts. Here, we implement the sample-size-based meta-analysis[27] into KnockoffTrio. Specifically, KnockoffTrio's meta-analysis procedure is defined as follows:

1. For the ith study, obtain $Z_{\phi_{kl},i}$ for a window ϕ_{kl} in the original cohort and $Z_{\phi_{kl},i}^m$ for the same window in the mth knockoff cohort; $Z_{\phi_{kl},i}$ and $Z_{\phi_{kl},i}^m$ are the standardized SNP-based FBAT statistics for a single-variant window or the set-based FBAT statistics for a multi-variant window.
2. Calculate $Z_{\phi_{kl},meta} = \frac{\Sigma_i w_i Z_{\phi_{kl},i}}{\sqrt{\Sigma_i w_i^2}}$ for the original cohort and $Z_{\phi_{kl},meta}^m = \frac{\Sigma_i w_i Z_{\phi_{kl},i}^m}{\sqrt{\Sigma_i w_i^2}}$ for the mth knockoff cohort, in which $w_i = \sqrt{N_i}$ is the weight and N_i is the sample size (i.e., the number of trios) for the ith study.
3. Calculate $p_{\phi_{kl},meta} = 2\Phi(-|Z_{\phi_{kl},meta}|)$ for the original cohort and $p_{\phi_{kl},meta}^m = 2\Phi(-|Z_{\phi_{kl},meta}^m|)$ for the mth knockoff cohort.
4. Calculate $W_{\phi_{kl},meta}$ and τ_{meta} using Equations 1 and 2.

Results

Simulation studies

We simulate genetic data based on the Autism Genome Project (AGP) cohort. The AGP cohort consists of 798,961 common (MAF $\geq$ 0.05) and low-frequency (0.01 $\leq$ MAF $<$ 0.05) variants for 1,266 trio families of European ancestry. For a simulation replicate, we simulate 10,000 trios with common and low-frequency variants sampled from a 1-Mb region (chr20: 15,981,843–16,981,842; 495 variants with MAF $\geq$ 0.01) near *MACROD2*. In line with previous studies,[18,28] we applied hierarchical clustering such that variants from different clusters have correlation no greater than 0.7 and then randomly selected one representative variant from each cluster to be included in the replicate. For a trio, we sampled four haplotypes from the phased AGP data for the parents and simulated the genotypes for the offspring using two of the four haplotypes, each randomly selected from a parent.

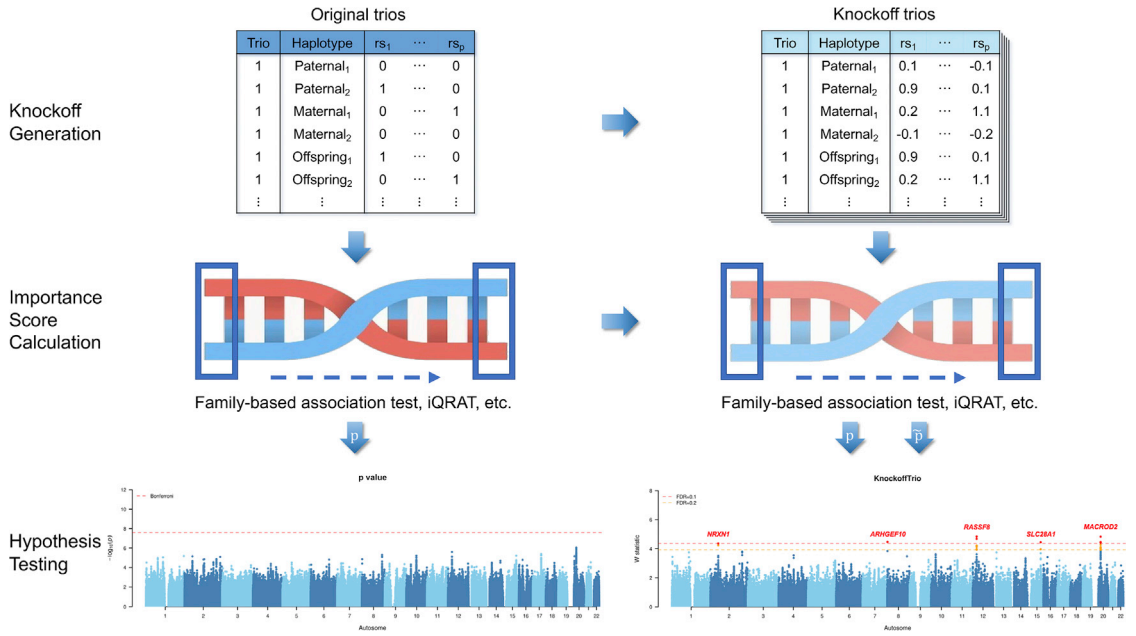

Figure 1. KnockoffTrio workflow
Knockoff generation based on original trios, calculation of importance scores using sliding windows, and examples of hypothesis testing using conventional association testing and KnockoffTrio.

KnockoffTrio preserves exchangeability in trio studies

The rationale of the proposed algorithm is to augment the original trios with synthetic trios. The knockoff construction proposed here ensures the exchangeability property between the original and synthetic genotypes: i.e., if we swap any subset of variants with their synthetic counterparts, the joint haplotype distribution for the trio remains the same (see formal proof in supplemental note 1). This exchangeability property is a necessary condition for the FDR control. We verify the exchangeability for the offspring haplotypes using simulations. We generated a replicate of 10,000 trios with variants sampled from a 1-Mb region as described above. To validate the exchangeability, we generated the offspring knockoff haplotypes using the proposed algorithm in KnockoffTrio and evaluated whether the covariance between each pair of variants is exchangeable for the common variants in the region. As shown in Figure S1, the exchangeability property holds in simulations.

Empirical power and FDR in single-locus simulations

We performed simulations to evaluate the power and empirical FDR of KnockoffTrio. We simulated 500 replicates as described above. We generated the dichotomous trait for the offspring using a logit model:

$$\text{logit}(Y_i) = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip},$$

and the quantitative trait using a linear model:

$$Y_i = \beta_1 X_{i1} + \ldots + \beta_p X_{ip} + \varepsilon_i,$$

where $\varepsilon_i \sim N(0,1)$, β_0 was set such that the disease prevalence is 1% and $\text{logit}(x) = \log\frac{x}{1-x}$. We randomly selected three variants within a 1-kb signal window to be causal with the causal effect $\beta_j = 0.2|\log_{10}\text{MAF}_j|$. For dichotomous traits, we include a trio only when $Y_i = 1$ to mimic the usual ascertainment in real trio design studies with dichotomous traits.

For each replicate, we generated multiple knockoffs ($M =$ 1, 4, 6, 8, and 10) and used several window sizes to scan the region (1 bp and 1, 5, 10, 20, and 50 kb). We evaluated the performance of KnockoffTrio in terms of different numbers of knockoffs for both dichotomous and quantitative traits. For each replicate, the power is the proportion of detected causal windows (i.e., windows that contain at least one causal variant) among all causal windows, and the FDR is the proportion of non-causal windows among all detected windows. The power and FDR were averaged over the 500 replicates. As shown in Figure 2, KnockoffTrio with multiple knockoffs controls the FDR at the target level in all scenarios considered. A slightly inflated FDR for a single knockoff is observed especially for dichotomous traits, which is consistent with previous literature showing inflated FDR for dichotomous traits under highly correlated designs;[15] see supplemental note 7 and Figure S7 for simulations where a single knockoff has no inflated FDR with lowered linkage disequilibrium. The power of KnockoffTrio increases when the number of knockoffs increases, especially at low target FDR levels as expected due to the detection threshold issue mentioned in the material and methods section.

KnockoffTrio prioritizes causal variants over false-positive associations due to linkage disequilibrium

Based on the single-locus simulations, we further compared KnockoffTrio with the conventional association

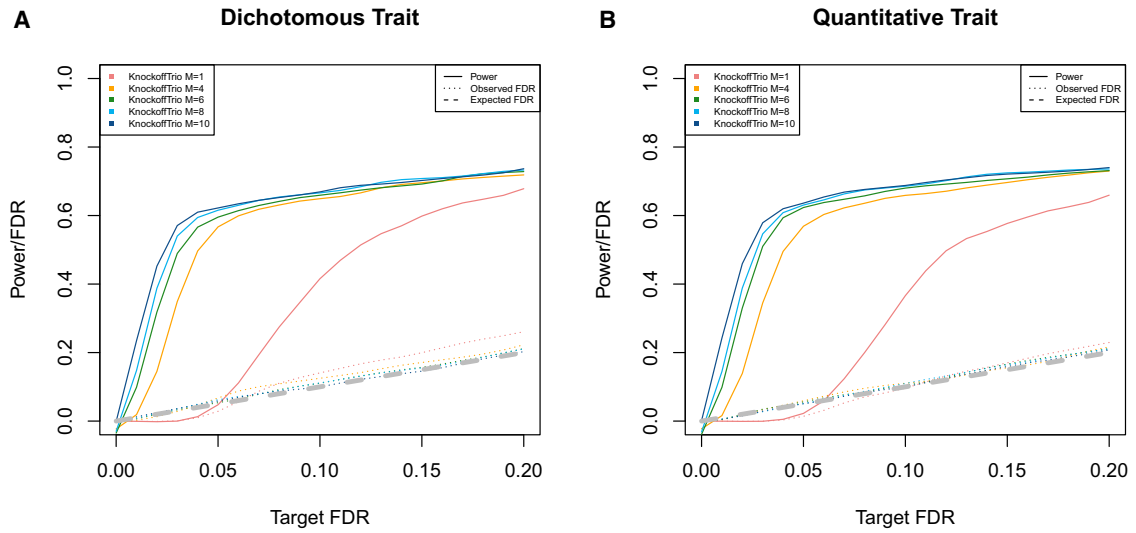

Figure 2. KnockoffTrio's power and FDR in single-locus simulations
(A and B) The two panels show the power and FDR for dichotomous and quantitative traits. We evaluate KnockoffTrio's power and FDR with a target FDR ranging from 0 to 0.2 and with different numbers of knockoffs. The solid lines indicate KnockoffTrio's power, and the dotted lines indicate KnockoffTrio's observed FDR. The different colors indicate different numbers of knockoffs. The gray dashed line indicates the expected FDR.

test that controls the FWER in terms of (1) the proportion of selected windows that overlap with the 1-kb signal window and (2) the median distance of selected windows to the 1-kb signal window. The distance was calculated as the absolute value of the difference between the middle point of a selected window and that of the signal window. For the conventional association test, we used the same aggregated Cauchy association test implemented in KnockoffTrio for each window and controlled the FWER using the Bonferroni correction. As shown in Figures 3A and 3B, the windows selected by KnockoffTrio have a substantially higher chance of overlapping with the signal window and a shorter distance to the signal window than the conventional method. We also randomly selected 200 false positives identified by the conventional association test with Bonferroni correction from all simulated replicates and showed the relationship between their significance and the maximum correlation with any causal variants in the left panel of Figure 3C. As the correlation increases, the conventional association test yields more significant p values for the false positives. On the other hand, for these same 200 variants, KnockoffTrio has a much higher chance of correctly identifying these non-causal variants as true negatives as shown in the right panel of Figure 3C, and thus is substantially more robust in controlling false positives in the presence of linkage disequilibrium between causal and non-causal variants.

Empirical power and FDR in multi-locus simulations in the presence of noise loci

We additionally conducted multi-locus simulations to compare KnockoffTrio with conventional FDR and FWER control methods in the presence of multiple causal and non-causal (noise) loci. We adopted the same simulation method in single-locus simulations to randomly generate 100 1-Mb causal loci and 2,000 200-kb non-causal loci. A causal locus contains a 1-kb signal window, in which three variants were randomly selected to be causal.

We compared KnockoffTrio with $M = 10$ to the Bonferroni correction that controls the FWER and the BH procedure that controls the FDR. Both the Bonferroni correction and the BH procedure were applied to the ACAT-combined p values used to compute importance scores in KnockoffTrio. We also applied the Bonferroni correction to the weighted burden FBAT, a commonly used test in family-based studies. A method's power is the proportion of detected causal windows (i.e., windows that contain at least one causal variant) among all causal windows. We evaluated power at a target FDR of 0.1 for FDR-control methods or a target FWER of 0.05 for FWER-control methods. The empirical FDR is defined as the proportion of non-causal windows at least 50/25/0 kb away from the nearest signal windows among all detected windows. As shown in Figure 4, KnockoffTrio was more powerful than the Bonferroni correction, as expected given the more liberal FDR control, while preserving the FDR at the target level of 0.1. The BH procedure failed to control the FDR at the target level due to the complex correlations among genetic variants. We also note that the FDR for each method decreased as the distance to the signal windows increased. This is expected because the non-causal windows closer to the signal windows are more likely to be false positives due to stronger linkage disequilibrium with variants in the signal windows. Such decrease in FDR is particularly evident for the BH procedure, which is more affected by the correlation among tests.

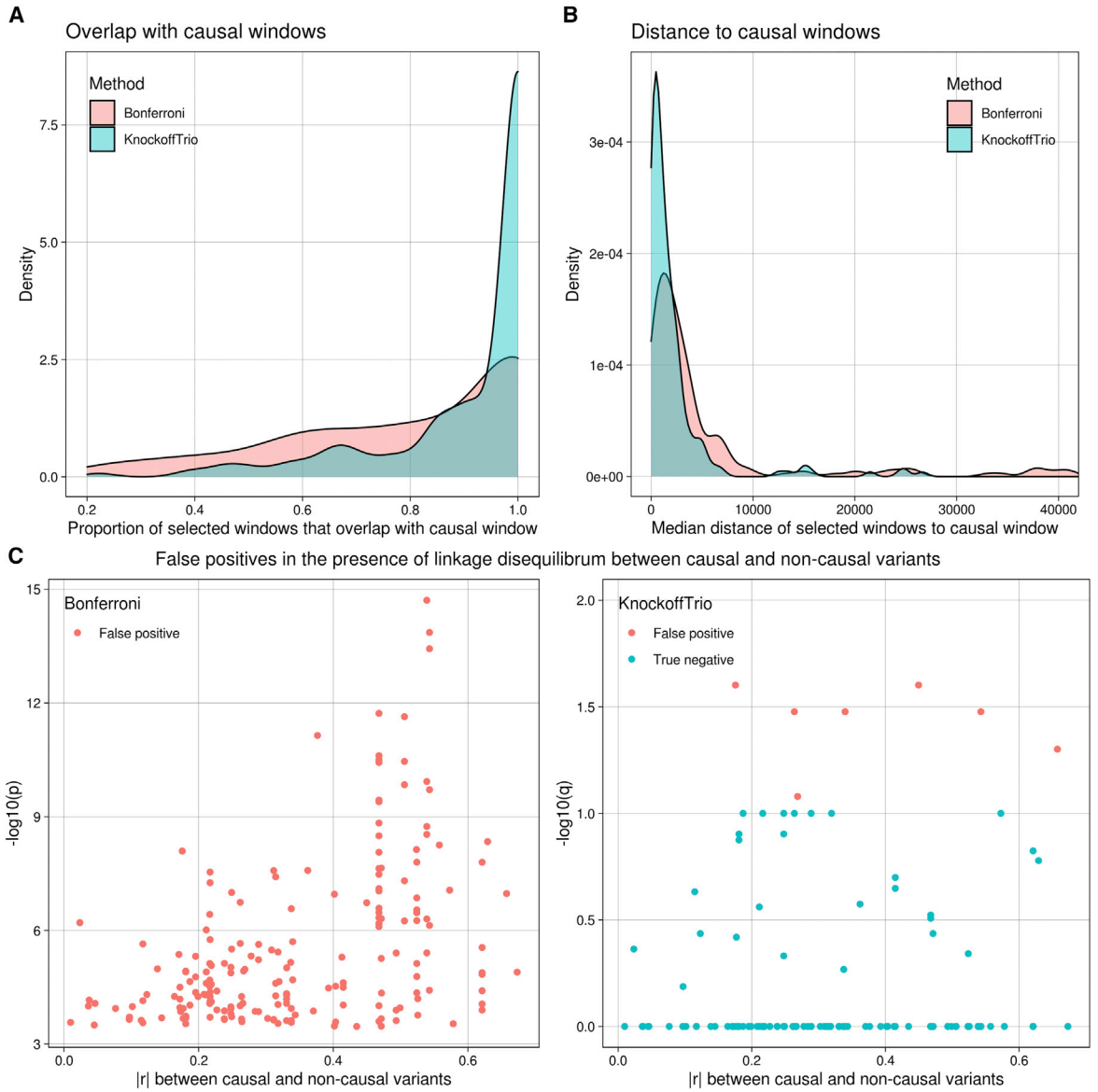

Figure 3. KnockoffTrio prioritizes causal variants over associations due to linkage disequilibrium

(A–C) Comparisons between KnockoffTrio and the conventional method in terms of (A) the proportion of selected windows that overlap with the signal window, (B) the median distance of selected windows to the signal window, and (C) the robustness in controlling false positives in the presence of linkage disequilibrium between causal and non-causal variants. The conventional method is the same aggregated Cauchy association test implemented in KnockoffTrio and controls the FWER using the Bonferroni correction. For KnockoffTrio, the target FDR is 0.1, and the number of multiple knockoffs is 10. The distance in (B) was calculated as the absolute value of the difference between the middle point of a selected window and that of the signal window. The $|r|$ in (C) is the maximum absolute correlation between the false positive and any causal variants. The variants in the right figure in (C) correspond to the variants in the left figure.

KnockoffTrio-iQRAT improves power in detecting complex associations

We performed simulations to compare the power of KnockoffTrio-iQRAT with KnockoffTrio-FBAT in complex scenarios where the normality of quantitative traits is violated. Specifically, we generated quantitative trait values using a location model:

$$Y_i = \beta_1 X_{i1} + \ldots + \beta_p X_{ip} + \varepsilon_i,$$

where $\varepsilon_i \sim \text{Cauchy}(\mu = 0, \gamma = 1)$, μ is the location parameter, and γ is the scale parameter for the Cauchy distribution. We generated 500 replicates, each of which consists of 1,000 trios and 500 variants near *MACROD2*, using the AGP cohort as above. We randomly selected three variants within a 1-kb window to be causal with the causal effect $\beta_j = 1.2|\log_{10}\text{MAF}_j|$. We applied quantile and rank normalization to Y_is before analysis. For KnockoffTrio-iQRAT, to make fair comparisons with FBAT, we only analyzed the offspring data, and we adjusted the offspring genotypes by subtracting the conditional expectation (conditional on parental genotypes), i.e., $X_i - E(X_i|P_i^1, P_i^2)$. As shown in Figure S5, KnockoffTrio-iQRAT is more powerful than KnockoffTrio-FBAT in the scenario with non-Gaussian errors as expected.

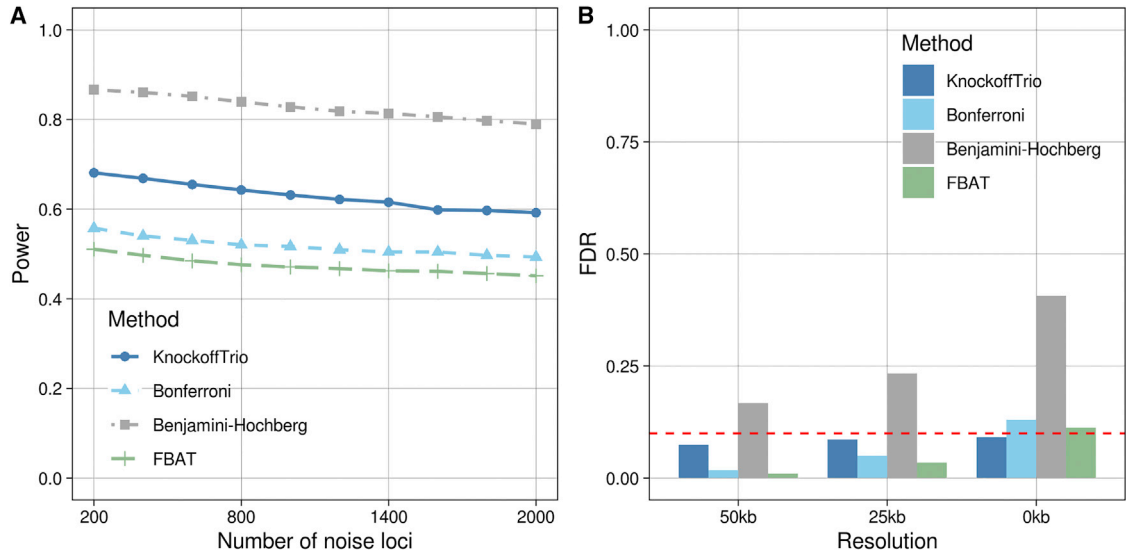

Figure 4. Genome-wide power and FDR in the presence of noise loci
The left panel presents each method's power (target FDR 0.1), defined as the proportion of detected causal windows among all causal windows. A causal window is a window that contains any causal variants. The right panel presents each method's false discovery rate (target FDR 0.1) at different resolutions, defined as the proportion of non-causal windows at least 50/25/0 kb away from the nearest signal windows among all detected windows. Note that KnockoffTrio and Benjamini-Hochberg have more liberal FDR control than Bonferroni and FBAT.

KnockoffTrio provides protection against external confounders such as population stratification

Population stratification is one of the most common confounders in genetic association studies and is often a source of spurious associations when a study cohort has individuals from different populations. We demonstrate KnockoffTrio's robustness in controlling FDR in the presence of population stratification through simulations (see also discussion in supplemental note 1). For a replicate, we simulated 10,000 trios with 500 common and low-frequency variants randomly selected from a 1-Mb region (chr20: 15,981,843–16,981,842) near *MACROD2* using the 1000 Genomes Phase III sequencing data,[29] where haplotype data are available on 2,504 samples across 26 (sub)populations. For these analyses, we focus on 1,006 haplotypes of European origin and 1,096 haplotypes of African origin. In line with previous simulations, we applied hierarchical clustering such that variants from different clusters have correlation no greater than 0.7 and then randomly selected one representative variant from each cluster to be included in the replicate. For 70% of the trios, we sampled four haplotypes from the European population to obtain the parental data in a trio and simulated the offspring genotypes using two of the four haplotypes, each randomly selected from a parent. For the remaining 30% of the trios, we did the same except that we sampled parental haplotypes from the African population. We then generated a quantitative trait using a linear model:

$$Y_i = Z_i + \varepsilon_i,$$

where $\varepsilon_i \sim N(0,1)$, $Z_i = 0$ if the ith trio is from the European population, and $Z_i = 1$ if the ith trio is from the African population. For dichotomous traits, we set $Y_i = 1$ to mimic the usual ascertainment in trio design studies with dichotomous traits.

We evaluated each method's FDR, defined as the proportion of replicates where any window was detected among 500 replicates. As shown in Figure S2, both KnockoffTrio and the conventional family-based methods control the FDR in the presence of population stratification at a target FDR of 0.1 for both dichotomous and quantitative traits.

Applications to trio data on ASDs

To study the risk genetic variants for ASD, we applied KnockoffTrio with multiple knockoffs ($M = 10$) to several ASD cohorts, including the family trio data from the AGP (dbGaP: phs000267.v5.p2)[30] and two cohorts collected by the Simons Foundation Autism Research Initiative (SFARI): the Simons Foundation Powering Autism Research (SPARK)[31] and the Simons Simplex Collection (SSC).[32] The details of the individual cohorts are described below. We have complied with the data-use agreements for each specific site. For comparisons, we also present results from the digital twin test and KnockoffGWAS (supplemental notes 4 and 5 and Tables S1 and S2).

Data descriptions

AGP

Our AGP analysis included 798,961 common (MAF $\geq$ 0.05) and low-frequency (0.01$\leq$ MAF$<$ 0.05) variants for 1,266 trio families of European ancestry, each of which consists of two parents and their offspring diagnosed with strict ASD, i.e., met the criteria for autism on both the Autism Diagnostic Interview-Revised[33] and the Autism Diagnostic Observation Schedule.[34]

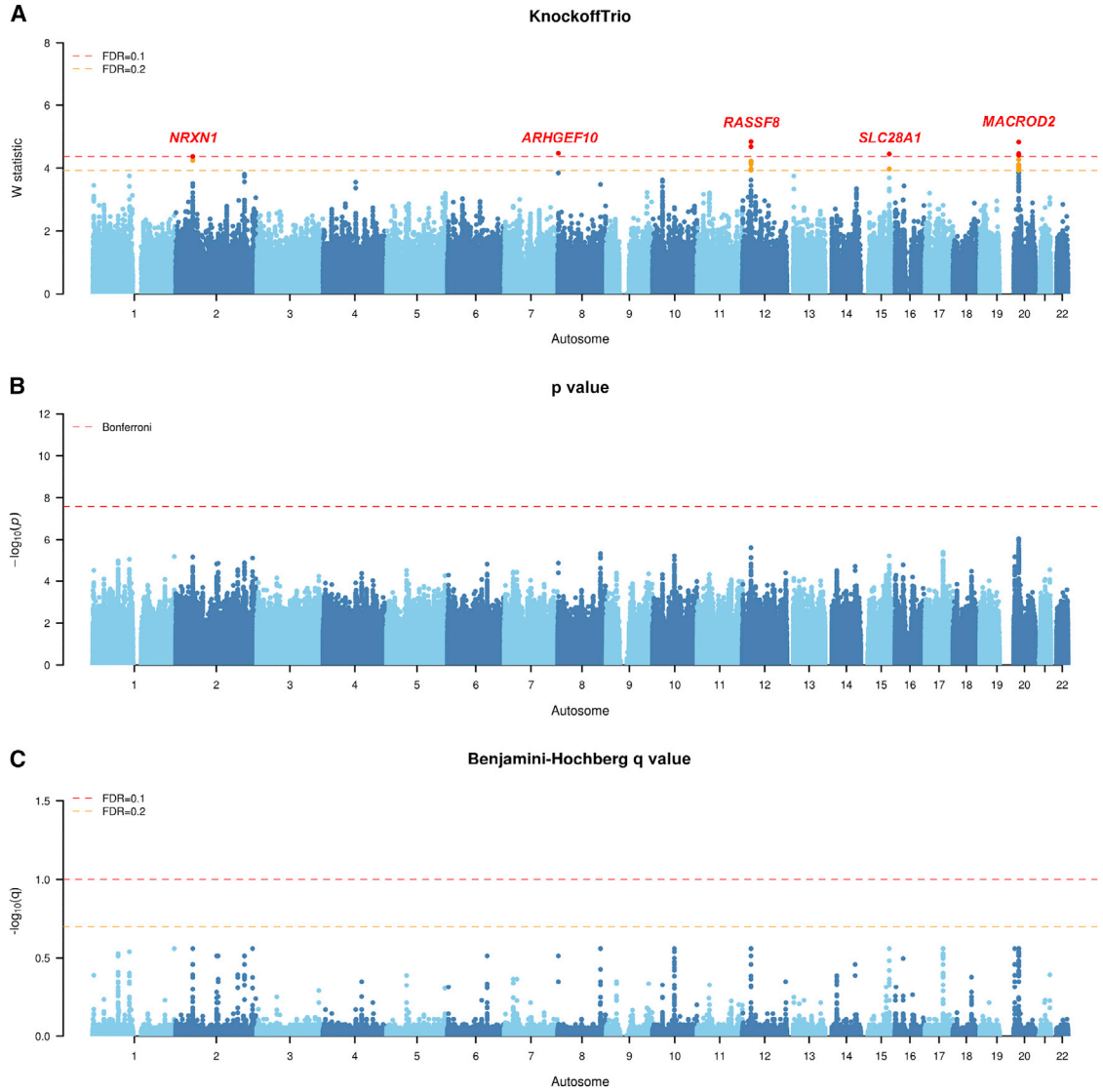

Figure 5. Manhattan plots from KnockoffTrio analysis for the Autism Genome Project (AGP)
The Manhattan plots of the W statistics from KnockoffTrio, the p values from the conventional association tests with the Bonferroni correction for controlling the FWER, and the q values from the BH procedure for controlling the FDR. The FDR target level for KnockoffTrio and the BH procedure is 0.1 or 0.2. Each locus is annotated with the closest gene name.

SPARK

Our SPARK analysis included 10,540 trio families from the first three releases of the SPARK cohort. The probands in the two SFARI cohorts received a professional diagnosis of ASD from a physician, psychologist, or therapist. We have focused on 381,063 common and low-frequency variants.

SSC

Our SSC analysis included 2,394 trio families from the pilot and phases 1, 2, 3-1, and 3-2 studies of the SSC cohort, with whole-genome sequencing data available. We have focused on 5,772,421 common and low-frequency variants.

KnockoffTrio analyses

We adopted a quality control (QC) procedure that excluded variants with MAFs $< 1\%$, missing call rates $> 5\%$, Mendelian error rates $> 0.1\%$, and Hardy-Weinberg equilibrium p values $< 10^{-7}$ for all cohorts. For each cohort, we performed the QC procedure using all available individuals and then broke families into all possible trios (if they were not already trios) for analyses. Genotype data were phased using SHAPEIT2.[35] The genomic coordinates in the AGP data were converted from hg18 to hg38 using the NCBI Genome Remapping Service. We adjusted for gender of offspring in all analyses. We present results from individual cohorts at a target FDR of 0.1 and 0.2 and compared them to the conventional association test with the Bonferroni correction and with the usual BH procedure for FDR control (Figures 5, 6, and 7 and Table 1). We also present results from meta-analyses of the three cohorts (Figure S11).

For the AGP cohort, the conventional association tests (Bonferroni and BH) did not identify any significant

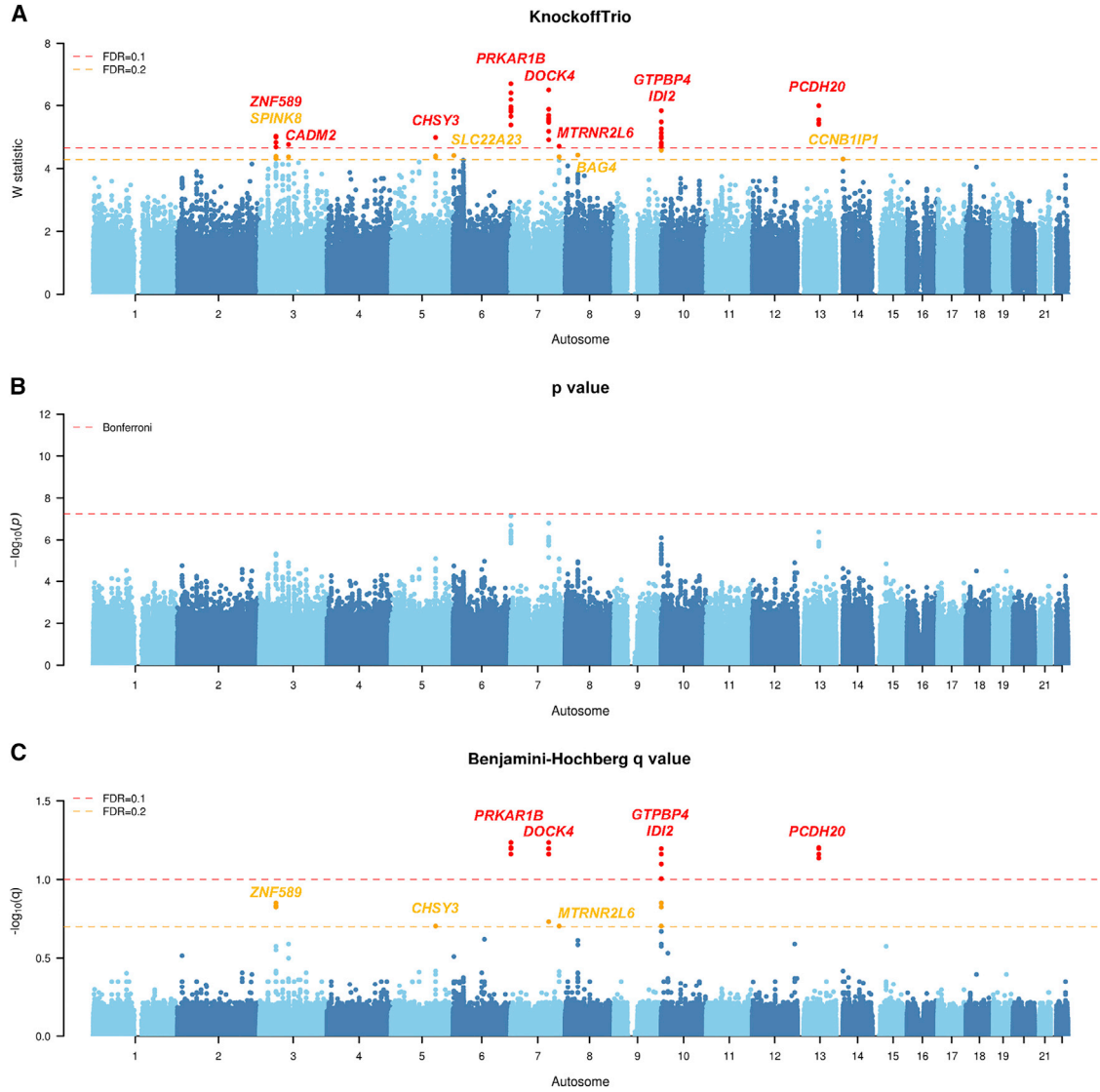

Figure 6. Manhattan plots from KnockoffTrio analysis for the Simons Foundation Powering Autism Research (SPARK)
The Manhattan plots of the W statistics from KnockoffTrio, the p values from the conventional association tests with the Bonferroni correction for controlling the FWER, and the q values from the BH procedure for controlling the FDR. The FDR target level for KnockoffTrio and the BH procedure is 0.1 or 0.2. Each locus is annotated with the closest gene name.

association, whereas KnockoffTrio identified five significant regions, including neurexin 1 (*NRXN1*), rho guanine nucleotide exchange factor 10 (*ARHGEF10*), lamin tail domain containing 1 (*LMNTD1*) - ras association domain family member 8 (*RASSF8*), alpha kinase 3 (*ALPK3*) - solute carrier family 28 member 1 (*SLC28A1*), and mono-ADP ribosylhydrolase 2 (*MACROD2*) at FDR = 0.1 (Figure 5). Among them, *MACROD2* and *NRXN1* have been reported in previous studies as risk genes associated with ASD.[36–39] *ARHGEF10* has been associated with impaired social interaction in mice,[40] one of the main features of ASD. *SLC28A1* has a brain-biased expression and shows an excess of introgressed segments in European and East Asian populations.[41] *SLC28A1* also belongs to the SLC (solute carrier) family, several members of which have previously been associated with behavioral traits (depression, mood disorders, and smoking behavior), autism susceptibility, and attention-deficit/hyperactivity disorder.[41] Furthermore, rs4842996, 8 kb upstream of *SLC28A1*, has been associated with ASD in a meta-analysis of GWAS findings from literature.[42]

For the SPARK cohort, KnockoffTrio identified nine significant loci, including zinc finger protein 589 (*ZNF589*), cell adhesion molecule 2 (*CADM2*), chondroitin sulfate synthase 3 (*CHSY3*) - histidine triad nucleotide binding protein 1 (*HINT1*), platelet derived growth factor subunit A (*PDGFA*) - protein kinase CAMP-dependent type I regulatory subunit beta (*PRKAR1B*), dedicator of cytokinesis 4 (*DOCK4*), MT-RNR2 like 6 (*MTRNR2L6*) - serine protease 1 (*PRSS1*), la ribonucleoprotein 4B (*LARP4B*) - GTP binding protein 4 (*GTPBP4*), isopentenyl-diphosphate delta isomerase 2 (*IDI2*), and protocadherin 20 (*PCDH20*) - protocadherin 9 (*PCDH9*) at FDR = 0.1 and, additionally, serine

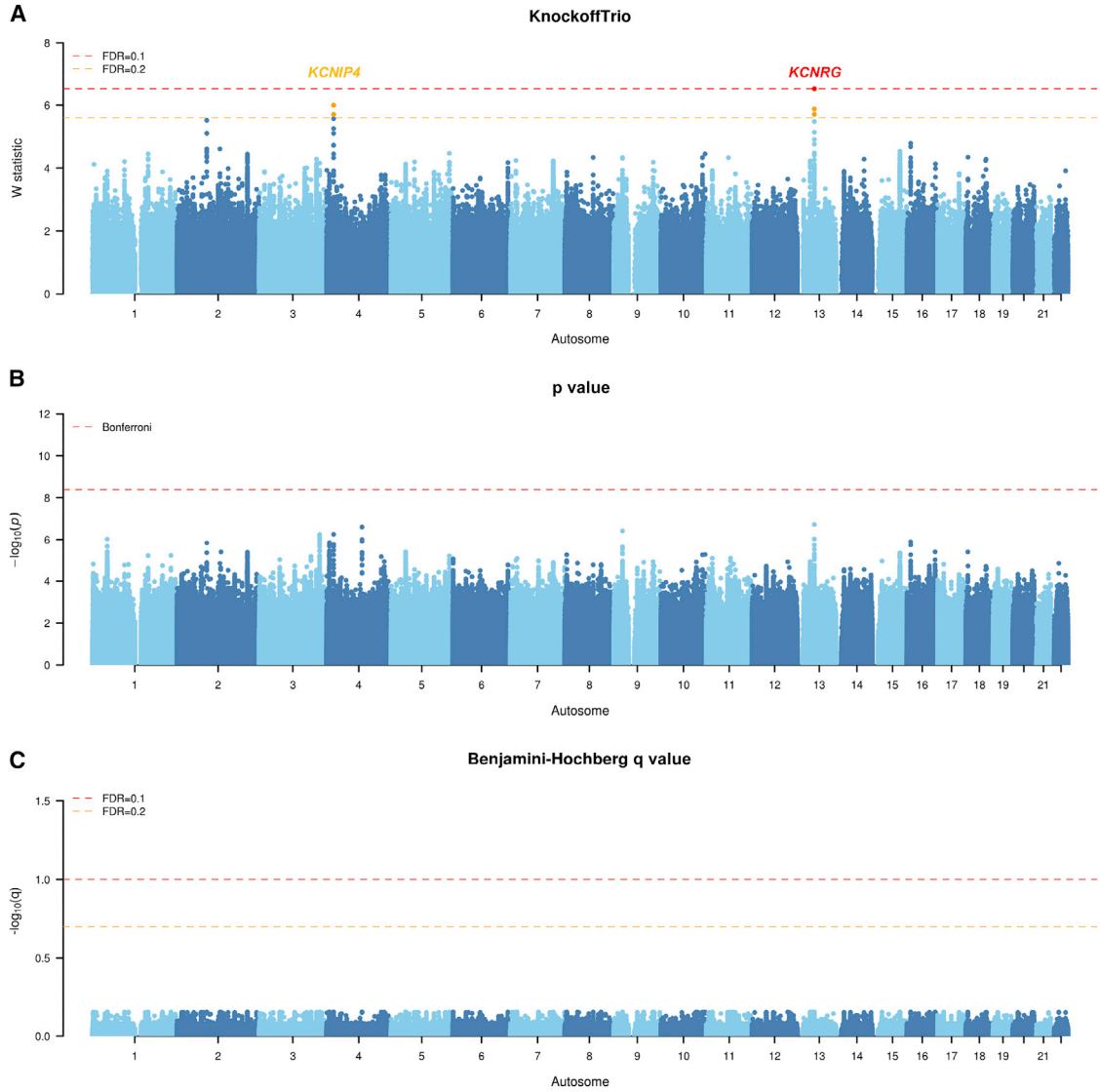

Figure 7. Manhattan plots from KnockoffTrio analysis for the Simons Simplex Collection (SSC)
The Manhattan plots of the W statistics from KnockoffTrio, the p values from the conventional association tests with the Bonferroni correction for controlling the FWER, and the q values from the BH procedure for controlling the FDR. The FDR target level for KnockoffTrio and the BH procedure is 0.1 or 0.2. Each locus is annotated with the closest gene name.

peptidase inhibitor kazal type 8 (*SPINK8*), solute carrier family 22 member 23 (*SLC22A23*)/proteasome assembly chaperone 4 (*PSMG4*), BAG cochaperone 4 (*BAG4*), and cyclin B1 interacting protein 1 (*CCNB1IP1*) - poly (ADP-ribose) polymerase 2 (*PARP2*) at FDR = 0.2 (Figure 6). *PRKAR1B* has been implicated in several neurodevelopmental disorders including ASD.[43–46] Similarly, *CADM2* has been associated with ASD in multiple studies.[47–50] *PCDH9* has been implicated as a genetic risk factor for multiple psychiatric disorders, including major depression[51] and ASD.[52] It is a cell adhesion molecule involved in neuronal migration, synaptic plasticity, and circuit formation. Previous studies have shown that homozygous knockout *PCDH9*-deficient mice have deficits in specific long-term social and object recognition.[53] *DOCK4* has been associated with ASD.[54,55] Furthermore, *DOCK4* knockout mice displayed a series of ASD-like behaviors, including impaired social novelty preference, abnormal isolation-induced pup vocalizations, elevated anxiety, and perturbed object and spatial learning.[56] *BAG4* resides at a locus that has been genome-wide significant in a combined ASD-schizophrenia GWAS.[57] A deleterious variant, c.956T>A (p.Leu319His) (GenBank: NM_016089.3), in *ZNF589* segregated with the phenotype (intellectual disability) and was identified as homozygous in two affected siblings in a consanguineous family from Northern Pakistan;[58] the variant was absent from 200 ethnically matched control individuals. *HINT1* regulates the function of protein kinase C (PKC), which is a prime gene to regulate regression in autism.[59,60] *SPINK8* resides at a GWAS-significant locus associated with multiple psychiatric disorders.[61] In comparison, the conventional association test

Table 1. Genome-wide significant loci from KnockoffTrio analysis

Gene	Chr	Position	Variant	Allele	MAF	p	*Z*	W	q	BH q
AGP (FDR = 0.1)										
NRXN1	2	50805721	rs9284756	A	0.03	7.10E−6	4.49	4.37	0.10	0.28
ARHGEF10	8	1920247–1920676	rs17756915-rs11136442	–	0.41	1.38E−5	–	4.47	0.10	0.31
LMNTD1-RASSF8	12	25946268	rs4963941	A	0.10	2.56E−6	4.70	4.84	0.10	0.28
ALPK3-SLC28A1	15	84881866	rs12917429	T	0.21	6.19E−6	−4.52	4.45	0.10	0.28
MACROD2	20	14781064	rs6074798	A	0.49	1.02E−6	4.89	4.83	0.10	0.28
SFARI: SPARK (FDR = 0.1)										
ZNF589	3	48262179	rs11709691	G	0.28	4.87E−6	−4.57	5.03	0.06	0.14
CADM2	3	85395534–85410981	rs75005531-rs1549979	–	0.22	1.30E−5	–	4.76	0.09	0.26
CHSY3-HINT1	5	130661503	rs17714209	C	0.28	8.25E−6	4.46	4.99	0.06	0.20
PDGFA-PRKAR1B	7	536383	rs62431385	C	0.10	7.20E−8	−5.39	6.71	0.02	0.06
DOCK4	7	111986531	rs73210911	A	0.12	1.59E−7	−5.24	6.51	0.02	0.06
MTRNR2L6-PRSS1	7	142688332	rs13223009	C	0.02	8.42E−6	−4.45	4.71	0.09	0.20
LARP4B-GTPBP4	10	975370	rs117732138	A	0.02	1.60E−6	4.80	5.48	0.02	0.07
IDI2	10	1020654	rs77782977	C	0.02	7.95E−7	4.94	5.84	0.02	0.06
PCDH20-PCDH9	13	63204555	rs12184522	T	0.23	4.21E−7	5.06	6.00	0.02	0.06
SFARI: SPARK (FDR = 0.2)										
SPINK8	3	48316110–48329279	rs74735576-rs13090538	–	0.17	1.58E−5	–	4.39	0.17	0.28
SLC22A23/PSMG4	6	3285062	rs41301847	G	0.02	1.85E−5	4.28	4.41	0.17	0.31
BAG4	8	38205717	rs7836805	A	0.24	2.83E−5	−4.19	4.43	0.17	0.40
CCNB1IP1-PARP2	14	20334133	rs72671266	T	0.02	2.45E−5	−4.22	4.30	0.19	0.38
SFARI: SSC (FDR = 0.1)										
KCNRG-DLEU7	13	50197099	rs2703087	A	0.04	1.88E−7	5.21	6.54	0.10	0.70
SFARI: SSC (FDR = 0.2)										
KCNIP4	4	20917151	rs185413018	T	0.02	5.59E−7	5.00	6.00	0.13	0.70

Only the top signal is shown if multiple signals were identified for a locus. Gene: A single gene name indicates the signal is within or overlaps with the gene. "Gene1/Gene2" indicates the signal overlaps with two genes. "Gene1-Gene2" indicates the signal is between two genes. MAF: minor allele frequency of a variant, or average minor allele frequency if a signal contains multiple variants. p: KnockoffTrio's ACAT-combined p values. For single variants, ACAT-combined p values are equivalent to FBAT p values. *Z*: FBAT *Z* scores for single variants. W: KnockoffTrio's feature statistics. q: KnockoffTrio's q values. BH q: Benjamini-Hochberg q values.

(BH) identified five loci (*PDGFA-PRKAR1B*, *DOCK4*, *LARP4B-GTPBP4*, *IDI2*, and *PCDH20-PCDH9*) at FDR = 0.1 and three loci (*ZNF589*, *CHSY3-HINT1*, and *MTRNR2L6-PRSS1*) at FDR = 0.2, all of which have been identified by KnockoffTrio as well.

For the SSC cohort, KnockoffTrio identified potassium channel regulator (*KCNRG*) - deleted in lymphocytic leukemia 7 (*DLEU7*) at FDR = 0.1 and, additionally, potassium voltage-gated channel interacting protein 4 (*KCNIP4*) at FDR = 0.2 (Figure 7). The finding of *KCNRG*, a gene in the potassium channel tetramerization domain (KCTD) family, provides further evidence for the role of KCTD family in neurodevelopmental and neuropsychiatric disorders.[62] *KCNIP4* is a gene with the largest number of differential RNA-editing sites that have been suggested for aberrant synaptic formation in ASD[63]; variants in *KCNIP4* have also been associated with nonverbal communication and social skills in ASD twins.[64] In comparison, the conventional association tests identified no significant loci.

Meta-analysis

We conducted meta-analysis of the AGP, SPARK, and SSC cohorts as a proof-of-principle because there are known differences across these studies in terms of phenotype definition (for example, AGP uses a very strict ASD definition) and study design (for example, SSC is expected to be enriched in *de novo* variants given its focus on discordant sibs), and such heterogeneity between cohorts makes it difficult to draw overall conclusions.[65] KnockoffTrio identified one significant locus, *DOCK4*, at FDR = 0.2 and, additionally, RANBP2 like and GRIP domain containing 2 (*RGPD2*), *KCNIP4*, and *CHSY3-HINT1* at FDR = 0.4

(Figure S11). In comparison, the conventional association tests identified no significant associations.

Replicability of analyses

Given the random nature of the knockoff procedure, we have attempted to assess the replicability of the results by re-analyzing the individual cohorts with different random seeds for knockoff generation. As shown in Figures S8–S10, the replications produced results that are in good concordance with the original results. For the AGP cohort, the replication analysis identified *NRXN1*, *ARHGEF10*, *LMNTD1-RASSF8*, and *MACROD2*, all of which were identified in the original analysis. For the SPARK cohort, the replication analysis identified *ZNF589*, *SPINK8*, *CHSY3-HINT1*, *PDGFA-PRKAR1B*, *DOCK4*, *MTRNR2L6-PRSS1*, *LARP4B-GTPBP4*, *IDI2*, and *PCDH20-PCDH9*, all of which were identified in the original analysis. For the SSC cohort, the replication analysis identified *RGPD2*, *KCNIP4*, and *KCNRG-DLEU7*, the latter two of which were identified in the original analysis. This shows the replicability of results from KnockoffTrio despite the randomness in knockoff generation.

Discussion

We propose KnockoffTrio, an association test with trio design for GWAS data built upon the knockoff framework. As an FDR-controlling procedure that accounts for arbitrary correlation structure, KnockoffTrio has been shown in both simulations and real data analyses to be more powerful than the conventional FWER-controlling methods while possessing better FDR control than the conventional FDR-controlling methods such as BH. We have also shown that KnockoffTrio protects against bias induced by population substructure using simulations and heuristic arguments. Furthermore, an important advantage of KnockoffTrio is that it can leverage more sophisticated machine learning methods to model the association between genotypes and phenotypes while maintaining valid FDR control and with potential increases in power. These properties make KnockoffTrio an appealing and promising strategy for the analysis of trio designs for which conventional methods are known to be underpowered.

Although we have focused the current manuscript on the complete trio design, the method can be naturally extended to more complex scenarios. In particular, KnockoffTrio can handle missing parental data and is robust to phasing errors in haplotypes as shown in supplemental note 6 and Figure S6. Furthermore, KnockoffTrio can be applied to large pedigrees by breaking each pedigree into all possible trios and applying KnockoffTrio on the individual trios. The method can also be extended to combine trios and population-based designs. For example, we can obtain the estimated coefficient $\widehat{\beta}_j$ for variant j from the external population-based GWAS and use it as weight w_j in the weighted FBAT when constructing the importance scores. Alternatively, we can perform knockoff analysis for population-based data as in He et al.[18] and use a meta-analysis approach as discussed in the material and methods section to combine the trio and population-based results. Note that this alternative approach is no longer robust to confounding due to population structure. Transfer learning methods that leverage information from such external population-based data could also be of interest.[66]

KnockoffTrio has been implemented in a computationally efficient R package. The runtime for completing the analyses of the AGP, SPARK, and SSC cohorts with 10 knockoffs is 8, 46, and 173 min, respectively, with 1,000 parallel jobs performed in a high-performance computing cluster environment of Intel(R) Xeon(R) CPU E5-2630 0 @ 2.30 GHz. In KnockoffTrio we have adopted the knockoff construction in KnockoffScreen, which has been shown to be a valid knockoff construction that is computationally efficient and can be applied to rare variants, but other, more sophisticated knockoff construction methods such as KnockoffZoom[17] can be applied as well. This demonstrates that KnockoffTrio is a highly scalable method and can be effectively used for any large-scale datasets in whole-genome sequencing studies.

KnockoffTrio reduces the randomness in the knockoff generation by using a multiple-knockoff generation procedure. As shown in the simulations and real-data applications, KnockoffTrio with 10 knockoffs is more powerful at lower target FDR levels than using a single knockoff and has good replicability in terms of identifying significant loci. In our experience, stronger signals are more likely to be reproduced across different runs, though results can be more variable for weaker signals. Further increasing M would help with the reproducibility for weak signals at the cost of lowered computational efficiency, which is a tradeoff that researchers should be aware of. Although the gain in power diminishes as the number of knockoffs increases, especially at larger target FDR levels, given the computational efficiency of KnockoffTrio, we recommend that researchers generate multiple knockoffs for improved reproducibility and potentially better power at stricter FDR targets.

To simplify the inference about the transmission pattern, KnockoffTrio assumes no recombination events given a 200-kb region. However, KnockoffTrio can be extended to handle recombination events at the cost of more complex construction of offspring knockoffs, which may potentially help improve performance. In addition to the haplotype-based knockoff generation algorithm that KnockoffTrio adopts, another possible approach is to use summary statistics and apply knockoff-based methods for summary statistics directly instead of generating knockoffs for individual trio data.[67] We leave these potential extensions to future studies.

GWASs with family-based designs are appealing due to their built-in robustness to population substructure, but they are underpowered due to limited sample sizes, much smaller than for GWASs with unrelated individuals. KnockoffTrio provides a more powerful alternative to classical FBATs in this setting while maintaining robustness to confounding due to population substructure. Furthermore,

by design, KnockoffTrio reduces the confounding effect of linkage disequilibrium and prioritizes causal variants over associations due to linkage disequilibrium.

We have focused our applications to genetic studies of ASD, a highly heterogeneous and complex genetic disease. Despite these challenges, KnockoffTrio has identified some well-known (i.e., robustly identified in previous ASD studies) signals such as *MACROD2*, *ARHGEF10*, and *NRXN1* in AGP; *CADM2*, *PRKAR1B*, *DOCK4*, and *PCDH20* in SPARK; and *KCNIP4* in SSC, suggesting that KnockoffTrio can have more power than conventional tests. Although the consistency across the different cohorts is low, that is not unexpected given the above-mentioned heterogeneity, the generally low power for each individual study with modest sample sizes, and inherent differences across studies in terms of phenotype definition (for example, AGP uses a very strict ASD definition) and study design (for example, SSC is expected to be enriched in *de novo* variants given its focus on discordant sibs).

In summary, KnockoffTrio provides a computationally efficient and more powerful association test for trio designs relative to commonly used family-based tests and has the added benefit of reducing confounding due to linkage disequilibrium. The method has been implemented in an R package.

Data and code availability

KnockoffTrio has been implemented in an R package available at https://cran.r-project.org/web/packages/KnockoffTrio. Researchers can apply for the AGP (dbGaP: phs000267.v5.p2) dataset at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000267.v5.p2 and the SPARK and the SSC datasets at https://base.sfari.org/.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2022.08.013.

Acknowledgments

This research was supported by NIH/National Institute of Mental Health Awards MH106910 and MH095797 (to I.I.-L.). We appreciate obtaining access to genetic and phenotypic data from dbGaP and SFARI Base and gratefully acknowledge the participants who provided data for the AGP, SPARK, and SSC projects.

Declaration of interests

The authors declare no competing interests.

Received: March 11, 2022
Accepted: August 24, 2022
Published: September 22, 2022

References

1. Al-Mubarak, B., Abouelhoda, M., Omar, A., AlDhalaan, H., Aldosari, M., Nester, M., Alshamrani, H.A., El-Kalioby, M., Goljan, E., Albar, R., et al. (2017). Whole exome sequencing reveals inherited and de novo variants in autism spectrum disorder: a trio study from saudi families. Sci. Rep. *7*, 5679–5714.
2. Wassink, T.H., Piven, J., Vieland, V.J., Huang, J., Swiderski, R.E., Pietila, J., Braun, T., Beck, G., Folstein, S.E., Haines, J.L., and Sheffield, V.C. (2001). Evidence supporting wnt2 as an autism susceptibility gene. Am. J. Med. Genet. *105*, 406–413.
3. O'Roak, B.J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J.J., Girirajan, S., Karakoc, E., MacKenzie, A.P., Ng, S.B., Baker, C., et al. (2011). Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat. Genet. *43*, 585–589.
4. Laird, N.M., and Lange, C. (2009). The role of family-based designs in genome-wide association studies. Stat. Sci. *24*, 388–397.
5. Laird, N.M., and Lange, C. (2006). Family-based designs in the age of large-scale gene-association studies. Nat. Rev. Genet. *7*, 385–394.
6. Kong, A., Thorleifsson, G., Frigge, M.L., Vilhjalmsson, B.J., Young, A.I., Thorgeirsson, T.E., Benonisdottir, S., Oddsson, A., Halldorsson, B.V., Masson, G., et al. (2018). The nature of nurture: Effects of parental genotypes. Science *359*, 424–428.
7. Price, A.L., Zaitlen, N.A., Reich, D., and Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. Nat. Rev. Genet. *11*, 459–463.
8. Chen, H., Huffman, J.E., Brody, J.A., Wang, C., Lee, S., Li, Z., Gogarten, S.M., Sofer, T., Bielak, L.F., Bis, J.C., et al. (2019). Efficient variant set mixed model association tests for continuous and binary traits in large-scale whole-genome sequencing studies. Am. J. Hum. Genet. *104*, 260–274.
9. Zhou, W., Zhao, Z., Nielsen, J.B., Fritsche, L.G., LeFaive, J., Gagliano Taliun, S.A., Bi, W., Gabrielsen, M.E., Daly, M.J., Neale, B.M., et al. (2020). Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts. Nat. Genet. *52*, 634–639.
10. Bates, S., Sesia, M., Sabatti, C., and Candès, E. (2020). Causal inference in genetic trio studies. Proc. Natl. Acad. Sci. USA. *117*, 24117–24126.
11. Nelson, C.P., Goel, A., Butterworth, A.S., Kanoni, S., Webb, T.R., Marouli, E., Zeng, L., Ntalla, I., Lai, F.Y., Hopewell, J.C., et al. (2017). Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nat. Genet. *49*, 1385–1391.
12. Sesia, M., Bates, S., Candès, E., Marchini, J., and Sabatti, C. (2020). Controlling the False Discovery Rate in Gwas with Population Structure. Preprint at bioRxiv. https://doi.org/10.1101/2020.08.04.236703.
13. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., et al. (2020). Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell *180*, 568–584.e23.
14. De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature *515*, 209–215.
15. Candès, E., Fan, Y., Janson, L., Lv, J., et al. (2018). Panning for gold: Model-x knockoffs for high-dimensional controlled variable selection. J. R. Stat. Soc. B *80*, 551–577.
16. Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Ann. Stat., 1165–1188.

17. Sesia, M., Katsevich, E., Bates, S., Candès, E., and Sabatti, C. (2020). Multi-resolution localization of causal variants across the genome. Nat. Commun. *11*, 1799.
18. He, Z., Liu, L., Wang, C., Le Guen, Y., Lee, J., Gogarten, S., Lu, F., Montgomery, S., Tang, H., Silverman, E.K., et al. (2021). Identification of putative causal loci in whole-genome sequencing data via knockoff statistics. Nat. Commun. *12*, 3152–3218.
19. Sesia, M., Bates, S., Candès, E., Marchini, J., and Sabatti, C. (2021). False discovery rate control in genome-wide association studies with population structure. Proc. Natl. Acad. Sci. USA. *118*. e2105841118.
20. Spielman, R.S., McGinnis, R.E., and Ewens, W.J. (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (iddm). Am. J. Hum. Genet. *52*, 506–516.
21. Chen, H., Meigs, J.B., and Dupuis, J. (2013). Sequence kernel association test for quantitative traits in family samples. Genet. Epidemiol. *37*, 196–204.
22. Yan, Q., Tiwari, H.K., Yi, N., Gao, G., Zhang, K., Lin, W.-Y., Lou, X.-Y., Cui, X., and Liu, N. (2015). A sequence kernel association test for dichotomous traits in family samples under a generalized linear mixed model. Hum. Hered. *79*, 60–68.
23. Marchini, J., Cutler, D., Patterson, N., Stephens, M., Eskin, E., Halperin, E., Lin, S., Qin, Z.S., Munro, H.M., Abecasis, G.R., et al. (2006). A comparison of phasing algorithms for trios and unrelated individuals. Am. J. Hum. Genet. *78*, 437–450.
24. De, G., Yip, W.-K., Ionita-Laza, I., and Laird, N. (2013). Rare variant analysis for family-based design. PLoS One *8*, e48495.
25. Liu, Y., Chen, S., Li, Z., Morrison, A.C., Boerwinkle, E., and Lin, X. (2019). Acat: A fast and powerful p value combination method for rare-variant analysis in sequencing studies. Am. J. Hum. Genet. *104*, 410–421.
26. Wang, T., Ionita-Laza, I., and Wei, Y. (2022). Integrated quantile rank test (iqrat) for gene-level associations. Ann. Appl. Stat. *16*, 1423–1444.
27. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). Metal: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.
28. Sesia, M., Sabatti, C., and Candès, E.J. (2019). Gene hunting with hidden Markov model knockoffs. Biometrika *106*, 1–18.
29. 1000 Genomes Project Consortium, Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74.
30. Autism Genome Project Consortium, Szatmari, P., Paterson, A.D., Zwaigenbaum, L., Roberts, W., Brian, J., Liu, X.-Q., Vincent, J.B., Skaug, J.L., Thompson, A.P., et al. (2007). Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat. Genet. *39*, 319–328.
31. SPARK Consortium Electronic address pfeliciano@simonsfoundationorg (2018). Spark: A US cohort of 50, 000 families to accelerate autism research. Neuron *97*, 488–493.
32. Fischbach, G.D., and Lord, C. (2010). The simons simplex collection: a resource for identification of autism genetic risk factors. Neuron *68*, 192–195.
33. Lord, C., Rutter, M., and Le Couteur, A. (1994). Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J. Autism Dev. Disord. *24*, 659–685.
34. Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Jr., Leventhal, B.L., DiLavore, P.C., Pickles, A., and Rutter, M. (2000). The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. J. Autism Dev. Disord. *30*, 205–223.
35. Delaneau, O., Marchini, J., and Zagury, J.-F. (2011). A linear complexity phasing method for thousands of genomes. Nat. Methods *9*, 179–181.
36. Anney, R., Klei, L., Pinto, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., Sykes, N., Pagnamenta, A.T., et al. (2010). A genome-wide scan for common alleles affecting risk for autism. Hum. Mol. Genet. *19*, 4072–4082.
37. Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R., et al. (2019). Identification of common genetic risk variants for autism spectrum disorder. Nat. Genet. *51*, 431–444.
38. Gauthier, J., Siddiqui, T.J., Huashan, P., Yokomaku, D., Hamdan, F.F., Champagne, N., Lapointe, M., Spiegelman, D., Noreau, A., Lafrenière, R.G., et al. (2011). Truncating mutations in nrxn2 and nrxn1 in autism spectrum disorders and schizophrenia. Hum. Genet. *130*, 563–573.
39. Kim, H.-G., Kishikawa, S., Higgins, A.W., Seong, I.-S., Donovan, D.J., Shen, Y., Lally, E., Weiss, L.A., Najm, J., Kutsche, K., et al. (2008). Disruption of neurexin 1 associated with autism spectrum disorder. Am. J. Hum. Genet. *82*, 199–207.
40. Lu, D.-H., Liao, H.-M., Chen, C.-H., Tu, H.-J., Liou, H.-C., Gau, S.S.-F., and Fu, W.-M. (2018). Impairment of social behaviors in arhgef10 knockout mice. Mol. Autism. *9*, 11.
41. Gouy, A., and Excoffier, L. (2020). Polygenic patterns of adaptive introgression in modern humans are mainly shaped by response to pathogens. Mol. Biol. Evol. *37*, 1420–1433.
42. Lee, J., Son, M.J., Son, C.Y., Jeong, G.H., Lee, K.H., Lee, K.S., Ko, Y., Kim, J.Y., Lee, J.Y., Radua, J., et al. (2020). Genetic variation and autism: A field synopsis and systematic meta-analysis. Brain Sci. *10*, E692.
43. Marbach, F., Stoyanov, G., Erger, F., Stratakis, C.A., Settas, N., London, E., Rosenfeld, J.A., Torti, E., Haldeman-Englert, C., Sklirou, E., et al. (2021). Variants in prkar1b cause a neurodevelopmental disorder with autism spectrum disorder, apraxia, and insensitivity to pain. Genet. Med. *23*, 1465–1473.
44. Ruzzo, E.K., Pérez-Cano, L., Jung, J.-Y., Wang, L.-K., Kashef-Haghighi, D., Hartl, C., Singh, C., Xu, J., Hoekstra, J.N., Leventhal, O., et al. (2019). Inherited and de novo genetic risk for autism impacts shared networks. Cell *178*, 850–866.e26.
45. Turner, T.N., Hormozdiari, F., Duyzend, M.H., McClymont, S.A., Hook, P.W., Iossifov, I., Raja, A., Baker, C., Hoekzema, K., Stessman, H.A., et al. (2016). Genome sequencing of autism-affected families reveals disruption of putative noncoding regulatory dna. Am. J. Hum. Genet. *98*, 58–74.
46. Chen, S., Zhou, X., Byington, E., Bruce, S.L., Zhang, H., and Shen, Y. (2020). Dissecting Autism Genetic Risk Using Single-Cell Rna-Seq Data. Preprint at bioRxiv. https://doi.org/10.1101/2020.06.15.153031.
47. Casey, J.P., Magalhaes, T., Conroy, J.M., Regan, R., Shah, N., Anney, R., Shields, D.C., Abrahams, B.S., Almeida, J., Bacchelli, E., et al. (2012). A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. Hum. Genet. *131*, 565–579.
48. Namjou, B., Marsolo, K., Caroll, R.J., Denny, J.C., Ritchie, M.D., Verma, S.S., Lingren, T., Porollo, A., Cobb, B.L., Perry, C., et al. (2014). Phenome-wide association study (phewas) in emr-linked pediatric cohorts, genetically links plcl1 to

speech language development and il5-il13 to eosinophilic esophagitis. Front. Genet. *5*, 401.
49. Gamsiz, E.D., Viscidi, E.W., Frederick, A.M., Nagpal, S., Sanders, S.J., Murtha, M.T., Schmidt, M., Simons Simplex Collection Genetics Consortium, Triche, E.W., Geschwind, D.H., et al. (2013). Intellectual disability is associated with increased runs of homozygosity in simplex autism. Am. J. Hum. Genet. *93*, 103–109.
50. Calderoni, S., Ricca, I., Balboni, G., Cagiano, R., Cassandrini, D., Doccini, S., Cosenza, A., Tolomeo, D., Tancredi, R., Santorelli, F.M., and Muratori, F. (2020). Evaluation of chromosome microarray analysis in a large cohort of females with autism spectrum disorders: a single center italian study. J. Pers. Med. *10*, 160.
51. Xiao, X., Zheng, F., Chang, H., Ma, Y., Yao, Y.-G., Luo, X.-J., and Li, M. (2018). The gene encoding protocadherin 9 (pcdh9), a novel risk factor for major depressive disorder. Neuropsychopharmacology *43*, 1128–1137.
52. Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., et al. (2008). Structural variation of chromosomes in autism spectrum disorder. Am. J. Hum. Genet. *82*, 477–488.
53. Bruining, H., Matsui, A., Oguro-Ando, A., Kahn, R.S., Van't Spijker, H.M., Akkermans, G., Stiedl, O., van Engeland, H., Koopmans, B., van Lith, H.A., et al. (2015). Genetic mapping in mice reveals the involvement of pcdh9 in long-term social and object recognition and sensorimotor development. Biol. Psychiatry *78*, 485–495.
54. Maestrini, E., Pagnamenta, A.T., Lamb, J.A., Bacchelli, E., Sykes, N.H., Sousa, I., Toma, C., Barnby, G., Butler, H., Winchester, L., et al. (2010). High-density snp association study and copy number variation analysis of the auts1 and auts5 loci implicate the immp2l-dock4 gene region in autism susceptibility. Mol. Psychiatry *15*, 954–968.
55. Pagnamenta, A.T., Bacchelli, E., de Jonge, M.V., Mirza, G., Scerri, T.S., Minopoli, F., Chiocchetti, A., Ludwig, K.U., Hoffmann, P., Paracchini, S., et al. (2010). Characterization of a family with rare deletions in cntnap5 and dock4 suggests novel risk loci for autism and dyslexia. Biol. Psychiatry *68*, 320–328.
56. Guo, D., Peng, Y., Wang, L., Sun, X., Wang, X., Liang, C., Yang, X., Li, S., Xu, J., Ye, W.-C., et al. (2021). Autism-like social deficit generated by dock4 deficiency is rescued by restoration of rac1 activity and nmda receptor function. Mol. Psychiatry *26*, 1505–1519.
57. The Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium (2017). Meta-analysis of gwas of over 16, 000 individuals with autism spectrum disorder highlights a novel locus at 10q24. 32 and a significant overlap with schizophrenia. Mol. Autism. *8*, 1–17.
58. Agha, Z., Iqbal, Z., Azam, M., Ayub, H., Vissers, L.E.L.M., Gilissen, C., Ali, S.H.B., Riaz, M., Veltman, J.A., Pfundt, R., et al. (2014). Exome sequencing identifies three novel candidate genes implicated in intellectual disability. PLoS One *9*, e112687.
59. Bemben, M.A., Nguyen, Q.-A., Wang, T., Li, Y., Nicoll, R.A., and Roche, K.W. (2015). Autism-associated mutation inhibits protein kinase c-mediated neuroligin-4x enhancement of excitatory synapses. Proc. Natl. Acad. Sci. USA. *112*, 2551–2556.
60. Ji, L., Chauhan, A., and Chauhan, V. (2012). Reduced activity of protein kinase c in the frontal cortex of subjects with regressive autism: relationship with developmental abnormalities. Int. J. Biol. Sci. *8*, 1075–1084.
61. Schork, A.J., Won, H., Appadurai, V., Nudel, R., Gandal, M., Delaneau, O., Revsbech Christiansen, M., Hougaard, D.M., Bækved-Hansen, M., Bybjerg-Grauholm, J., et al. (2019). A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment. Nat. Neurosci. *22*, 353–361.
62. Teng, X., Aouacheria, A., Lionnard, L., Metz, K.A., Soane, L., Kamiya, A., and Hardwick, J.M. (2019). Kctd: A new gene family involved in neurodevelopmental and neuropsychiatric disorders. CNS Neurosci. Ther. *25*, 887–902.
63. Tran, S.S., Jun, H.-I., Bahn, J.H., Azghadi, A., Ramaswami, G., Van Nostrand, E.L., Nguyen, T.B., Hsiao, Y.-H.E., Lee, C., Pratt, G.A., et al. (2019). Widespread rna editing dysregulation in brains from autistic individuals. Nat. Neurosci. *22*, 25–36.
64. Hu, V.W., Devlin, C.A., and Debski, J.J. (2019). Asd phenotype-genotype associations in concordant and discordant monozygotic and dizygotic twins stratified by severity of autistic traits. Int. J. Mol. Sci. *20*, E3804.
65. Higgins, J.P.T., and Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. Stat. Med. *21*, 1539–1558.
66. Li, S., Ren, Z., Sabatti, C., and Sesia, M. (2021). Transfer learning in genome-wide association studies with knockoffs. Preprint at arXiv. https://doi.org/10.48550/arXiv.2108.08813.
67. He, Z., Liu, L., Belloy, M.E., Le Guen, Y., Sossin, A., Liu, X., Qi, X., Ma, S., Wyss-Coray, T., Tang, H., et al. (2021). Summary statistics knockoff inference empowers identification of putative causal variants in genome-wide association studies. Preprint at bioRxiv. https://doi.org/10.1101/2021.12.06.471440.

The American Journal of Human Genetics, Volume *109*

Supplemental information

KnockoffTrio: A knockoff framework for the identification of putative causal variants in genome-wide association studies with trio design

Yi Yang, Chen Wang, Linxi Liu, Joseph Buxbaum, Zihuai He, and Iuliana Ionita-Laza

Supplemental Note 1: Exchangeability and FDR control in trio studies

We now formally define the exchangeability and FDR control in trio studies.

1 Notations

Let $\mathbf{P}_i$, a $4 \times p$ matrix, denote the parental haplotypes for the i-th trio:

$$\mathbf{P}_i = \begin{bmatrix} H_i^{f,1} \\ H_i^{f,2} \\ H_i^{m,1} \\ H_i^{m,2} \end{bmatrix}.$$

We assume that the genome has been divided into K contiguous regions of equal sizes, e.g., 200kb, and that no recombination occurs within each region. The indices of genetic variants in the k-th region are denoted as $\mathcal{T}_k$, and $t_k = |\mathcal{T}_k|$. We have $\cup_{k=1}^{K} \mathcal{T}_k = [p]$. Let $\mathbf{P}_{i,k}$, a $4 \times t_k$ matrix, denote the parental haplotypes in the k-th region.

Given the parental haplotypes, the offspring haplotypes within a region are a function of the parental haplotypes, either viewed as deterministic (observed) or random. As we have assumed no recombination events, the form of the function remains the same for all genetic variables in $\mathcal{T}_k$. A further observation is that this function is a linear form of $\mathbf{P}_{i,k}$. Therefore, it can be represented by $\mathbf{\Gamma}_{i,k}$, a 2×4 matrix. Now let $\mathbf{\Gamma}_i$ be a $2K \times 4K$ block diagonal matrix of $\mathbf{\Gamma}_{i,k}$'s

$$\mathbf{\Gamma}_i = \begin{bmatrix} \mathbf{\Gamma}_{i,1} & & \\ & \ddots & \\ & & \mathbf{\Gamma}_{i,K} \end{bmatrix},$$

and $\mathbf{P}_i^{\text{block}}$ be a $4K \times p$ block diagonal matrix of $\mathbf{P}_{i,k}$'s

$$\mathbf{P}_i^{\text{block}} = \begin{bmatrix} \mathbf{P}_{i,1} & & \\ & \ddots & \\ & & \mathbf{P}_{i,K} \end{bmatrix}.$$

Then, the offspring haplotypes can be represented as

$$\begin{bmatrix} X_i^f \\ X_i^m \end{bmatrix} = \mathbf{A}\mathbf{\Gamma}_i\mathbf{P}_i^{\text{block}},$$

where A is a $2 \times 2K$ matrix of 0's and 1's, with 1's at odd positions in the first row and at even positions in the second row.

2 Exchangeability for trios

We first construct the knockoff variables for parental haplotypes using the SCIT algorithms as described in the Methods section. Let $\tilde{\mathbf{P}}_i$, a $4 \times p$ matrix, denote the knockoff parental haplotypes for the i-th trio

$$\tilde{\mathbf{P}}_i = \begin{bmatrix} \tilde{H}_i^{f,1} \\ \tilde{H}_i^{f,2} \\ \tilde{H}_i^{m,1} \\ \tilde{H}_i^{m,2} \end{bmatrix}.$$

We also define $\tilde{\mathbf{P}}_i^{\text{block}}$ similarly. The knockoff offspring haplotypes are obtained by setting

$$\begin{bmatrix} \tilde{X}_i^f \\ \tilde{X}_i^m \end{bmatrix} = \mathbf{A}\mathbf{\Gamma}_i\tilde{\mathbf{P}}_i^{\text{block}},$$

where we assume the transmission patterns are the same for both the original and synthetic trios.

The exchangeability is defined at the matrix level. We say that $[\mathbf{P}_i, \tilde{\mathbf{P}}_i]$ satisfies the exchangeability condition if for any $\mathcal{S} \subset [p]$

$$[\mathbf{P}_i, \tilde{\mathbf{P}}_i] \overset{D}{=} [\mathbf{P}_i, \tilde{\mathbf{P}}_i]_{\text{swap}(\mathcal{S})}$$

where $[\mathbf{P}_i, \tilde{\mathbf{P}}_i]_{\text{swap}(\mathcal{S})}$ is obtained by swapping the j-th column of $\mathbf{P}_i$ and $\tilde{\mathbf{P}}_i$ for $j \in \mathcal{S}$. Note that if the exchangeability condition holds for $[\mathbf{P}_i, \tilde{\mathbf{P}}_i]$, then it also holds for $[\mathbf{P}_i^{\text{block}}, \tilde{\mathbf{P}}_i^{\text{block}}]$. Therefore,

$$\begin{bmatrix} X_i^f & \tilde{X}_i^f \\ X_i^m & \tilde{X}_i^m \end{bmatrix}_{\text{swap}(\mathcal{S})} = \mathbf{A}\mathbf{\Gamma}_i[\mathbf{P}_i^{\text{block}}, \tilde{\mathbf{P}}_i^{\text{block}}]_{\text{swap}(\mathcal{S})}.$$

This implies that, if we consider all haplotypes in a trio, the exchangeability holds in the sense that

$$\begin{bmatrix} \mathbf{P}_i & \tilde{\mathbf{P}}_i \\ X_i^f & \tilde{X}_i^f \\ X_i^m & \tilde{X}_i^m \end{bmatrix} \overset{D}{=} \begin{bmatrix} \mathbf{P}_i & \tilde{\mathbf{P}}_i \\ X_i^f & \tilde{X}_i^f \\ X_i^m & \tilde{X}_i^m \end{bmatrix}_{\text{swap}(\mathcal{S})}. \tag{1}$$

Here, we treat the transmission pattern $\mathbf{\Gamma}_i$ as given. In the case that we also consider the randomness of $\mathbf{\Gamma}_i$, the exchangeability can be understood as (1) holds conditional on $\mathbf{\Gamma}_i$.

As shown in Figure S1, the exchangeability property holds in simulations.

3 Exchangeability for trios with missing parents

If one of the parents is missing, a valid construction of knockoff variables can still be obtained by conceptually setting the knockoff variables for the missing parent to the original variables. Specifically, we can assume $H_i^{m,1}$ and $H_i^{m,2}$ are missing in $\mathbf{P}_i$. For the knockoff counterparts, we can then hypothetically set $\tilde{H}_i^{m,1} = H_i^{m,1}$, $\tilde{H}_i^{m,2} = H_i^{m,2}$, and consequently $\tilde{X}_i^m = X_i^m$, which is a trivial construction of knockoff variables. Therefore, the exchangeability condition (1) is still satisfied.

4 FDR control for trios

Our goal is to test the conditional null hypothesis

$$H_{0,g} : \boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{G}_g | \boldsymbol{G}_{-g}$$

where $\boldsymbol{G} \in \{0,1,2\}^{3n \times p}$ is the matrix of trio genotypes and $g \subset [p]$ is a continuous block. For trios, the null hypothesis is essentially

$$H_{0,l} : \boldsymbol{Y} \perp\!\!\!\perp (\mathbf{X}_{gl}, \mathbf{P}_{gl}) | (\mathbf{X}_{-gl}, \mathbf{P}_{-gl}), l = 1, \ldots, L$$

where $\mathbf{X} \in \{0,1\}^{2n \times p}$ are the offspring haplotypes, $\mathbf{P} \in \{0,1\}^{4n \times p}$ are the parental haplotypes, and $g_1, \ldots, g_L \subset [p]$ are a collection of continuous blocks of p genetic variables. Let KnockoffTrio's feature importance statistic for a window be

$$W_l = w_l\left(\begin{bmatrix} \mathbf{P} & \tilde{\mathbf{P}} \\ \mathbf{X} & \tilde{\mathbf{X}} \end{bmatrix}, \boldsymbol{y} \right)$$

for some anti-symmetric function w_l. Because the p-values for calculating W_l's are obtained from marginal tests for each genetic variable in a window, we can see that for any $S \subset 1, \ldots, L$

$$w_l\left(\begin{bmatrix}\mathbf{P} & \tilde{\mathbf{P}} \\ \mathbf{X} & \tilde{\mathbf{X}}\end{bmatrix}_{\text{swap}(S)}, \boldsymbol{y}\right) = \begin{cases} w_l\left(\begin{bmatrix}\mathbf{P} & \tilde{\mathbf{P}} \\ \mathbf{X} & \tilde{\mathbf{X}}\end{bmatrix}, \boldsymbol{y}\right), & l \notin S, \\ -w_l\left(\begin{bmatrix}\mathbf{P} & \tilde{\mathbf{P}} \\ \mathbf{X} & \tilde{\mathbf{X}}\end{bmatrix}, \boldsymbol{y}\right), & l \in S \end{cases} \tag{2}$$

where $\begin{bmatrix}\mathbf{P} & \tilde{\mathbf{P}} \\ \mathbf{X} & \tilde{\mathbf{X}}\end{bmatrix}_{\text{swap}(S)}$ is defined by swapping original genetic variables in all windows $g_l, l \in S$ with their knockoffs. The flip-sign property (2) in combination with the exchangeability (1) leads to valid FDR control for trios. The proof is for single knockoff construction and can be easily generalized to the case of multiple knockoffs with similar arguments.

5 Protection against external confounders

Based on the exchangeability condition (1), when a hypothesis $H_{0,g}$ is rejected, at least one of the following two cases is true: (1) $X_g \not\perp\!\!\!\perp Y \mid (X_{-g}, \mathbf{P}_{-g})$; (2) $\mathbf{P}_g \not\perp\!\!\!\perp Y \mid (X_{-g}, \mathbf{P}_{-g})$. Under the definition of external confounders introduced in [1], the claim from case (1) is robust to external confounders, such as population stratification, while case (2) is not. While it is possible that the observed association between P_g and Y is due to population stratification, because the family-based association test (FBAT) is applied to calculate feature importance statistics in KnockoffTrio, the p-value for the original cohort can only be small when case (1) holds, and is expected to be relatively large for case (2). This feature leads to protection against external confounders: when a discovery is made by KnockoffTrio, the family-based association test helps distinguish case (1) from case (2), and the discovery is most likely due to the association of case (1).

As shown in Figure S2, KnockoffTrio controls the FDR in the presence of population stratification at a target FDR of 0.1 for both dichotomous and quantitative traits.

Supplemental Note 2: Empirical power and FDR in meta-analysis simulations

We adopted the same simulation settings as in the "Empirical power and FDR in single-locus simulations" Section except that in each replicate we partitioned the trios into two subcohorts of 5,000 trios each. We then applied KnockoffTrio to the two subcohorts respectively and used KnockoffTrio's meta-analysis procedure to combine the results from the two subcohorts. In Figure S3, we show the empirical power and FDR for KnockoffTrio's meta-analysis of the two subcohorts, compared to the corresponding mega-analysis of the combined cohort. KnockoffTrio's meta-analysis has comparable power to the mega-analysis (when $M = 10$) while preserving the FDR in all scenarios.

Supplemental Note 3: KnockoffScreen in trio studies

KnockoffScreen was designed for independent individuals in population-based studies. We investigated the power and FDR of KnockoffScreen in trio studies. Specifically, we used KnockoffScreen to generate knockoffs of trio data disregarding the family structure and treating family members as unrelated individuals. We adopted the same simulation settings as in the "Empirical power and FDR in single-locus simulations" Section. As shown in Figure S4, KnockoffScreen has inflated FDR when applied to trio data.

Supplemental Note 4: Analyses of AGP, SPARK, and SSC cohorts with digital twin test

We applied the digital twin test to the AGP, SPARK, and SSC cohorts. We applied the digital twin test only to those loci identified by KnockoffTrio because of its demanding computational cost and because many of those loci have independent literature support so we believe they are *bona fide* ASD loci. This focused analysis is also consistent with the analysis in the original digital twin manuscript [1]. The digital twin test was performed with default parameters, 10,000 permutations, and testing windows ranging from 1kb to 2Mb (1kb, 2kb, 5kb, 10kb, 20kb, 50kb, 100kb, 200kb, 500kb, 1Mb, and 2Mb) centered around the loci identified by KnockoffTrio. We present the results in Table S1. For the AGP cohort, the digital twin test identified *ARHGEF10* ($p = 0.045$, window size = 2Mb) at the $\alpha = 0.05$ level and found suggestive significance for *LMNTD1-RASSF8* ($p = 0.053$, window size = 5kb). For the SPARK cohort, the digital twin test identified *SLC22A23/PSMG4* ($p = 0.012$, window size = 1kb), *CHSY3-HINT1* ($p = 0.024$, window size = 1Mb), *BAG4* ($p = 0.040$, window size = 2Mb), and *CCNB1IP1-PARP2* ($p = 0.048$, window size = 500kb) at the $\alpha = 0.05$ level. For the SSC cohort, the digital twin test identified no loci at the $\alpha = 0.05$ level.

In Table S1, we show the results from the digital twin test in addition to KnockoffTrio. For the digital twin test, we show the most significant p-value and corresponding testing window size for a locus.

Supplemental Note 5: Analyses of AGP and SPARK cohorts with KnockoffGWAS

We applied KnockoffGWAS to the AGP and SPARK cohorts. We did not apply KnockoffGWAS to the SSC cohort because we estimated the runtime to be more than a month in addition to unpredictable queue waiting time on the clusters. We defined the ASD children as cases and non-ASD parents as controls (N case/control: AGP = 1266/2522 and SPARK = 10540/17989). Therefore, the cases in the case-control KnockoffGWAS are identical to the cases tested by the KnockoffTrio. We performed the analyses using the default parameters in KnockoffGWAS (number of references for each haplotype mosaic = 10, size of haplotype clusters between 1k and 10k, and 7 levels of knockoff filter resolution). In line with KnockoffGWAS, we also used RaPID [2] to generate identical-by-descent (IBD) segments longer than 3 cM for each cohort. For the AGP cohort, KnockoffGWAS identified no loci significant at FDR=0.5 at any level of resolution. For the SPARK cohort, KnockoffGWAS identified seven loci at FDR=0.3 at a resolution of 41 kb. However, these loci do not overlap with the loci identified by KnockoffTrio. Among these loci, we have found that several have some level of support on their potential association with ASD. Specifically, a *de novo* deleterious variant in *PLA2G4A* was identified in one female individual from the Faroe Islands with autism without intellectual disability [3]. Similarly, *de novo* mutations in *SLC12A2* were identified in six children with neurodevelopmental disorders [4]. No loci were significant at FDR=0.1 or 0.2 (Table S2).

Supplemental Note 6: KnockoffTrio is robust in the presence of phasing errors

We evaluate KnockoffTrio's performance in the presence of phasing errors. We first adopted the same simulation settings as in the "Empirical power and FDR in single-locus simulations" Section to generate haplotype and genotype data and then in each replicate we produced switch errors in the haplotype data. A switch error is defined as the switch of haplotypes at a heterozygous site. The switch error rate (SER) is estimated to be <0.1% for SHAPEIT when parental genotype data are available[5]. Therefore, we randomly selected 0.1% of all heterozygous sites in a replicate to produce switch errors in the haplotype data. We then analyzed the haplotype data with switch errors and the genotype data using KnockoffTrio. As shown in Figure S6, KnockoffTrio is robust to phasing errors particularly when multiple knockoffs were generated.

Supplemental Note 7: Power and FDR for KnockoffTrio in the presence of lowered linkage disequilibrium

The inflated FDR for M=1 in Figure 2 is partly caused by the strong correlation between causal and non-causal variants, and is especially apparent for dichotomous traits. As a comparison, we used the same simulation settings as in the Empirical power and FDR in single-locus simulations except that we applied hierarchical clustering such that variants from different clusters have correlation no greater than 0.6 instead of 0.7. As shown in Figure S7, a single knockoff no longer has inflated FDR when the correlation between variants is lowered.

Supplemental Figures

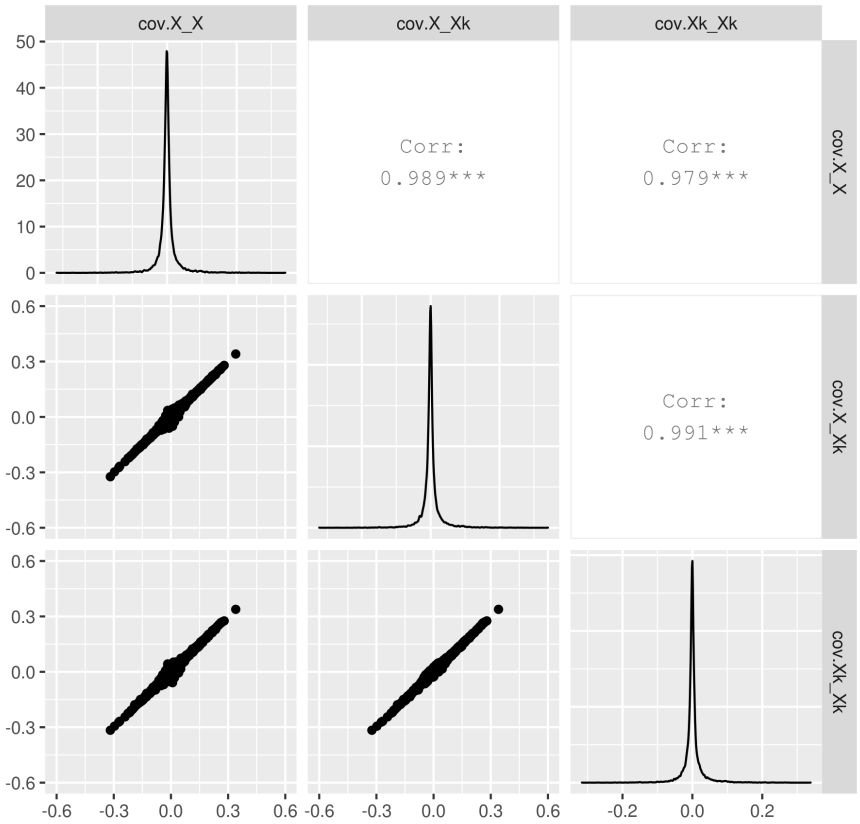

Figure S1: Empirical validation for exchangeability property in KnockoffTrio. To validate the exchangeability, we generated offspring knockoff genotypes (Xk) using the proposed algorithm and evaluated whether the second order (covariance between each pair of genetic variants) is exchangeable for common variants in the region. "Cov.X_X" is the covariance between each pair of original variants, "cov.Xk_Xk" is the covariance between each pair of knockoff variants, and "cov.X_Xk" is the covariance between each pair of original and knockoff variants.

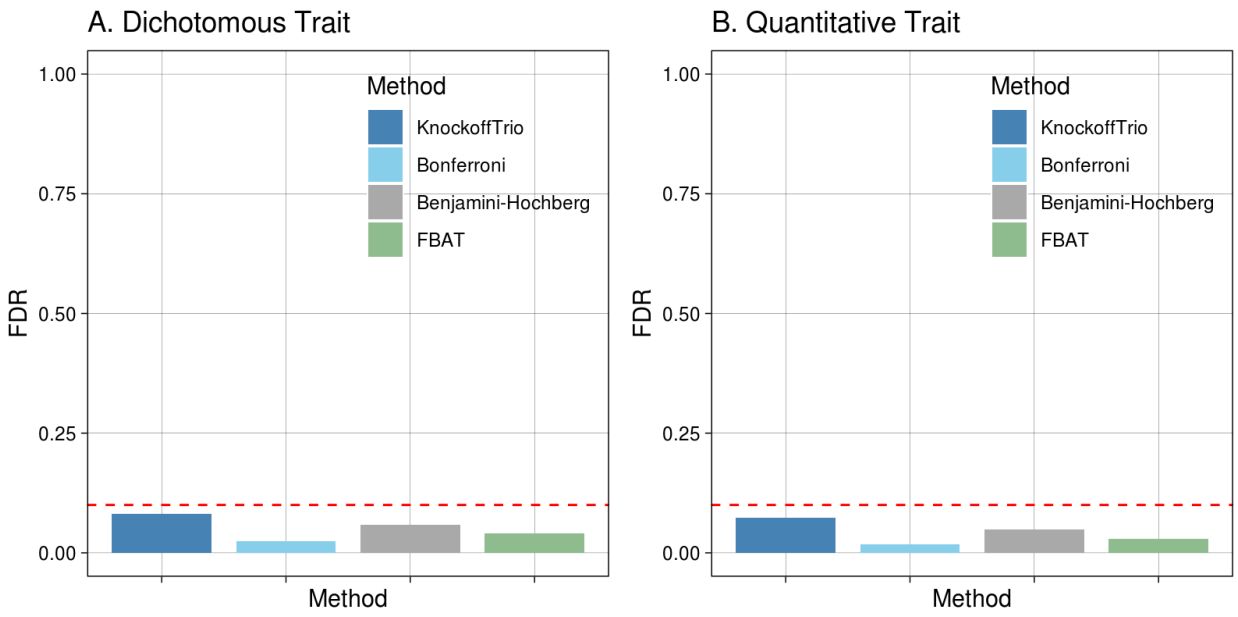

Figure S2: KnockoffTrio controls FDR in the presence of population stratification. The two panels show the FDR in the presence of population stratification for dichotomous and quantitative traits. A method's FDR is defined as the proportion of replicates where any window is detected among 500 replicates. The target FDR is 0.1.

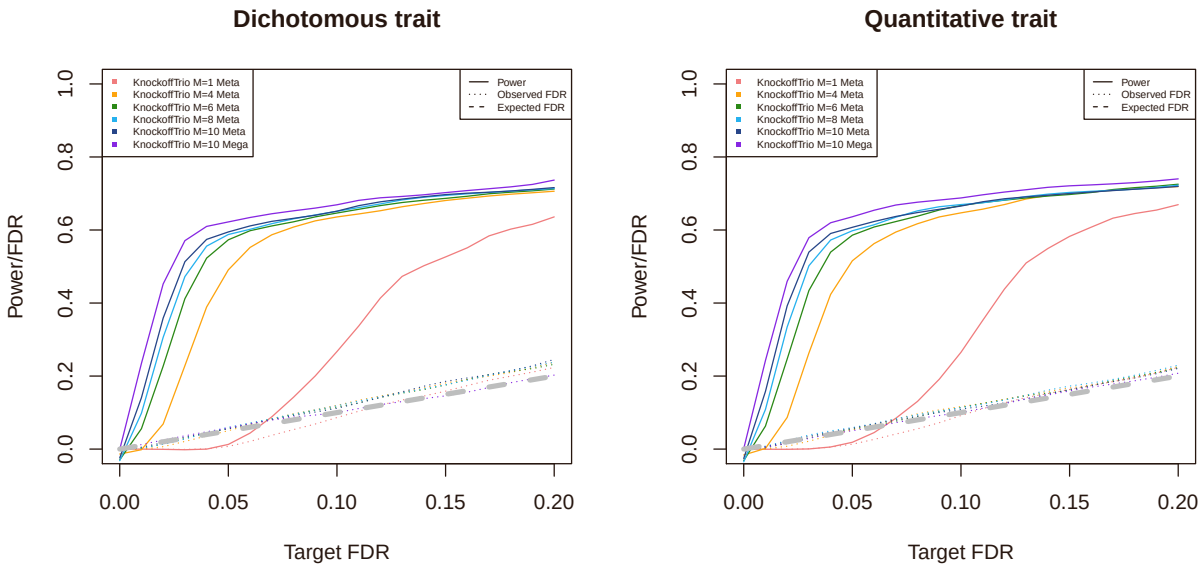

Figure S3: KnockoffTrio's power and FDR in meta-analysis. The two panels show the power and FDR for dichotomous and quantitative traits. We evaluate KnockoffTrio's power and FDR with a target FDR ranging from 0 to 0.2 and with different numbers of knockoffs. The solid lines indicate KnockoffTrio's power and the dotted lines indicate KnockoffTrio's observed FDR. The different colors indicate different numbers of knockoffs. The grey dashed line indicates the expected FDR.

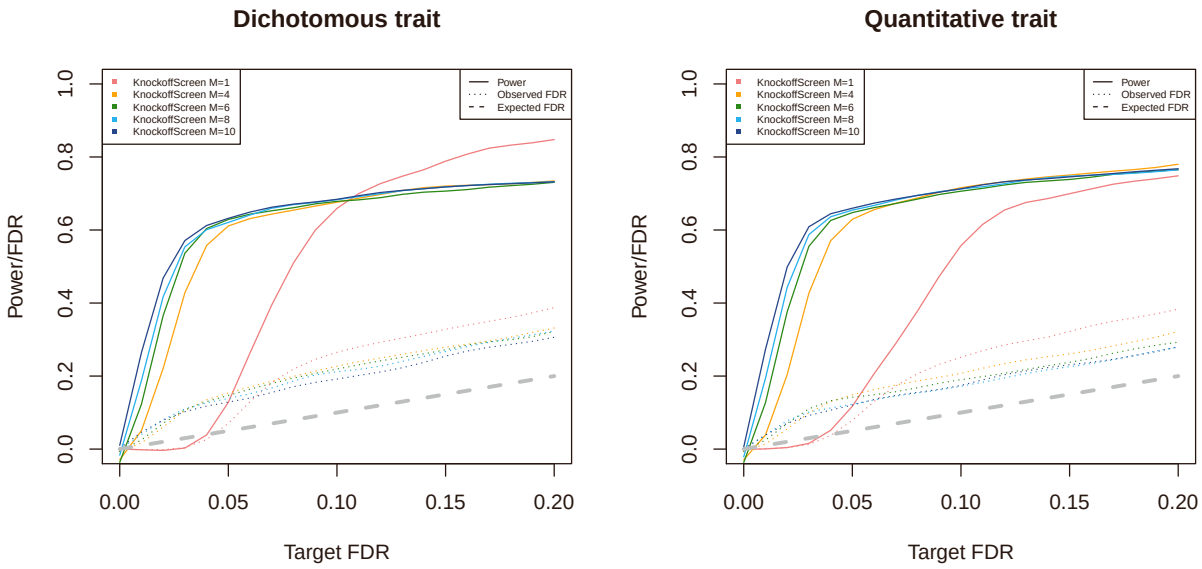

Figure S4: KnockoffScreen's power and FDR in single-locus simulations. The two panels show the power and FDR for dichotomous and quantitative traits. We evaluate KnockoffScreen's power and FDR with a target FDR ranging from 0 to 0.2 and with different numbers of knockoffs. The solid lines indicate KnockoffTrio's power and the dotted lines indicate KnockoffScreen's observed FDR. The different colors indicate different numbers of knockoffs. The grey dashed line indicates the expected FDR.

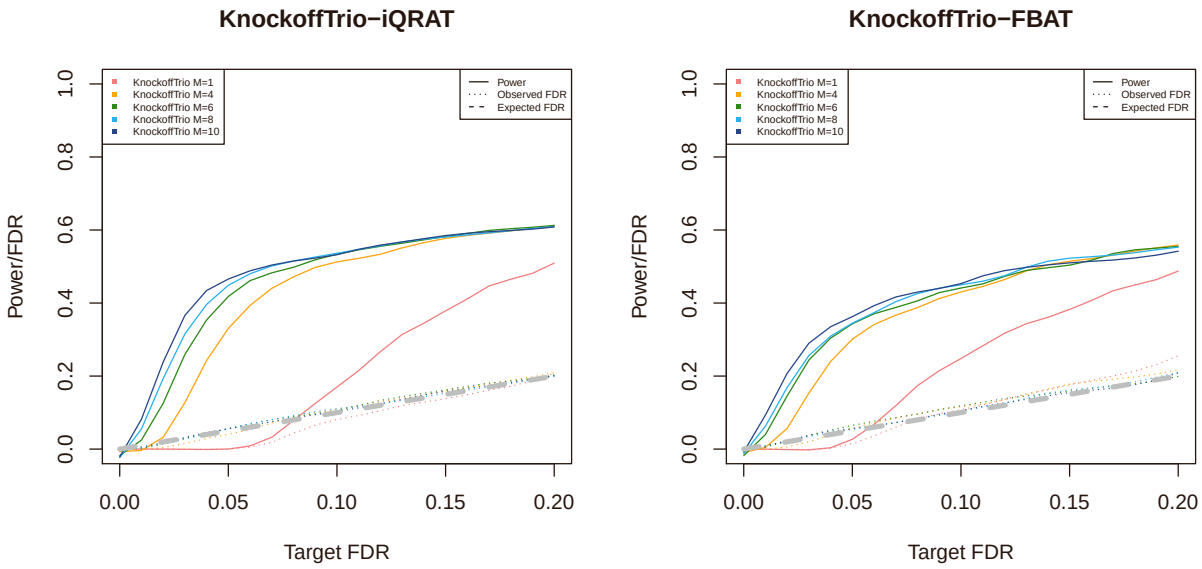

Figure S5: KnockoffTrio-iQRAT improves power in detecting complex associations. The two panels show the power and FDR for quantitative traits with Cauchy error terms using KnockoffTrio-iQRAT and KnockoffTrio-FBAT, respectively. We evaluate KnockoffTrio's power and FDR with a target FDR ranging from 0 to 0.2 and with different numbers of knockoffs. The solid lines indicate KnockoffTrio's power and the dotted lines indicate KnockoffTrio's observed FDR. The different colors indicate different numbers of knockoffs. The grey dashed line indicates the expected FDR.

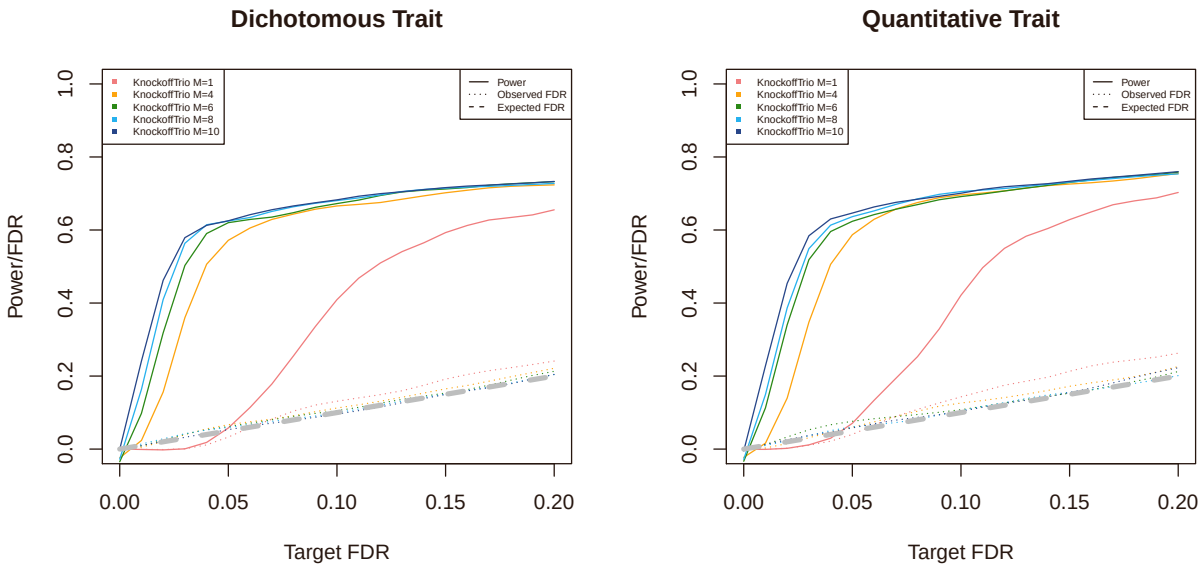

Figure S6: KnockoffTrio's power and FDR in the presence of phasing errors. The two panels show the power and FDR for dichotomous and quantitative traits. We evaluate KnockoffTrio's power and FDR with a target FDR ranging from 0 to 0.2 and with different numbers of knockoffs. The solid lines indicate KnockoffTrio's power and the dotted lines indicate KnockoffTrio's observed FDR. The different colors indicate different numbers of knockoffs. The grey dashed line indicates the expected FDR.

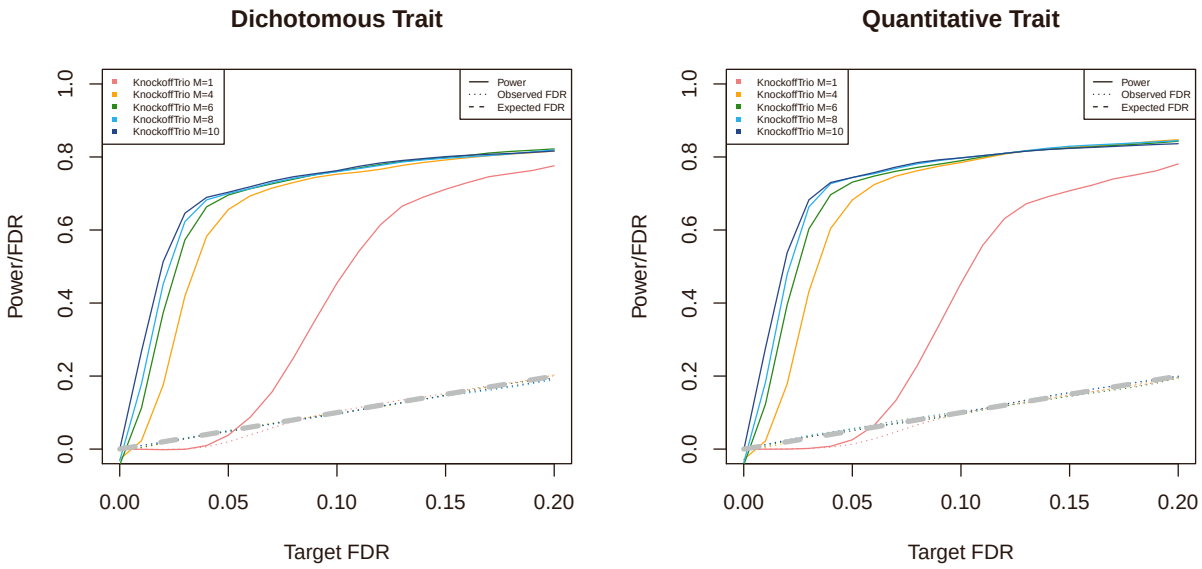

Figure S7: KnockoffTrio's power and FDR with reduced inter-cluster LD. The two panels show the power and FDR for dichotomous and quantitative traits. We evaluate KnockoffTrio's power and FDR with a target FDR ranging from 0 to 0.2 and with different numbers of knockoffs. The solid lines indicate KnockoffTrio's power and the dotted lines indicate KnockoffTrio's observed FDR. The different colors indicate different numbers of knockoffs. The grey dashed line indicates the expected FDR.

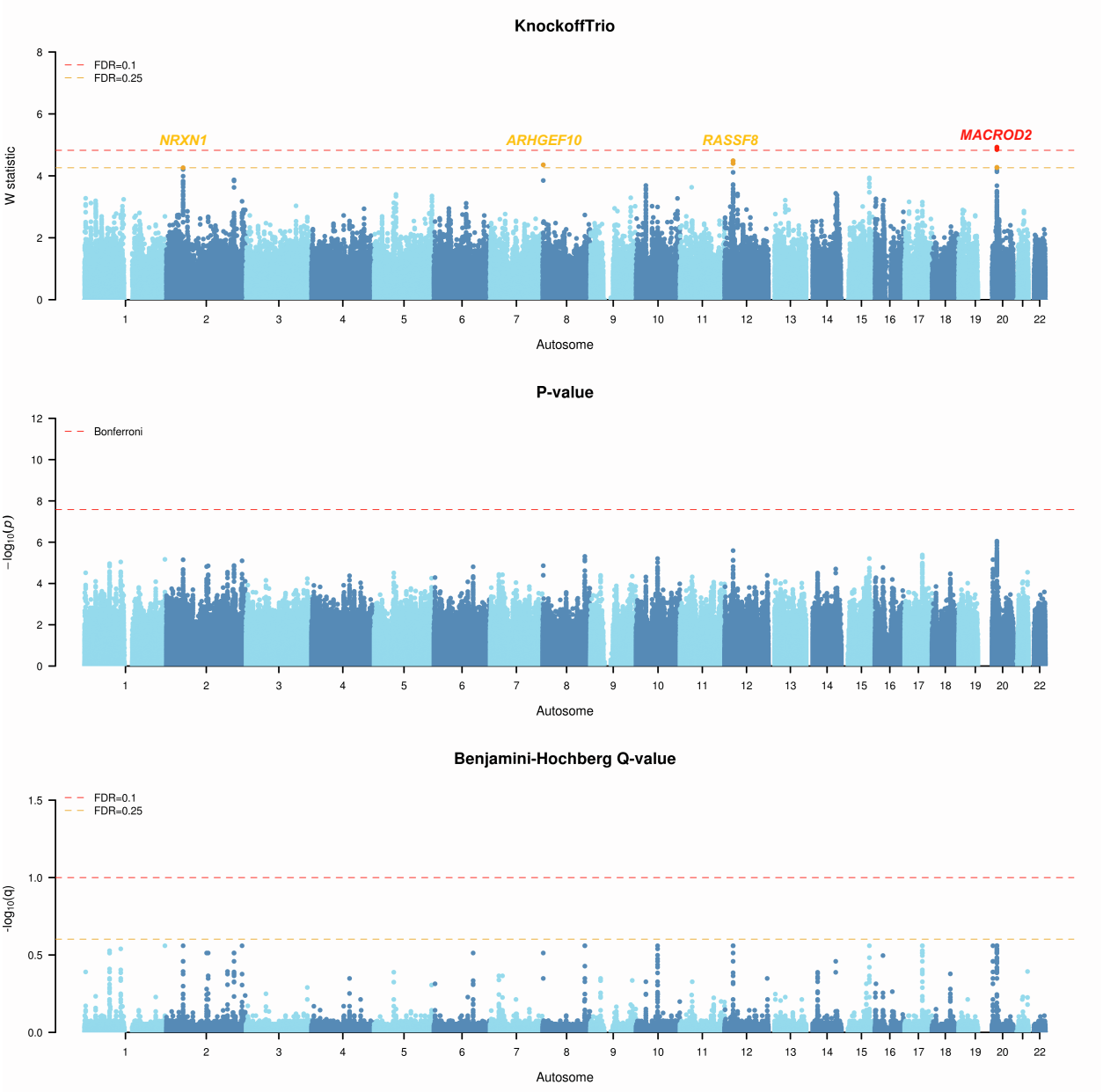

Figure S8: Replication of Autism Genome Project (AGP) KnockoffTrio results. Manhattan plots from KnockoffTrio analysis for the Autism Genome Project (AGP) with different random seeds in knockoff generation. The Manhattan plots of the W statistics from KnockoffTrio, the p-values from the conventional association tests with the Bonferroni correction for controlling the FWER, and the Q-values from the BH procedure for controlling the FDR. Different random seeds were used to generate knockoffs than in the main manuscript. The FDR target level for KnockoffTrio and the BH procedure is 0.1 or 0.25. Each locus is annotated with the closest gene name.

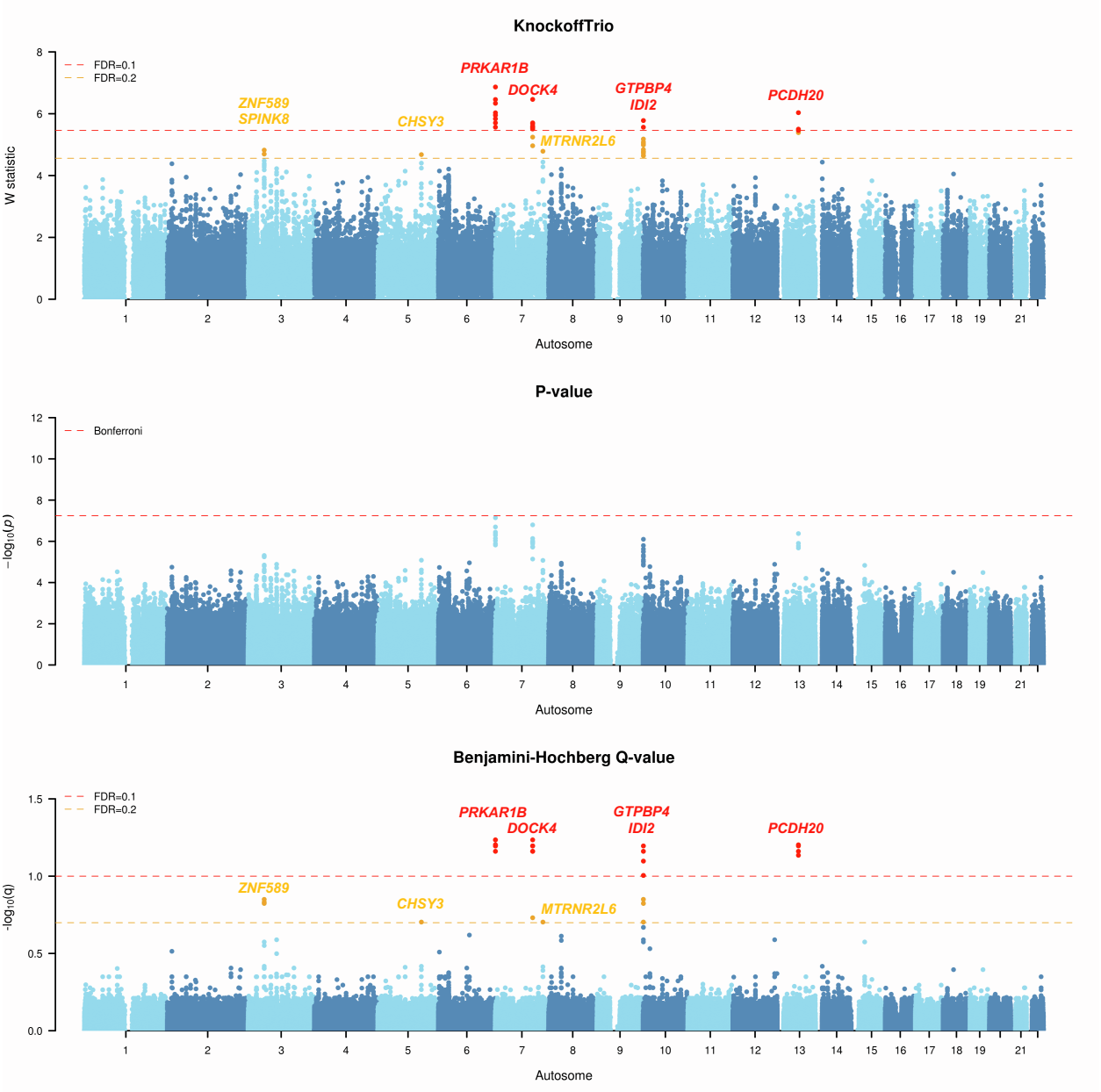

Figure S9: Replication of the Simons Foundation Powering Autism Research (SPARK) KnockoffTrio results. Manhattan plots from KnockoffTrio analysis for the Simons Foundation Powering Autism Research (SPARK) with different random seeds in knockoff generation. The Manhattan plots of the W statistics from KnockoffTrio, the p-values from the conventional association tests with the Bonferroni correction for controlling the FWER, and the Q-values from the BH procedure for controlling the FDR. Different random seeds were used to generate knockoffs than in the main manuscript. The FDR target level for KnockoffTrio and the BH procedure is 0.1 or 0.2. Each locus is annotated with the closest gene name.

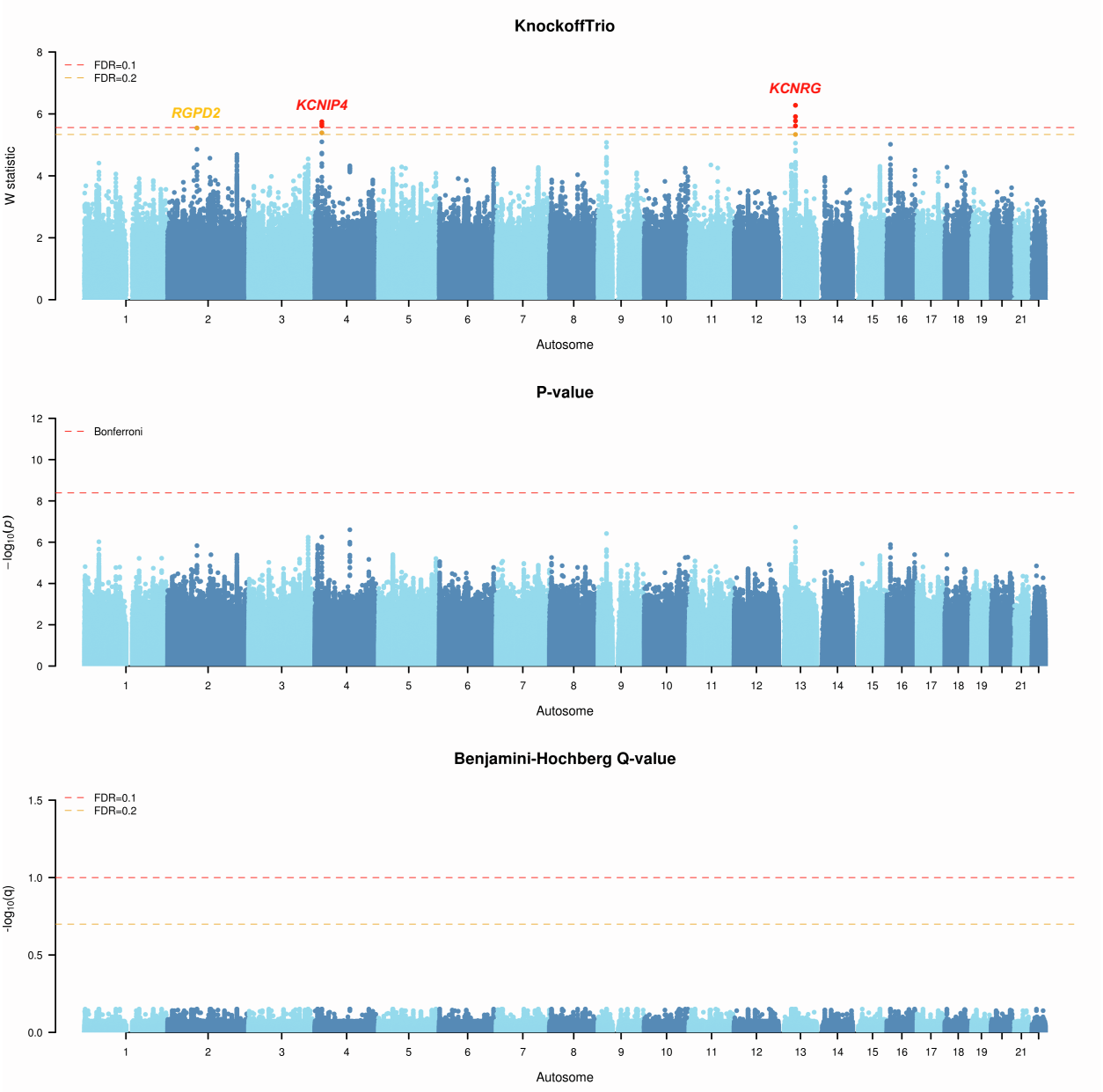

Figure S10: Replication of the Simons Simplex Collection (SSC) KnockoffTrio results. Manhattan plots from KnockoffTrio analysis for the Simons Simplex Collection (SSC) with different random seeds in knockoff generation. The Manhattan plots of the W statistics from KnockoffTrio, the p-values from the conventional association tests with the Bonferroni correction for controlling the FWER, and the Q-values from the BH procedure for controlling the FDR. Different random seeds were used to generate knockoffs than in the main manuscript. The FDR target level for KnockoffTrio and the BH procedure is 0.1 or 0.2. Each locus is annotated with the closest gene name.

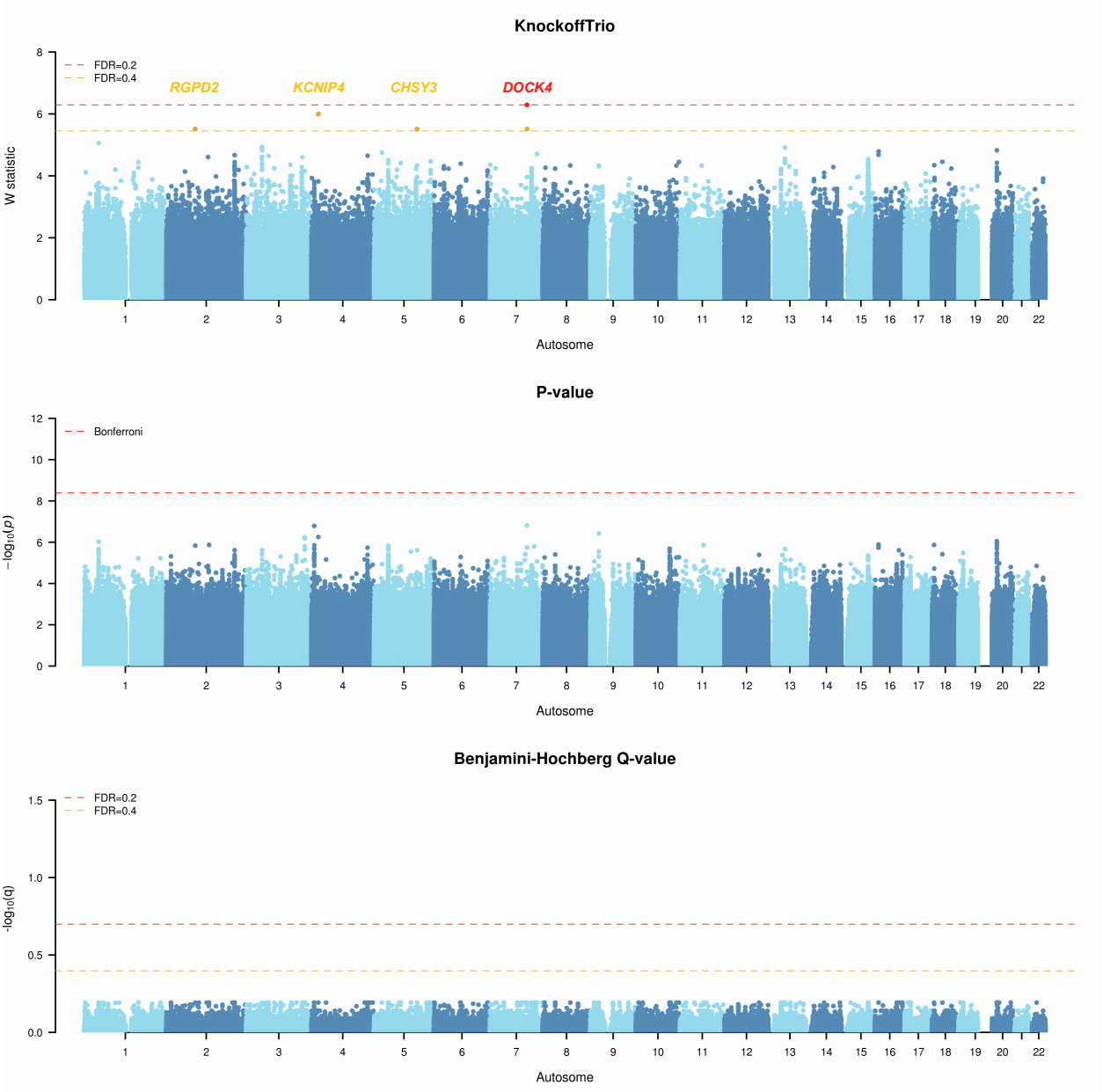

Figure S11: Manhattan plots from KnockoffTrio analysis for the meta-analysis of the AGP, SPARK, and SSC cohorts. The Manhattan plots of the W statistics from KnockoffTrio, the p-values from the conventional association tests with the Bonferroni correction for controlling the FWER, and the Q-values from the BH procedure for controlling the FDR. The FDR for KnockoffTrio and the BH procedure is 0.2 and 0.4. Each locus is annotated with the closest gene name.

Supplemental Tables

Gene	Chr	Position	Variant	Allele	MAF	P	Z	W	Q	BH Q	DTT Size	DTT P
AGP (FDR=0.1)												
NRXN1	2	50805721	rs9284756	A	0.03	7.10E-6	4.49	4.37	0.10	0.28	200K	8.56E-2
ARHGEF10	8	1920247-1920676	rs17756915-rs11136442	-	0.41	1.38E-5	-	4.47	0.10	0.31	2M	4.50E-2
LMNTD1-RASSF8	12	25946268	rs4963941	A	0.10	2.56E-6	4.70	4.84	0.10	0.28	5K	5.30E-2
ALPK3-SLC28A1	15	84881866	rs12917429	T	0.21	6.19E-6	-4.52	4.45	0.10	0.28	100K	2.93E-1
MACROD2	20	14781064	rs6074798	A	0.49	1.02E-6	4.89	4.83	0.10	0.28	20K	2.83E-1
SFARI: SPARK (FDR=0.1)												
ZNF589	3	48262179	rs11709691	G	0.28	4.87E-6	-4.57	5.03	0.06	0.14	2M	9.04E-1
CADM2	3	85395534-85410981	rs75005531-rs1549979	-	0.22	1.30E-5	-	4.76	0.09	0.26	1M	6.77E-1
CHSY3-HINT1	5	130661503	rs17714209	C	0.28	8.25E-6	4.46	4.99	0.06	0.20	1M	2.40E-2
PDGFA-PRKAR1B	7	536383	rs62431385	C	0.10	7.20E-8	-5.39	6.71	0.02	0.06	2M	2.49E-1
DOCK4	7	111986531	rs73210911	A	0.12	1.59E-7	-5.24	6.51	0.02	0.06	2M	4.67E-1
MTRNR2L6-PRSS1	7	142688332	rs13223009	C	0.02	8.42E-6	-4.45	4.71	0.09	0.20	1M	1.01E-1
LARP4B-GTPBP4	10	975370	rs117732138	A	0.02	1.60E-6	4.80	5.48	0.02	0.07	50K	3.18E-1
IDI2	10	1020654	rs77782977	C	0.02	7.95E-7	4.94	5.84	0.02	0.06	20K	1.83E-1
PCDH20-PCDH9	13	63204555	rs12184522	T	0.23	4.21E-7	5.06	6.00	0.02	0.06	2M	2.20E-1
SFARI: SPARK (FDR=0.2)												
SPINK8	3	48316110-48329279	rs74735576-rs13090538	-	0.17	1.58E-5	-	4.39	0.17	0.28	100K	8.28E-1
SLC22A23/PSMG4	6	3285062	rs41301847	G	0.02	1.85E-5	4.28	4.41	0.17	0.31	1K	1.18E-2
BAG4	8	38205717	rs7836805	A	0.24	2.83E-5	-4.19	4.43	0.17	0.40	2M	4.03E-2
CCNB1IP1-PARP2	14	20334133	rs72671266	T	0.02	2.45E-5	-4.22	4.30	0.19	0.38	500K	4.75E-2
SFARI: SSC (FDR=0.1)												
KCNRG-DLEU7	13	50197099	rs2703087	A	0.04	1.88E-7	5.21	6.54	0.10	0.70	1M	9.30E-2
SFARI: SSC (FDR=0.2)												
KCNIP4	4	20917151	rs185413018	T	0.02	5.59E-7	5.00	6.00	0.13	0.70	10K	1.52E-1

Table S1: Analyses of AGP, SPARK, and SSC cohorts with digital twin test (DTT). Gene: loci identified by KnockoffTrio. A single gene name indicates the signal is within or overlaps with the gene. "Gene1/Gene2" indicates the signal overlaps with two genes. "Gene1-Gene2" indicates the signal is between two genes. **MAF**: minor allele frequency of a variant, or average minor allele frequency if a signal contains multiple variants. **P**: KnockoffTrio's ACAT-combined p-values. For single variants, ACAT-combined p-values are equivalent to FBAT p-values. **Z**: FBAT Z-scores for single variants. **W**: KnockoffTrio's feature statistics. **Q**: KnockoffTrio's Q-values. **BH Q**: Benjamini-Hochberg Q-values. **DTT Size**: testing window size for the digital twin test. The unit is base pair. **DTT P**: digital twin test's p-values.

Gene	Chr	Position	Variant	W
SPARK (FDR=0.3, resolution=41 kb)				
ZYG11B	1	52781682	rs74911353	0.002
PLA2G4A-BRINP3	1	190041115	rs17374565	0.004
DYSF-CYP26B1	2	71858486	rs12469485	0.003
FZD5	2	207767140	2-207767140-T-G	0.019
PCDH7-ARAP2	4	33081025	rs78314717	0.002
CCDC192-SLC12A2	5	127948101	rs72792235	0.002
TMEM71-PHF20L1	8	132773968	rs28550258	0.002

Table S2: Analyses of AGP and SPARK cohorts with KnockoffGWAS. Gene: loci identified by KnockoffGWAS. A single gene name indicates the signal is within or overlaps with the gene. "Gene1-Gene2" indicates the signal is between two genes. **Position**: Position of the lead variant of a locus. **Variant**: The lead variant of a locus. **W**: KnockoffGWAS's feature statistics. KnockoffGWAS's feature importance scores were calculated by fitting the Lasso with cross-validation and taking the absolute value of the average estimated regression coefficients, and W, the feature statistics, were feature importance scores for the original minus those for the knockoff cohort.

Supplemental References

[1] Bates, S., Sesia, M., Sabatti, C., and Candès, E. (2020). Causal inference in genetic trio studies. Proc. Natl. Acad. Sci. U. S. A. 117, 24117–24126.

[2] Naseri, A., Liu, X., Tang, K., Zhang, S., and Zhi, D. (2019). Rapid: ultra-fast, powerful, and accurate detection of segments identical by descent (ibd) in biobank-scale cohorts. Genome Biol. 20, 143.

[3] Leblond, C. S., Cliquet, F., Carton, C., Huguet, G., Mathieu, A., Kergrohen, T., Buratti, J., Lemière, N., Cuisset, L., Bienvenu, T., et al. (2019). Both rare and common genetic variants contribute to autism in the faroe islands. NPJ Genom. Med. 4, 1.

[4] McNeill, A., Iovino, E., Mansard, L., Vache, C., Baux, D., Bedoukian, E., Cox, H., Dean, J., Goudie, D., Kumar, A., et al. (2020). Slc12a2 variants cause a neurodevelopmental disorder or cochleovestibular defect. Brain 143, 2380–2387.

[5] Choi, Y., Chan, A. P., Kirkness, E., Telenti, A., and Schork, N. J. (2018). Comparison of phasing strategies for whole human genomes. PLoS genetics 14, e1007308.