A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases

Citation

Published Version

Permanent link

Terms of Use

Share Your Story

Accessibility

A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases

Buhm Han[1,2,3,4,25], **Jennie G. Pouget**[1,5,6,7,25], **Kamil Slowikowski**[1,3,4,8], **Eli Stahl**[9], **Cue Hyunkyu Lee**[10], **Dorothee Diogo**[1,3,4], **Xinli Hu**[1,3,4,11], **Yu Rang Park**[10,12], **Eunji Kim**[10,13], **Peter K. Gregersen**[14], **Solbritt Rantapää Dahlqvist**[15], **Jane Worthington**[16,17], **Javier Martin**[18], **Steve Eyre**[16,17], **Lars Klareskog**[19], **Tom Huizinga**[20], **Wei-Min Chen**[21], **Suna Onengut-Gumuscu**[21], **Stephen S. Rich**[21], **Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium**[22], **Naomi R. Wray**[23], and **Soumya Raychaudhuri**[1,3,4,19,24]

[1]Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, USA [2]Department of Convergence Medicine, University of Ulsan College of Medicine & Asan Institute for Life Sciences, Asan Medical Center, Seoul, Republic of Korea [3]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, USA [4]Partners Center for Personalized Genetic Medicine, Boston, USA [5]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Canada [6]Institute of Medical Sciences, University of Toronto, Toronto, Canada [7]Department of Psychiatry, University of Toronto, Toronto, Canada [8]Bioinformatics and Integrative Genomics, Harvard University, Cambridge, USA [9]Department of Psychiatry, Mount Sinai School of Medicine, New York, USA [10]Asan Institute for Life Sciences, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea [11]Harvard-MIT Division of Health Sciences and Technology, Boston, USA [12]Department of Biomedical Informatics, Asan Medical Center, Seoul, Republic of Korea [13]Department of Chemistry, Seoul National University, Seoul, Republic of Korea [14]Robert S. Boas Center for Genomics and Human Genetics, The Feinstein Institute for Medical Research, Manhasset, USA [15]Department of Public Health and Clinical Medicine, Rheumatology, Umeå University, Umeå,

Correspondence to: Soumya Raychaudhuri, 77 Avenue Louis Pasteur, Harvard New Research Building, Suite 250D, Boston, MA 02446, USA. soumya@broadinstitute.org; 617-525-4484 (tel); 617-525-4488 (fax). Buhm Han, Asan Institute for Life Sciences, Asan Medical Center, 88, Olympic-ro 43-gil, Songpa-gu, Seoul 138-736, Korea. buhm.han@amc.seoul.kr; 82-2-3010-2648 (tel); 82-2-3010-2619 (fax).
[22]A full list of members and affiliations appears in the Supplementary Note
[25]These authors contributed equally to this work

URLs

BUHMBOX software, https://www.broadinstitute.org/mpg/buhmbox/; ImmunoBase, http://www.immunobase.org.

Sweden [16]Arthritis Research UK Centre for Genetics and Genomics, Musculoskeletal Research Centre, Institute for Inflammation and Repair, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK [17]National Institute for Health Research, Manchester Musculoskeletal Biomedical Research Unit, Central Manchester University Hospitals National Health Service Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, UK [18]Instituto de Parasitologíay Biomedicina López-Neyra, Consejo Superior de Investigaciones Científicas, Granada, Spain [19]Rheumatology Unit, Department of Medicine, Karolinska Institutet and Karolinska University Hospital Solna, Stockholm, Sweden [20]Department of Rheumatology, Leiden University Medical Centre, Leiden, the Netherlands [21]Center for Public Health Genomics, University of Virginia, Charlottesville, USA [23]The University of Queensland, Queensland Brain Institute, Brisbane, Australia [24]Institute of Inflammation and Repair, University of Manchester, Manchester, UK

Abstract

There is growing evidence of shared risk alleles between complex traits (pleiotropy), including autoimmune and neuropsychiatric diseases. This might be due to sharing between all individuals (whole-group pleiotropy), or a subset of individuals within a genetically heterogeneous cohort (subgroup heterogeneity). BUHMBOX is a well-powered statistic distinguishing between these two situations using genotype data. We observed a shared genetic basis between 11 autoimmune diseases and type 1 diabetes (T1D, $p<10^{-4}$), and 11 autoimmune diseases and rheumatoid arthritis (RA, $p<10^{-3}$). This sharing was not explained by subgroup heterogeneity (corrected $p_{BUHMBOX}>0.2$, 6,670 T1D cases and 7,279 RA cases). Genetic sharing between seronegative and seropostive RA ($p<10^{-9}$) had significant evidence of subgroup heterogeneity, suggesting a subgroup of seropositive-like cases within seronegative cases ($p_{BUHMBOX}=0.008$, 2,406 seronegative RA cases). We also observed a shared genetic basis between major depressive disorder (MDD) and schizophrenia ($p<10^{-4}$) that was not explained by subgroup heterogeneity ($p_{BUHMBOX}=0.28$ in 9,238 MDD cases).

INTRODUCTION

Recent studies have demonstrated that many diseases share risk alleles[1–4] and exhibit significant coheritability[5–7]. Coheritability studies are defining the relationship between complex traits, and providing new insights into disease mechanisms. Critically, as the number of phenotypes studied with genetics expands in the context of emerging deeply phenotyped population-wide cohorts[8], including the Precision Medicine Initiative[9], coheritablity between traits will become even more apparent. In the genomic era, methods for detecting coheritability have moved beyond traditional approaches such as twin or family studies[10, 11]. Now, alternative approaches using genome-wide association study (GWAS) data from unrelated individuals are widely used. Polygenic risk score approaches[3, 12, 13] build genetic risk scores (GRSs) for one phenotype and test their association with a second phenotype. Mixed-model approaches[5, 6, 14] can estimate the genetic covariance between two traits on the observed scale. Genetic covariance can be used to calculate genetic correlation and coheritability[6]. Cross-trait LD Score regression (LDSC) utilizes linkage disequilibrium

(LD) and summary statistics obtained from GWAS to estimate genetic correlation attributable to SNPs[7]. In addition, the p-values of independent SNPs associated with multiple phenotypes can be tested for a significant deviation from the null distribution[2]. These approaches have been applied to demonstrate significant shared genetic structure among many phenotypes[5, 7, 15] including autoimmune[2] and neuropsychiatric diseases[3, 6, 13]. The observed coheritability and genetic sharing suggests the possibility of *pleiotropy*, defined here as the sharing of risk alleles across traits at specific loci or at a genome-wide level. An example of pleiotropy is the *PTPN22* variant R620W, which is associated with multiple autoimmune diseases[16].

Shared risk alleles across diseases can be driven by all individuals or by a subset of individuals. In the former, the sharing is clearly driven by pleiotropy (*whole-group pleiotropy*). In the latter, only a subset of individuals is genetically similar to another disease. We call this *subgroup heterogeneity* – a situation where a patient cohort consists of genetically distinct subgroups that may or may not result in distinct symptom profiles and treatment outcomes. Subgroup heterogeneity can occur in the context of misclassifications (e.g. cases with atypical clinical presentations for a different disease are erroneously included), molecular subtypes (e.g. two different etiologies cause a disease, resulting in a subset of cases that share pathogenesis with a different disease), asymmetric causal relationships (e.g. one disease causes another disease, resulting in a subset of cases that also have the causal disease; often called *mediated pleiotropy*), or ascertainment bias (e.g. cases also affected with a different disease are more likely to come to clinical attention and be included in the study). These situations result in a subset of cases that is genetically similar to another disease, creating shared genetic structure[17]. Indeed, there is now evidence that misclassifications[18–21], etiological diversity[22], and ascertainment bias[23] are prevalent across certain human diseases, leading to the conclusion that significant heterogeneity may exist[24–27]. Since the potential contribution of subgroup heterogeneity to any genetic sharing observed between diseases represents a critical disease insight, statistical methods are needed to distinguish subgroup heterogeneity from whole-group pleiotropy. For the purposes of this paper, we will use the term *pleiotropy* to refer to whole-group pleiotropy and *heterogeneity* to refer to subgroup heterogeneity.

RESULTS

Overview of BUHMBOX

Genetic sharing between disease A (D_A) and disease B (D_B) could be due to pleiotropy, but could also be due to heterogeneity (i.e. a subset of D_A cases are genetically more similar to D_B cases). If we calculated GRSs for D_A cases using D_B-associated loci and their effect sizes (GRS_B), the mean of GRS_B would be statistically different between D_A cases and controls under either pleiotropy or heterogeneity. Under pleiotropy, some D_B risk alleles impose D_A risk, and D_B risk alleles will be enriched in D_A cases compared to controls. Under heterogeneity, a subset of D_A cases will have genetic characteristics of D_B, and therefore D_B risk alleles will also be enriched in those individuals. In both situations, the enriched D_B risk alleles in D_A cases will result in an increased mean GRS_B in individuals that are D_A cases. For the same reasons, if we calculated the r_g of D_A and D_B using cross-

trait LDSC[7] in this scenario, the r_g would be positive under both pleiotropy and heterogeneity.

To detect heterogeneity, even in the presence of pleiotropy, we developed BUHMBOX (Breaking Up Heterogeneous Mixture Based On Cross-locus correlations). Our method tests for the presence of heterogeneous subgroups (i.e. D_B-like cases) in an otherwise homogenous phenotype (i.e. D_A). To do this, BUHMBOX requires (1) a list of known D_B-associated SNPs with corresponding risk alleles, risk allele frequencies, and effect sizes, and (2) individual-level genotype data for D_B SNPs in D_A cases. BUHMBOX leverages the fact that in the setting of heterogeneity, D_B risk alleles have higher allele frequencies only in a specific subset of D_A cases. In contrast, under true pleiotropy, D_B risk alleles are expected to have higher allele frequencies across all D_A cases (Figure 1). If D_B risk alleles are enriched in one subgroup, the expected correlations of risk allele dosages between loci will be consistently positive (for details see Supplementary Table 1 and Supplementary Note). BUHMBOX combines these pairwise correlations into one statistic and tests for it; heterogeneity can lead to a significant BUHMBOX test statistic. In contrast, the lack of true heterogeneity or insufficient power to detect the presence of heterogeneity (type II error) can lead to a non-significant BUHMBOX test statistic. Insufficient power occurs when the number of D_A cases, heterogeneity proportion, or number of known risk alleles and/or their effect sizes for D_B are small.

BUHMBOX discriminates between heterogeneity and pleiotropy

To demonstrate that BUHMBOX detects heterogeneity (even in the presence of pleiotropy), we conducted simulations with the following parameters: sample size of D_A case individuals (N), number of risk loci associated to D_B (M), and the proportion of D_A cases that actually show genetic characteristics of D_B (heterogeneity proportion, or π). To simulate realistic distributions of effect sizes and allele frequencies, we sampled odds ratio (OR) and risk allele frequency (RAF) pairs from reported associations in the GWAS catalog[28] (**Online Methods**).

To characterize the false positive rate (FPR) of BUHMBOX we simulated 1,000,000 studies (N=2,000 and M=50) where there was no heterogeneity (π=0, **Online Methods**) or pleiotropy. BUHMBOX obtained a 5.1% FPR at p<0.05; it also obtained appropriate FPRs at a wide range of statistical significance thresholds (p<0.05 to 0.0005, Supplementary Table 2).

To evaluate the FPR of BUHMBOX where there actually was pleiotropy without heterogeneity (π=0), we simulated 1,000 studies (N=2,000 and M=50) assuming D_A and D_B shared 10% of risk loci (five loci). We quantified the proportion of instances where BUHMBOX and GRS approaches obtained p-values smaller than the threshold p<0.05. GRS appropriately demonstrated 64.8% power to detect shared genetic structure. BUHMBOX demonstrated an appropriate false positive rate of 4.3% to detect heterogeneity (Supplementary Figure 1).

Finally, to evaluate BUHMBOX's power to detect heterogeneity we repeated these simulations assuming there was no pleiotropy, but that there was indeed subtle

heterogeneity. We assumed that 10% of D_A cases were actually D_B (π=0.1). Here, BUHMBOX demonstrated 81.7% power to detect heterogeneity at p<0.05 (Supplementary Figure 1). The GRS approach demonstrated 100% power to detect shared genetic structure. Note that the power difference of the GRS approach in the pleiotropy and heterogeneity simulations is because of the stochastic chance that sampled effect sizes of all five loci may be small in the pleiotropy simulation; in simulations where we fixed the OR (1.25) and RAF (0.3) for all loci, the power of GRS was similar: 91.8% in pleiotropy and 92.0% in heterogeneity.

Together, these simulations illustrate that BUHMBOX is sensitive to heterogeneity but robust to pleiotropy, while the GRS detects both scenarios and cannot discriminate between the two. Thus, BUHMBOX complements methods for detecting pleiotropy by helping to interpret shared genetic structure (Supplementary Table 1).

Weighting pairwise correlations increases power

BUHMBOX combines multiple pairwise correlations into one statistic. A pair of loci with larger allele frequencies and effect sizes will show larger expected correlation given the same π, and may be more informative than other pairs of loci (Supplementary Figure 2). We hypothesized that accounting for this unequal information between SNP pairs could increase power. We defined a scheme to weight pairwise correlations between loci as a function of their effect sizes and allele frequencies (**Online Methods**). In simulations we observed substantial power gain with this weighting scheme. Assuming 1,000 cases and 50 loci, we compared the BUHMBOX power implemented with and without weighting correlations (equation (12) in Supplementary Note). Across a wide range of π we observed that weighting dramatically increased power (Figure 2). For example, at π=0.1 the weighted implementation of BUHMBOX obtained 74% compared to the unweighted implementation which obtained only 36% power.

Power is proportional to number of samples and loci

The statistical power of BUHMBOX is a function of many factors including sample size N of the cases we are testing for heterogeneity in, heterogeneity proportion π, number of loci M for the coheritable disease, RAF, and OR. We sampled pairs of RAF and OR from the GWAS catalog. Given a sample size of N=2,000 cases and 2,000 controls, assuming π=0.2 and 50 risk loci, BUHMBOX achieved 92% power at p<0.05 (Figure 3). As many GWAS now consist of more than 2,000 cases, and many diseases are approaching 50 known associated loci[28], BUHMBOX is currently well powered to detect a moderate amount heterogeneity (π=0.2) for many human traits. Modest heterogeneity is more challenging to detect at this sample size; power decreased to 67% at π=0.1 and to 38% at π=0.05. Power can be augmented with larger sample size (Figure 3) and larger effect sizes (Supplementary Figure 3). Power can also be increased by including large numbers of loci with even nominal evidence of association in addition to established genome-wide significant loci (Supplementary Note and Supplementary Figure 4).

Controlling for linkage disequilibrium

Although BUHMBOX adequately controlled the FPR when loci were truly independent, we were concerned that long-range LD between apparently independent loci may introduce false positives[29]. To ensure BUHMBOX was robust to LD, we implemented the following strategies: (1) stringent LD-pruning of D_B loci to exclude SNPs with $r^2 > 0.1$, and (2) accounting for any remaining residual LD by assessing the relative increase of correlations in cases compared to controls (*delta-correlations*). We evaluated these strategies by measuring FPR using the RA Immunochip Consortium data[30]. In 1,000 different loosely pruned ($r^2 < 0.5$) SNP sets constructed using the Sweden EIRA data (**Online Methods**), the FPR without using delta-correlations was high (22.4% at p<0.05). Applying delta-correlations reduced this FPR to 9.5%. When we used stringent pruning ($r^2 < 0.1$), FPR was appropriately controlled (FPR 5.9% and FPR 5.3% with and without delta correlations, respectively). Although LD pruning alone was sufficiently effective for FPR control in this simulation, we used both strategies throughout the paper to be conservative.

Accounting for population stratification

Another potential confounding factor is population stratification. If population stratification exists, weak correlations between unlinked loci may occur, leading to inappropriate significance. If similar population stratification exists in cases and controls, the use of delta-correlations mitigates this effect. To more aggressively control for the effect of stratification at the individual level, we implemented BUHMBOX to regress out principal components (PCs) from risk allele dosages before calculating correlation statistics. To evaluate this strategy, we simulated extreme population stratification using HapMap[31] data (60 CEU and 60 YRI founders as cases, and 90 JPG+CHB founders as controls; $\lambda_{GC}=26.5$). Unsurprisingly, in 5,000 randomly sampled sets of independent SNPs we observed an inflated BUHMBOX FPR (14.1% at p<0.05). After regressing the effect of ten PCs from risk allele dosages, we observed that the FPR was appropriately controlled (5.7% at p<0.05). As an additional test under a more realistic scenario, we merged genotype data from Northern Europe (Sweden EIRA cohort; 2,762 cases/1,940 controls) and Southern Europe (Spain cohort; 807 cases/399 controls) in the RA Immunochip Consortium case-control dataset[30] (**Online Methods**) to create a highly stratified dataset. In 1,000 sets of randomly sampled independent SNPs, we observed an inflation of the FPR (8.6% at p<0.05); this was appropriately corrected (5.9% at p<0.05) when we regressed out the effect of ten PCs.

Application to autoimmune diseases

Autoimmune diseases share genetic loci[2, 4, 32–36], clustering in specific immune pathways[2, 27, 36]. We used the GRS approach to evaluate shared genetic structure between autoimmune diseases, and then applied BUHMBOX to assess heterogeneity. We obtained individual-level genotype data from the Type 1 Diabetes Genetics Consortium (T1DGC) UK case-control cohort (6,670 cases and 9,416 controls)[37] and the RA Immunochip Consortium's six RA case-control cohorts (7,279 seropositive RA cases and 15,870 controls)[30] (**Online Methods**). We evaluated genetic sharing between a spectrum of autoimmune diseases with T1D and RA. We obtained associated independent loci for all 18 autoimmune diseases ($r^2 < 0.1$, including MHC SNPs) from ImmunoBase (see URLs and

Supplementary Table 3), and tested the association of GRSs for these autoimmune diseases with T1D and RA case status.

We observed substantial genetic sharing between autoimmune diseases. T1D demonstrated significant sharing with alopecia areata (AA), autoimmune thyroid disease (ATD), celiac disease (CEL), Crohn's disease (CRO), juvenile idiopathic arthritis (JIA), primary biliary cirrhosis (PBC), primary sclerosing cholangitis (PSC), RA, Sjögren's syndrome (SJO), systemic lupus erythematosus (SLE), and vitiligo (VIT) (positive association, $p<10^{-4}$). RA exhibited significant sharing with AA, ankylosing spondylitis (AS), ATD, CEL, JIA, PBC, PSC, SLE, systemic sclerosis (SSC), T1D and VIT ($p<10^{-3}$). Overall, GRSs showed significant positive associations for 11 autoimmune diseases each in T1D and RA cohorts, respectively (GRS $p<2.9\times10^{-3}$ [=0.05/17 correcting for 17 diseases tested]; Table 1, Supplementary Table 4). We considered only these traits for subsequent analyses.

To evaluate the degree of heterogeneity necessary to achieve the observed genetic sharing for these autoimmune diseases, we calculated the GRS regression coefficient, which we previously showed approximates the expected heterogeneity proportion π[38] assuming no pleiotropy. Based on the GRS coefficients, we observed π estimates ranging from 0.08–0.76 across the different autoimmune diseases in T1D and from 0.10–0.43 in RA (Figure 4, Table 1).

We estimated the power of BUHMBOX to detect heterogeneity, correcting for 11 tests ($p<4.5\times10^{-3}$). BUHMBOX was well powered for some autoimmune traits; at $\pi=0.2$, four traits had >90% power for T1D, and four traits had >90% power for RA (Figure 5). Despite this, we observed no evidence of heterogeneity at all (corrected p>0.2; Figure 6, Table 1). Our findings suggest that autoimmune diseases share similar risk alleles and pathways with T1D and RA, and not by subgroups of genetically similar cases resulting from misclassifications or molecular subtypes.

Application to subtype misclassifications in RA

RA consists of two subtypes, seropositive and seronegative, with distinct clinical outcomes and MHC associations[38]. These two subtypes are classified by whether patients are reactive to anti-CCP antibody. While anti-CCP testing is specific, its lack of sensitivity can result in some seropositive RA patients being misclassified as seronegative RA[20]. We previously demonstrated that there is shared genetic structure between seropositive and seronegative RA using the GRS approach[38], which could imply misclassifications of up to 26.3% between the two RA subtypes.

We used BUHMBOX to evaluate whether seropositive RA misclassifications are present in a seronegative RA cohort. We used the seronegative RA cohort (2,406 cases/15,870 controls) from the RA Immunochip Consortium[30]. Among 68 RA-associated independent loci, we chose SNPs that are associated to seropositive RA ($p<5\times10^{-8}$) but not seronegative RA ($p>5\times10^{-8}$) in our Immunochip data. This criterion resulted in 14 specific loci exclusively associated to seropositive RA (Supplementary Table 3). The seropositive RA GRS was significantly associated with seronegative RA case status ($\beta=0.30$, $p=1.1\times10^{-10}$). The regression coefficient ($\beta=0.30$) represents an upper bound for π (Figure 4). BUHMBOX

suggested that heterogeneity was indeed present (p=0.008, Figure 6, Table 1, Supplementary Table 4), consistent with potential subtype misclassifications. As a more stringent test, we selected SNPs based on between-RA-subtype heterogeneity test results; for this test we obtained p-values by assigning seropositive RA as cases and seronegative RA as controls. We chose SNPs that are associated to seropositive RA ($p < 5 \times 10^{-8}$) and show nominally significant between-RA-subtype heterogeneity (p<0.05, Supplementary Table 3). Applying BUHMBOX to these 12 loci still showed significant heterogeneity within the seronegative RA cohort (p=0.017).

Application to major depressive disorder and schizophrenia

Current definitions of psychiatric disorders reflect clinical syndromes, with overlapping clinical features. As a result, psychiatric diagnoses for a patient may change as their symptoms evolve[21]. In addition to the potential for misdiagnosis, a subset of true MDD cases may be genetically more similar to schizophrenia. If heterogeneity with respect to schizophrenia risk alleles exists among MDD cases, then genetic studies would suggest evidence of coheritability between the two disorders[17] as has been observed in previous studies[3, 6, 7]. The unintentional inclusion of "schizophrenia-like" MDD cases, due to diagnostic misclassification or genetically distinct subgroups, has been acknowledged and explored as a potential source of bias in coheritability studies by previous investigators[3, 17].

We used BUHMBOX to test for a subgroup of "schizophrenia-like" cases in MDD. If a subset of MDD cases are misdiagnosed and in fact have schizophrenia, or are more genetically similar to schizophrenia, we would expect to see subgroup heterogeneity among MDD cases with respect to schizophrenia risk loci. We first evaluated evidence of shared genetic structure among 90 known schizophrenia associated loci[39] (Supplementary Table 3) in 9,238 MDD cases and 7,521 controls from the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium[40] (Supplementary Table 5). Consistent with previous findings (Supplementary Table 6)[3, 6, 7], the GRS was associated with MDD case status ($p = 1.54 \times 10^{-5}$) indicating shared genetic structure (Figure 4). For the GRS analysis we used a refined subset of the total sample (6,382 MDD cases and 5,614 controls), excluding samples that overlapped with the schizophrenia GWAS[39] (**Online Methods**). Application of cross-trait LDSC[7] to estimate the genetic correlation obtained further evidence of shared genetic structure between MDD and SCZ ($r_g = 0.47$, SE=0.07, $p = 1.61 \times 10^{-10}$), of similar magnitude to previous reports[7]. However, the BUHMBOX p-value was not significant (p=0.28), indicating no excess positive correlations among schizophrenia loci within MDD cases (Figure 6, Supplementary Table 4). Our findings suggest no evidence of a subgroup of schizophrenia-like MDD cases. However, we note that we lacked adequate statistical power to detect heterogeneity in the context of a small heterogeneity proportion. Given the MDD sample size and the number of currently known schizophrenia risk loci, there was 53% power at $\pi = 0.20$ but only 25% power at $\pi = 0.10$ (Figure 5).

DISCUSSION

BUHMBOX can distinguish whether shared genetic structure between traits is the consequence of heterogeneity or pleiotropy based on SNP genotype data alone. It can help to

interpret recent observations of shared genetic structures in complex traits including autoimmune, neuropsychiatric, and metabolic diseases. The intuition behind BUHMBOX is that if heterogeneity exists, independent loci will show non-random positive correlations. Hence, correcting for population structure and long-range LD is critical for this approach to be effective. We emphasize that it is necessary to appropriately interpret the source of heterogeneity, which will depend on the biological and clinical relationship between the two traits. We provide detailed information to guide interpretation in the Supplementary Note.

We demonstrated that genetic sharing between autoimmune diseases is due to pleiotropy, noting that for a few traits we had only modest power (Figure 5). One notable exception was seronegative RA, which might contain misclassified seropositive RA cases. The results presented here demonstrate that seronegative RA is a heterogeneous phenotype with respect to genetic overlap with seropositive RA, bringing clarity to an ongoing debate about the nature of this disease. In contrast we were underpowered to draw more definitive conclusions as to whether a subset of MDD cases are genetically similar to schizophrenia cases; as MDD cohorts increase in size we will be able to reassess more accurately whether smaller proportions of heterogeneity might partially explain observed coheritability. Our results are consistent with recent analyses concluding that pleiotropy between psychiatric diseases is unlikely explained by misclassifications alone[17].

We showed that the power of BUHMBOX is a function of sample size, heterogeneity proportion π, and the number, effect sizes and allele frequencies of loci. Power for subtle heterogeneity ($\pi<0.1$) in current datasets is limited. But, in future studies, increasing sample size and number of known associated loci will augment power. One potential strategy to augment power is to use a polygenic modeling[3, 12, 13] approach, including a larger number of SNPs with less stringent significance thresholds (Supplementary Note and Supplementary Figure 4).

BUHMBOX has certain key caveats. First, it is designed to detect a specific type of heterogeneity resulting from the presence of a subgroup comprising a known second trait. Thus, BUHMBOX cannot currently be applied agnostically to detect the presence of heterogeneity within a dataset. Second, BUHMBOX requires prior knowledge of associated loci and their effect sizes. For diseases with few known loci, BUHMBOX may perform suboptimally. Also, if known effect size estimates are inaccurate, power may decrease because appropriate weighting is crucial (Figure 2). Third, BUHMBOX requires individual-level genotype data for a limited number of loci. Fourth, BUHMBOX can be sensitive to confounding factors. We recommend careful control of LD and population structure using LD pruning and PCs. Fifth, interpretation of the BUHMBOX test statistic is not simplistic. Positive findings indicate the presence of heterogeneity but cannot distinguish between the various causes of this (e.g. misclassifications, molecular subtypes, mediated pleiotropy, ascertainment bias), and negative findings may indicate no heterogeneity or low power. To aid interpretation, BUHMBOX provides a power calculation based on sample size and risk allele information, but it may not always be accurate. For example, if pleiotropy and heterogeneity co-exist, power may be overestimated. Sixth, if the heterogeneity proportion π is small (e.g. 0.05), BUHMBOX's ability to detect heterogeneity is limited. We expect that π will vary between situations, and further clinical and biological investigations are necessary

to uncover true π. Finally, there is the unlikely possibility that real epistasis can manifest as positive signal for BUHMBOX. Broadly, BUHMBOX can be thought of as capturing a specific form of epistasis where risk alleles correlate positively within the additive model. As such, if this specific form of epistasis occurs naturally between D_B-associated SNPs, and if this epistasis structure is shared with D_A, it has the potential to create a significant BUHMBOX test result and confound these analyses. However, this specific type of epistasis seems unlikely; were it present, application of BUHMBOX using D_B-associated SNPs in D_B cases to detect apparent "heterogeneity" might yield a significant result.

When comparing BUHMBOX to existing approaches, we focused on the GRS method. However, the results of our comparison also apply to other existing methods such as mixed-model-based approaches[5, 6] and LD-score-based approaches[7], which are similar to the GRS approach in the sense that they detect both pleiotropy and heterogeneity. We expect that BUHMBOX will complement any of these methods to facilitate interpretation of observed genetic sharing between traits. Our statistical approach may be extended to have application beyond heterogeneity, including identification of missing heritability resulting from this type of heterogeneity[41]. These applications will become more feasible as functional annotations of SNPs advance in the coming years.

ONLINE METHODS

Genetic risk score approach

Given M independent risk loci associated to D_B, we calculated the GRS of individual i as

$$GRS_i = \sum_{j=1}^{M} x_{ij}\beta_j$$

where x_{ij} is individual i's risk allele dosage at marker j, and β_j is the effect size (log odds ratio) of risk allele at marker j for disease D_B. The GRS approach calculates GRSs for all individuals and associates GRSs to the case/control status of D_A. In the logistic regression framework for associating GRSs and D_A status, we can obtain the regression coefficient for GRS (β_{GRS}). We previously showed that β_{GRS} approximates the proportion of D_A cases that are genetically D_B (heterogeneity proportion π), if we assume is no pleiotropy and the GRS association is solely driven by a subgroup[38]. Thus, β_{GRS} represents an upper bound of π.

The BUHMBOX approach

To detect heterogeneity within D_A cases driven by a subgroup that is genetically similar to D_B patients, we utilize the following procedure:

1. Prepare genotype data of D_A cases and controls, and information about SNPs associated to D_B (risk allele, RAF, and OR).

2. Prune SNPs associated to D_B based on LD in control samples (excluding SNPs with $r^2 > 0.1$ or within ± 1Mb of other SNPs)

3. Obtain risk allele dosages of pruned SNPs from D_A cases and controls

4. Regress out PCs from risk allele dosages to obtain residual dosages, each locus at a time

5. Calculate $\mathbf{R}$, the correlation matrix of residual dosages of risk alleles in N cases with D_A and $\mathbf{R}'$, in N' controls

6. Calculate $\mathbf{Y}$, a z-score matrix from delta-correlations:

$$\mathbf{Y} = \sqrt{\frac{N \cdot N'}{N+N'}})\mathbf{R} - \mathbf{R}'$$

7. Calculate the BUHMBOX statistic:

$$S_{BUHMBOX} = \frac{\sum_{i<j} w_{ij} y_{ij}}{\sqrt{\sum_{i<j} w_{ij}^2}}$$

where y_{ij} is the element in $\mathbf{Y}$ at row i and column j. Given M pruned SNPs, (i,j) iterates $M(M-1)/2$ non-diagonal elements of $\mathbf{Y}$. The w_{ij} term is a weighting function that is designed to maximize power, such that (equation (13) in Supplementary Note):

$$w_{ij} = \frac{\sqrt{p_i(1-p_i)p_j(1-p_j)}(\gamma_i-1)(\gamma_j-1)}{((\gamma_i-1)p_i+1)((\gamma_j-1)p_j+1)}$$

where p_i is RAF of SNP i, and γ_i is the OR of SNP i for D_B. The BUHMBOX statistic follows $N(0,1)$ under the null hypothesis. We calculate the significance of this statistic as a positive one-sided test; the p-value is $p_{BUHMBOX} = 1 - \Phi(S_{BUHMBOX})$ where Φ is the cumulative density function of the standard normal distribution. In the context of heterogeneity, excessive positive correlations among D_B risk alleles in D_A cases result in $p_{BUHMBOX} < \alpha$. See Supplementary Table 1 for a comparison of BUHMBOX and GRS approaches. The BUHMBOX test statistic was inspired by previous work deriving covariance between correlation estimates[42] and on combining dependent estimates[43, 44]. For details of the intuition, derivation, optimization, and interpretation of the BUHMBOX test statistic, see Supplementary Note.

Code availability

BUHMBOX has been fully implemented as a publicly available R script (see URLs).

Power and false positive rate simulations

Given sample size of D_A cases (N), proportion of D_A cases that actually show genetic characteristics of D_B (heterogeneity proportion π), and number of risk loci associated to D_B (M), we simulated studies to estimate power of our method as follows. To simulate a reasonable joint distribution of RAFs and ORs, we downloaded the GWAS catalog (as of 29

April 2014). Among all binary traits in the catalog, we selected traits with 50 reported SNPs resulting in 22 traits with 1,480 SNPs. From these SNPs, we sampled M pairs of RAF (p) and their corresponding OR (γ). To simulate genotypes, we set the RAF of a subgroup ($N\pi$ individuals) to $\gamma p/((\gamma-1)p+1)$ and p for the other subgroup ($N(1-\pi)$ individuals), because $N\pi$ individuals can be thought of as D_B cases. Within each subgroup, we generated genotypes assuming that risk alleles are distributed according to the Hardy-Weinberg equilibrium (HWE) and risk loci are independent. We assumed HWE in cases because we assumed an additive disease model. Then we applied BUHMBOX to calculate the p-value. We repeated this 1,000 times to approximate power as the proportion of simulations with p-values 0.05. We evaluated power for different values of N, M, and π.

Under the assumption that the loci are independent, the FPR simulation was equivalent to the power simulation described above with the only difference being that π was set to zero, which forced the null hypothesis. We measured the FPR by assuming $N=1,000$ and $M=20$, and constructing 1,000,000 such studies.

Linkage disequilibrium simulations

To simulate realistic LD, we used chromosome 22 data from control individuals in the Swedish EIRA cohort of the RA dataset (2,762 cases/1,940 controls)[30]. We assigned half of control individuals as cases and the rest as controls. To generate 1,000 random sets of SNPs, we began from all SNPs and thinned the SNP set by 10-fold with different seed numbers using PLINK[45] (with the command --thin 0.1). We then pruned each of the 1,000 datasets using PLINK[45] with r^2 criterion of 0.5 or 0.1.

Population stratification simulations

To assess the effects of population stratification, we conducted two sets of simulations. First, used data from HapMap[31] release 23 data (60 CEU founders, 60 YRI founders, and 90 JPT+CHB founders) setting CEU+YRI as cases and JPT+CHB as controls. We calculated PCs after LD pruning ($r^2<0.1$). For D_B SNPs we randomly selected 5,000 sets of 22 independent SNPs; we selected a single SNP from each autosome. Second, we used genotype data from a Northern Europe RA cohort (Swedish EIRA; 2,762 cases/1,940 controls) and a Southern Europe cohort (Spain; 807 cases/399 controls) from the RA dataset[30]. For this simulation we used SNPs that we had generated for LD simulations (described above, thinned from Swedish EIRA chromosome 22 with criterion $r^2<0.1$), by setting them as cases and adding Spain samples as controls.

Application to specific phenotypes

Type 1 diabetes dataset—To evaluate pleiotropy and heterogeneity between 18 autoimmune diseases and T1D, we applied GRS and BUHMBOX approaches to the UK case-control dataset provided by the T1DGC[37], which consisted of a total of 16,086 samples (6,670 cases and 9,416 controls) from three collections: (1) cases from the UK-GRID, (2) shared controls from the British 1958 Birth Cohort and (3) shared controls from Blood Services controls (data release 4 February 2012, hg18). The samples were collected from 13 regions. All samples were collected after obtaining informed consent, and were genotyped

on the Immunochip array. GRS and BUHMBOX analyses were conducted using the region index as covariates.

Rheumatoid arthritis dataset—To evaluate pleiotropy and heterogeneity between 18 autoimmune diseases and RA, we used the RA Immunochip consortium data from six RA case-control cohorts (UK, US, Dutch, Spanish, Swedish Umea, and Swedish EIRA)[30]. To evaluate pleiotropy to autoimmune diseases, we used 7,279 seropositive RA cases and 15,870 controls. To evaluate misclassifications of RA subtypes, we used 2,406 seronegative RA samples and the same controls. Seropositive and seronegative RA patients were defined in each cohort using standard clinical practices to assess whether patients were reactive to anti-CCP antibody[38]. All samples were obtained with informed consent, and were collected through institutional review board approved protocols. All individuals self-reported as white and of European descent. Samples were genotyped with the Immunochip array. We merged the data of six cohorts into one, and used binary variables representing cohorts as well as 10 PCs as covariates in the analysis.

Defining autoimmune risk loci—We accessed ImmunoBase (7 June 2015 version) to define genome-wide significant risk loci for 18 autoimmune diseases. We did not include inflammatory bowel disease, due to its redundancy with Crohn's disease and ulcerative colitis. For each of the 18 autoimmune diseases analyzed we pruned the list of index SNPs obtained from ImmunoBase in PLINK[45] with options --r2 --ld-window-r2 0.1, using the 1000 Genomes Phase 1 European reference panel for LD. For all pairs of SNPs with $r^2 > 0.1$, we kept the most strongly associated SNP. To ensure completely independent risk loci we also removed SNPs annotated as being located in the same chromosomal region in ImmunoBase, again keeping the most strongly associated index SNP (Supplementary Table 3). When a locus was not in the Immunochip datasets, we looked for a proxy ($r^2 > 0.2$) based on the 1000 Genomes data.

Major depressive disorder dataset—We used BUHMBOX to investigate the relationship between MDD and schizophrenia, which have been previously reported to share genetic etiology based on polygenic risk scoring[3] and coheritability analyses[6]. The full MDD sample analyzed comprised nine GWAS datasets collected from eight separate studies (Supplementary Table 5) as previously described[40]. All samples were collected through institutional review board approved protocols and were obtained with informed consent. Independence of the training (SCZ) and target (MDD) datasets is crucial in GRS analyses; GRSs are constructed using effect size estimates obtained using allele frequency differences between cases and controls in the training GWAS, and overlapping cases or controls will therefore bias the association of GRSs to the target dataset in the positive direction. In contrast the BUHMBOX test statistic is based on the correlation of risk allele dosages among cases, which is orthogonal to allele frequency differences in cases and controls, and is therefore not inflated by sample overlap. Thus, for the GRS analysis individual MDD samples (four cases, 886 controls) that overlapped with those in the schizophrenia GWAS[39] were removed from the analysis; three GWAS cohorts with an insufficient number of independent control samples ($N < 5$) were also removed from the analysis. GRS analyses were conducted in each of the remaining six GWAS datasets (Supplementary Table 5),

followed by meta-analysis of the GRS. To obtain the overall GRS effect size (β) and test statistic we used the inverse-variance weighted fixed effects method. For BUHMBOX, we used the full dataset; analyses were conducted in each of the nine GWAS datasets (Supplementary Table 5) followed by meta-analysis. Because the BUHMBOX statistic is a z-score, we meta-analyzed BUHMBOX results across the datasets using the standard weighted sum of z-score approach, where z-scores are weighted by the square root of the sample size.

Defining schizophrenia risk loci—Schizophrenia associated SNPs were selected as those showing genome-wide significant association with schizophrenia ($p < 5 \times 10^{-8}$) in the most recent Psychiatric Genomics Consortium[39] GWAS. For schizophrenia associated SNPs not directly genotyped in the MDD GWAS datasets, we selected proxy SNPs as those with the highest r^2 from the list of all proxies with $r^2 > 0.2$ using the 1000 Genomes Phase 1 European reference panel. Of the 97 schizophrenia associated SNPs (11 indels were not considered in our analysis), 90 LD-independent SNPs ($r^2 > 0.1$, distance to each other is >1Mb) were available for analysis in the MDD GWAS datasets either via direct genotyping or by a proxy SNP (see Supplementary Table 3 for a detailed list of SNPs).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES FOR MAIN TEXT

1. Sivakumaran S, et al. Abundant pleiotropy in human complex diseases and traits. Am J Hum Genet. 2011; 89:607–618. [PubMed: 22077970]

2. Cotsapas C, et al. Pervasive sharing of genetic effects in autoimmune disease. PLoS Genet. 2011; 7:e1002254. [PubMed: 21852963]

3. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet. 2013; 381:1371–1379. [PubMed: 23453885]

4. Fortune MD, et al. Statistical colocalization of genetic risk variants for related autoimmune diseases in the context of common controls. Nat Genet. 2015; 47:839–846. [PubMed: 26053495]

5. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics. 2012; 28:2540–2542. [PubMed: 22843982]

6. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet. 2013; 45:984–994. [PubMed: 23933821]

7. Bulik-Sullivan B, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015; 47:1236–1241. [PubMed: 26414676]

8. Pendergrass SA, et al. Phenome-wide association study (PheWAS) for detection of pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) Network. PLoS Genet. 2013; 9:e1003087. [PubMed: 23382687]

9. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med. 2015; 372:793–795. [PubMed: 25635347]

10. Criswell LA, et al. Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. Am J Hum Genet. 2005; 76:561–571. [PubMed: 15719322]

11. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ. Major depression and generalized anxiety disorder. Same genes, (partly) different environments? Arch Gen Psychiatry. 1992; 49:716–722. [PubMed: 1514877]

12. Wray NR, Goddard ME, Visscher PM. Prediction of individual genetic risk to disease from genome-wide association studies. Genome Res. 2007; 17:1520–1528. [PubMed: 17785532]

13. Purcell SM, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009; 460:748–752. [PubMed: 19571811]

14. Lee SH, et al. New data and an old puzzle: the negative association between schizophrenia and rheumatoid arthritis. Int J Epidemiol. 2015; 44:1706–21. [PubMed: 26286434]

15. Power RA, et al. Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. Nat Neurosci. 2015; 18:953–955. [PubMed: 26053403]

16. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet. 2013; 14:483–495. [PubMed: 23752797]

17. Wray NR, Lee SH, Kendler KS. Impact of diagnostic misclassification on estimation of genetic correlations using genome-wide genotypes. Eur J Hum Genet. 2012; 20:668–674. [PubMed: 22258521]

18. Silverberg MS, et al. Diagnostic misclassification reduces the ability to detect linkage in inflammatory bowel disease genetic studies. Gut. 2001; 49:773–776. [PubMed: 11709510]

19. van der Linden MP, et al. Value of anti-modified citrullinated vimentin and third-generation anti-cyclic citrullinated peptide compared with second-generation anti-cyclic citrullinated peptide and rheumatoid factor in predicting disease outcome in undifferentiated arthritis and rheumatoid arthritis. Arthritis Rheum. 2009; 60:2232–2241. [PubMed: 19644872]

20. Wiik AS, van Venrooij WJ, Pruijn GJ. All you wanted to know about anti-CCP but were afraid to ask. Autoimmun Rev. 2010; 10:90–93. [PubMed: 20727426]

21. Bromet EJ, et al. Diagnostic shifts during the decade following first admission for psychosis. Am J Psychiatry. 2011; 168:1186–1194. [PubMed: 21676994]

22. Gibson P, et al. Subtypes of medulloblastoma have distinct developmental origins. Nature. 2010; 468:1095–1099. [PubMed: 21150899]

23. Smoller JW, Lunetta KL, Robins J. Implications of comorbidity and ascertainment bias for identifying disease genes. Am J Med Genet. 2000; 96:817–822. [PubMed: 11121189]

24. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. Nature. 2013; 501:338–345. [PubMed: 24048066]

25. Jeste SS, Geschwind DH. Disentangling the heterogeneity of autism spectrum disorder through genetic findings. Nat Rev Neurol. 2014; 10:74–81. [PubMed: 24468882]

26. Flint J, Kendler KS. The genetics of major depression. Neuron. 2014; 81:484–503. [PubMed: 24507187]

27. Cho JH, Feldman M. Heterogeneity of autoimmune diseases: pathophysiologic insights from genetics and implications for new therapies. Nat Med. 2015; 21:730–738. [PubMed: 26121193]

28. Welter D, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 2014; 42:D1001–6. [PubMed: 24316577]

29. Raychaudhuri S, et al. Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk. Nat Genet. 2009; 41:1313–1318. [PubMed: 19898481]

30. Eyre S, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat Genet. 2012; 44:1336–1340. [PubMed: 23143596]

31. The International HapMap Consortium. The International HapMap Project. Nature. 2003; 426:789–796. [PubMed: 14685227]

32. Smyth DJ, et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. N Engl J Med. 2008; 359:2767–2777. [PubMed: 19073967]

33. Festen EA, et al. A meta-analysis of genome-wide association scans identifies IL18RAP, PTPN2, TAGAP, and PUS10 as shared risk loci for Crohn's disease and celiac disease. PLoS Genet. 2011; 7:e1001283. [PubMed: 21298027]

34. Zhernakova A, et al. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS Genet. 2011; 7:e1002004. [PubMed: 21383967]

35. Jostins L, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012; 491:119–124. [PubMed: 23128233]

36. Cotsapas C, Hafler DA. Immune-mediated disease genetics: the shared basis of pathogenesis. Trends Immunol. 2013; 34:22–26. [PubMed: 23031829]

37. Onengut-Gumuscu S, et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat Genet. 2015; 47:381–386. [PubMed: 25751624]

38. Han B, et al. Fine mapping seronegative and seropositive rheumatoid arthritis to shared and distinct HLA alleles by adjusting for the effects of heterogeneity. Am J Hum Genet. 2014; 94:522–532. [PubMed: 24656864]

39. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014; 511:421–427. [PubMed: 25056061]

40. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium. A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry. 2013; 18:497–511. [PubMed: 22472876]

41. Wray NR, Maier R. Genetic basis of complex genetic disease: The contribution of disease heterogeneity to missing heritability. Curr Epidemiol Rep. 2014; 1:220–227.

42. Jennrich RI. An asymptotic χ^2 test for the equality of two correlation matrices. J Am Statist Assoc. 1970; 65:904–912.

43. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. J Am Statist Assoc. 1989; 84:1065–1073.

44. Lin DY, Sullivan PF. Meta-analysis of genome-wide association studies with overlapping subjects. Am J Hum Genet. 2009; 85:862–872. [PubMed: 20004761]

45. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]
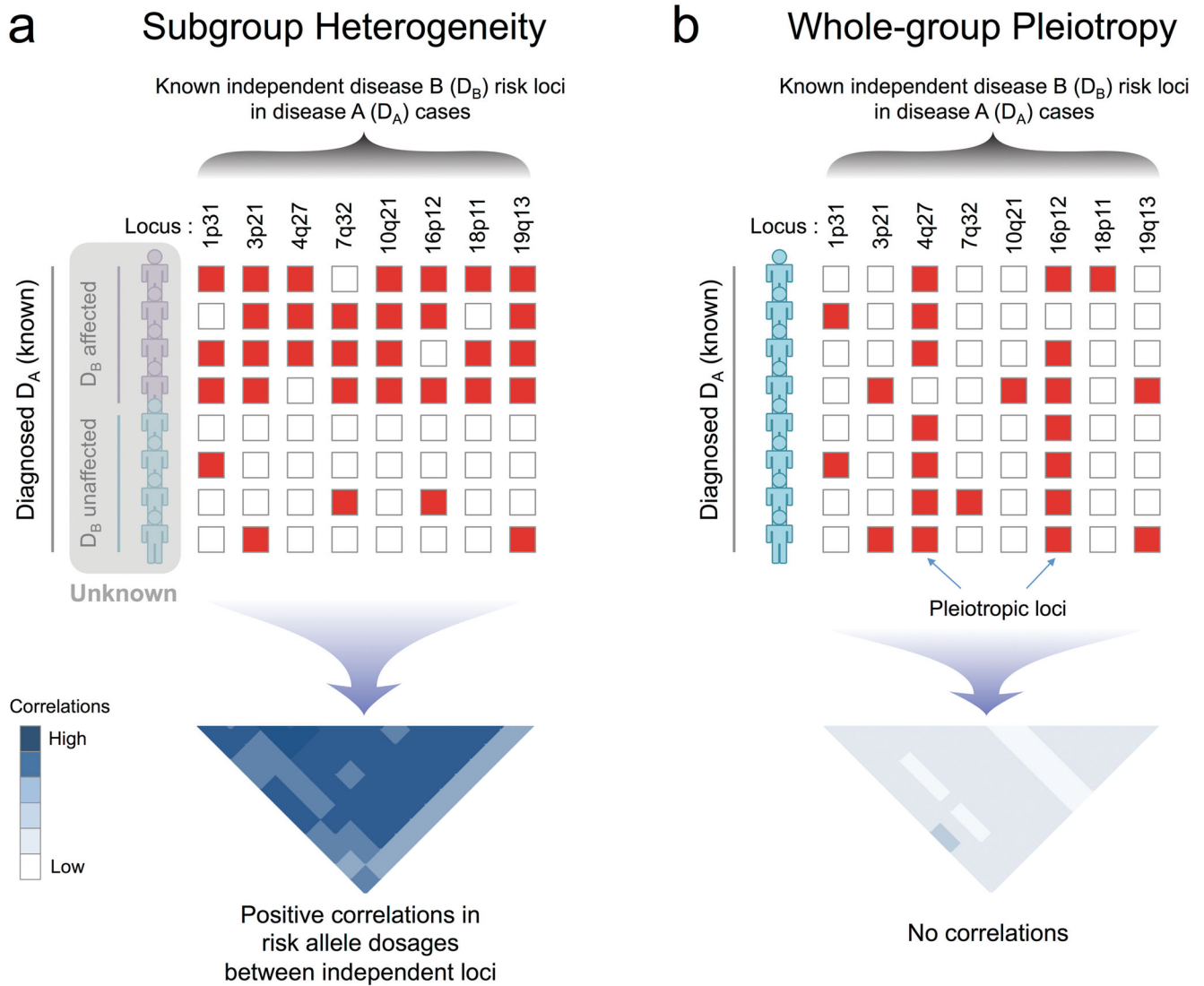
Figure 1. Overview of BUHMBOX

(a) Under the scenario of subgroup heterogeneity, risk alleles of disease B (D_B)-associated loci will be enriched in a subgroup of disease A (D_A) cases, producing positive correlations between D_B risk allele dosages from independent loci. (b) Under the scenario where there is no heterogeneity and D_A and D_B share alleles due to pleiotropy (i.e. whole-group pleiotropy), D_B risk alleles will be uniformly distributed and have no correlations. Red boxes: risk alleles; white boxes: non-risk alleles.
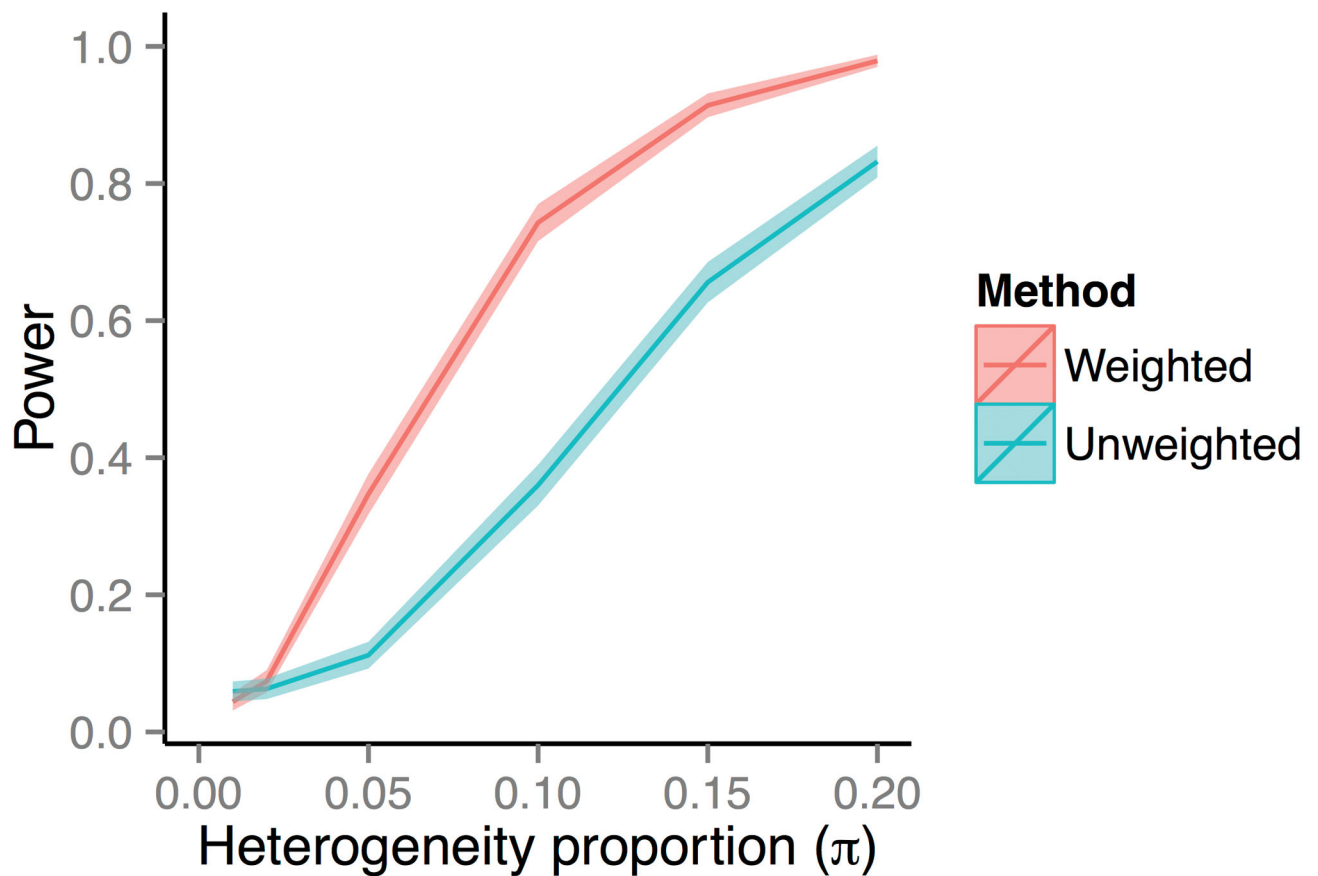
Figure 2. Power gain by weighting SNPs by allele frequency and effect size
We compared the statistical power of BUHMBOX with a weighting scheme that optimally weights correlations between SNPs (weighted) to an alternative approach that weights correlations uniformly (unweighted; equation (12) in Supplementary Note). We simulated 1,000 case individuals and assumed 50 risk loci, whose OR and RAFs were sampled from the GWAS catalog. Colored bands denote 95% confidence intervals of power estimates.
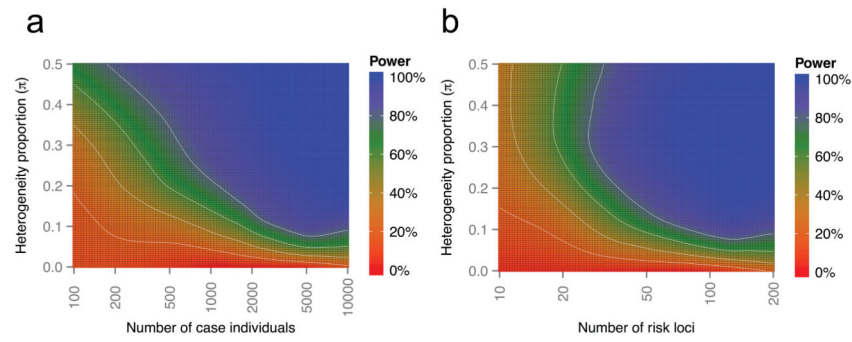
Figure 3. BUHMBOX power analysis

Power of BUHMBOX for detecting heterogeneity as a function of the number of risk loci, number of case samples, and the proportion of samples that actually have different phenotype (heterogeneity proportion, π). We assume that we have the same number of controls as cases. White lines denote 20, 40, 60, and 80% power. (**a**) Power as a function of number of case individuals and heterogeneity proportion, when the number of risk loci is fixed at 50. (**b**) Power as a function of number of risk loci and heterogeneity proportion, when the case sample size is fixed at 2,000.
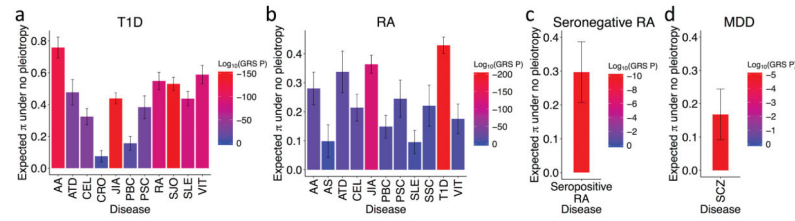
Figure 4. Genetic sharing between autoimmune diseases and psychiatric disorders
In (a) and (b), we show only the diseases that have significantly positive GRS p-values out of the 17 tested. Y-axis denotes the expected heterogeneity proportion (π) to explain observed genetic sharing. Vertical bars indicate 95% confidence intervals. Heterogeneity proportion estimates are based on GRS analysis, assuming no pleiotropy for (a) T1D, (b) RA, (c) seronegative RA, and (d) MDD.

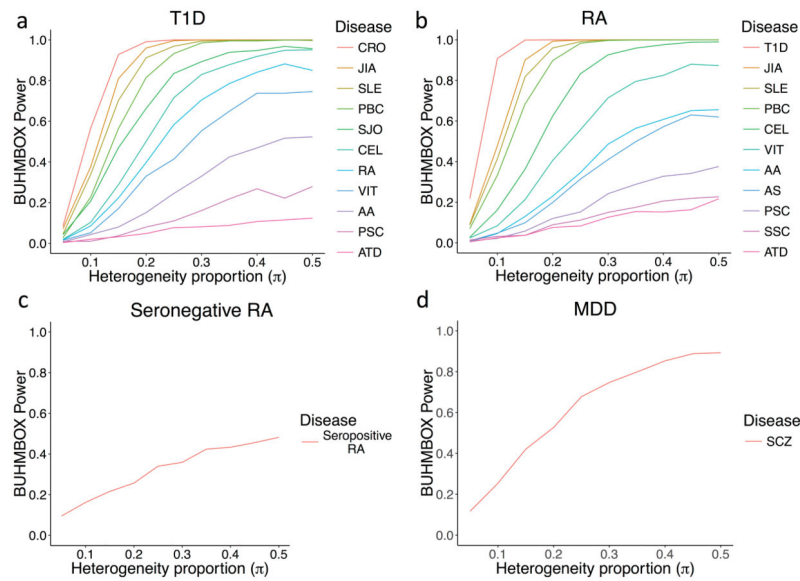**Figure 5. Statistical power of BUHMBOX to detect heterogeneity**

We calculated power by performing 1,000 simulations with corresponding sample size, number of risk alleles, risk allele frequencies, and odds ratios. To calculate power for (c) and (d), we used a significance threshold of 0.05. For (a) and (b), the threshold was adjusted using the Bonferroni correction accounting for 11 tests in T1D and RA, respectively.
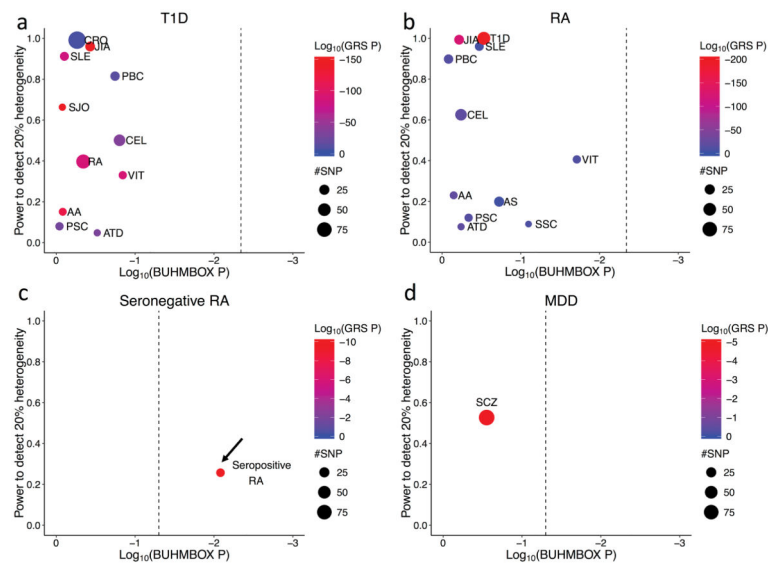
Figure 6. BUHMBOX results

We show only diseases with significantly positive GRS p-values (for complete results for all traits tested, see Supplementary Table 4). Significant GRS p-values indicate evidence of shared genetic structure; significant BUHMBOX p-value indicates evidence of heterogeneity. Point size represents the number of D_B-associated SNPs included in the analysis. Dashed vertical lines denote the Bonferroni-adjusted significance threshold for the BUHMBOX test statistic. Arrow indicates significant BUHMBOX test statistic.

Table 1

Summary of genetic overlap using GRS and BUHMBOX

Only the traits that have significant GRS p-values in positive directions are shown. Significant GRS p-value indicates evidence of shared genetic structure; significant BUHMBOX p-value indicates evidence of heterogeneity. See Supplementary Table 4 for the full results for all traits tested.

Cohort data	Test trait	#SNP	GRS p-value	GRS Beta (95% CI)	BUHMBOX p-value	BUHMBOX power at π=0.20
T1D	AA	10	1.4×10^{-120}	0.76 (0.69 – 0.82)	0.83	0.15
	ATD	7	1.4×10^{-31}	0.48 (0.40 – 0.56)	0.30	0.05
	CEL	38	2.2×10^{-35}	0.32 (0.27 – 0.38)	0.16	0.50
	CRO	119	2.4×10^{-05}	0.08 (0.04 – 0.11)	0.54	0.99
	JIA	22	3.6×10^{-151}	0.44 (0.40 – 0.47)	0.37	0.96
	PBC	19	1.1×10^{-12}	0.16 (0.11 – 0.20)	0.18	0.82
	PSC	12	4.1×10^{-26}	0.38 (0.31 – 0.45)	0.91	0.08
	RA	68	6.6×10^{-89}	0.55 (0.49 – 0.60)	0.45	0.40
	SJO	7	3.9×10^{-146}	0.53 (0.49 – 0.57)	0.84	0.66
	SLE	16	1.1×10^{-83}	0.44 (0.39 – 0.48)	0.79	0.91
	VIT	12	2.5×10^{-90}	0.59 (0.53 – 0.65)	0.14	0.33
RA	AA	10	1.5×10^{-22}	0.28 (0.22 – 0.34)	0.71	0.23
	AS	24	6.1×10^{-04}	0.10 (0.04 – 0.15)	0.19	0.20
	ATD	7	3.9×10^{-20}	0.34 (0.27 – 0.41)	0.57	0.08
	CEL	38	6.4×10^{-20}	0.21 (0.17 – 0.26)	0.57	0.63
	JIA	22	8.9×10^{-125}	0.36 (0.33 – 0.39)	0.61	0.99
	PBC	19	1.5×10^{-13}	0.15 (0.11 – 0.19)	0.83	0.90
	PSC	12	6.2×10^{-14}	0.24 (0.18 – 0.31)	0.46	0.12
	SLE	16	4.3×10^{-06}	0.10 (0.05 – 0.14)	0.34	0.96
	SSC	5	9.6×10^{-10}	0.22 (0.15 – 0.29)	0.08	0.09
	T1D	53	9.6×10^{-207}	0.43 (0.40 – 0.46)	0.29	1.00
	VIT	12	1.8×10^{-11}	0.18 (0.12 – 0.23)	0.02	0.41
Seroneg.RA	Seropos.RA	14	1.1×10^{-10}	0.30 (0.21 – 0.39)	0.008	0.26

Cohort data	Test trait	#SNP	GRS p-value	GRS Beta (95% CI)	BUHMBOX p-value	BUHMBOX power at π=0.20
MDD	SCZ	90	1.5×10^{-5}	0.17 (0.09 − 0.24)	0.28	0.53

AA, Alopecia areata; AS, Ankylosing spondylitis; ATD, Autoimmune thyroid disease; CEL, celiac disease; CRO, Crohn's disease; JIA, juvenile idiopathic arthritis; MS, multiple sclerosis; PBC, primary biliary cirrhosis; PSC, primary sclerosing cholangitis; SJO, Sjögren's syndrome; SLE, systemic lupus erythematosus; SSC, Systemic sclerosis; UC, ulcerative colitis; VIT: Vitiligo; MDD, major depressive disorder; SCZ, schizophrenia; Seroneg., seronegative; Seropos., seropositive.