Supplementary Note

$h^2_{\mathbf{g}}$ from narrowed DHS peaks

Based on the hypothesis that most regulatory sites lie at the center of the called DHS peaks, we considered the enrichment after progressively narrowing the DHS annotations. Specifically, we trimmed the ends of each DHS peak to a maximum length such that the resulting DHS annotation size is 1%, 5%, and 10% of the genome (without removing any individual peaks). We then tested these three narrowed annotations in two models: a univariate model where h^2_{g} is inferred from only the narrowed DHS component, thereby including any tagged heritability from other functional categories; and a six component model where the full DHS component was replaced with the narrowed DHS component and remaining DHS SNPs distributed into the intron and other components. In both models, we find the DHS centers to capture substantially more heritability than their size (Table S6), with the 1% annotation explaining 61.0% of the total h^2_{g} in the univariate model and 19.8% of the total h^2_{g} in the multivariate model ($P = 2.6 \times 10^{-6}$). For comparison, the coding component covering roughly 1% of the genome explains 30.0% of the total h^2_{g} in a univariate model.

Information content of functional enrichment

Our estimates of the relative significance of different h^2_{g} enrichment scenarios were directly dependent on the standard error and overall sample size analyzed. Here, we consider an alternative figure of merit which relies only on the fraction of h^2_{g} in each category. We borrow from information theory the concept of entropy, which is a measure of uncertainty in the distribution of a random variable. Given $P(X_i)$, the probability mass function of a random variable, entropy can be quantified as $H = -\sum_{i=1}^{a} P(X_i)\log P(X_i)$. Depending on the distribution and log-base, this is equivalent to the number of bits required to encode an observation, with higher entropy implying lower predictability. Applying this to functional categories, we define $P(X_i)$ as the normalized probability of a SNP in the category being causal, equal to the product of the % h^2_{g} and the % SNPs for that category. We then compute the entropy as outlined previously. Table S20 demonstrates the resulting entropy from multiple enrichment scenarios, with entropy inversely correlated to the individual category significance. Highest entropy was observed for an enrichment scenario that only accounted for the (least significant) promoter category, and lowest entropy was observed for an enrichment scenario that accounted for all six categories. Interestingly, the six-category genotyped enrichment yielded higher entropy than a hypothetical DHS-only imputed enrichment. This formulation of "functional entropy provides a standard metric for comparing real and hypothetical enrichment scenarios completely independent of sample size and data platform.

Robustness of variance-component estimates

Jackknife estimates of $h^2_{\mathbf{g}}$ variance

The analytical standard error used for significance testing was accurate in our simulations (Table S21) and has previously been shown to be robust in real data [1], but can be biased when the number of causal variants is very small [2]. We assessed this directly with a weighted block-jackknife estimate of the enrichment by dropping each chromosome in turn and re-computing the h^2_{g} [2,3]. This estimate also captures true variation in enrichment across the chromosomes, and is therefore highly conservative. Though we observe little difference in genotyped data (Table S8), the jack-knife estimate of imputed % h^2_{g} (71% s.e. 7.7%, Table S9) is indeed more conservative than the analytical estimate (79% s.e. 6.6%), but the enrichment is still highly significant ($P = 5.5 \times 10^{-13}$) and the overall results not substantially effected.

Impact of shared controls

We evaluated potential biases due to the use of shared controls by shifting the functional categories and performing the entire genotyped meta-analysis procedure to achieve an empirical null distribution. Specifically, over 1,000 consecutive indices, we shifted all functional annotations ahead by 2MB (moving regions that crossed the chromosome boundary into the next chromosome) thereby preserving the total h^2_g, total sample relatedness, and relative dependence between categories but permuting any relationship to true function. For each shifted annotation, we re-computed GRMs from the genotyped data and estimated functional enrichment within each trait, as well as the meta-analysis value across all 11 traits, yielding $1,000 \times 6$ shifted meta-analysis estimates. We observed no inflation of p-values within each study, further supporting the robustness of the empirical standard error. As expected, we did observe inflation in the meta-analysis p-values ranging from λ_{GC} of 1.26 (coding) to 1.70 (intergenic). We adjusted the standard errors observed in real data by the corresponding $\sqrt{\lambda_{GC}}$, which yielded adjusted p-values that remained significant for all categories but UTR (Table S10).

Impact of case-control ascertainment

Recent work[4–6] has shown that liability-scale estimates of h^2_g from REML can be biased downward in dichotomous traits with strong ascertainment.[4,5] propose an alternative estimator based on Haseman-Elston regression[7] and show that it eliminates bias. Briefly, this approach regresses the product of normalized phenotypes on the genetic covariance (off-diagonal GRM entries) for all unique pairs of samples; with the resulting slope used as an estimate of observed-scale h^2_g and converted to liability-scale. This method can be extended naturally to multiple components, where the product of phenotypes is regressed onto GRM entries from each analyzed component in a multiple linear regression. Here, we compared the method and transformation of[5] to the transformed REML estimator described in the main text. We also evaluated the impact of incorporating principal components as fixed-effects to account for genetic ancestry. This is particularly important for the schizophrenia (SP) and multiple sclerosis (MS) cohorts, which were ascertained in a way that induces correlations between ancestry and phenotype. All analyses were performed using the same set of GRMs computed from 1000G imputed data, with regression and regression fixed-effects implemented as described in[5]. In all instances, analytical error covariance estimates were used and rescaled with the delta method to compute standard errors. Note that the standard error for regression makes assumptions about independence that are strongly violated and are therefore only presented for completeness.

We observed little difference between the two methods, with regression yielding an average estimate of $1.05\times$ higher than REML and an overall $R^2 = 0.95$ between the two methods (across 11 traits, Table S11). The relative performance was similar when considering only the % h^2_g from the DHS component (Tabel S12), with regression yielding $1.04\times$ higher estimates than REML on average and an overall $R^2 = 0.94$. Meta-analysis across traits within each method did not yield significant differences, with regression identifying DHS enrichment of $5.8\times$ (s.e. 0.45) compared to REML identifying DHS enrichment of $5.1\times$ (s.e. 0.42). A large difference between the two methods was observed in the SP and MS cohorts without fixed effects, where liability-scale regression estimates were 10.00 and 2.91 respectively (Table S11), significantly higher than REML without fixed-effects. This suggests that regression-based estimates may be particularly sensitive to the confounding effects of ancestry.

Potential confounding from principal components

Recently,[8] demonstrated that h^2_g can vary significantly when principal components are also included as fixed-effects, as a function of the number of included eigenvectors. To assess the presence of this bias in our data, we re-compute the previous joint estimates of $\hat{h}^2_g$ with an increasing number of eigenvectors included as fixed-effects. We observe no significant fluctuation of the $\hat{h}^2_g$, with the average estimate over 1-20 eigenvector covariates of 0.184 having a standard deviation of 0.002 suggesting a tight estimate unbiased by the fixed effects.

Unbiased estimates of h^2_g from rare and common variants

Variant normalization

We considered weather the SNPs used in construct the GRM should be normalized by their observed variance or the expected variance $2p(1-p)$ based on the minor allele frequency p. We performed simulations for the two normalization schemes and two effect-size distributions. Under the infinitesimal model where every variant explains the same amount of phenotypic variance in expectation, we observed no differences between the normalizations for any class of SNPs. Under the neutral model where effect-size is proportional to the minor allele frequency, we observed a significant difference between the two normalizations when rare variants were included in the analysis, with the $2p(1-p)$ scaling resulting in a significant upwards bias. These findings indicate that rare variants have slight but consistent deviations from Hardy-Weinberg equilibrium that can affect the variance-component estimate under the $2p(1-p)$ normalization. To account for this, we use the observed variance to normalize markers in all analyses of rare variants.

LD-induced bias

Previous work using GWAS chip data has shown that genetic architectures with systematically unusual LD patterns at causal variants can yield biased estimates of h^2_g, and that this bias can be reduced by adjusting the input GRM to account for LD [1,2,9]. Unlike GWAS chip data, which is a relatively uniform sample of common variants, exome chip consists predominantly of densely-typed rare variants in short exons (Table S18), potentially exacerbating this bias.

As in previous work, we evaluate potential biases by simulating phenotypes from diverse disease architectures using the real exome chip variants. We randomly sampled 1,000 causal variants from coding SNPs with minor allele frequency below a fixed threshold ranging from 0.01 to 0.1. Effect sizes for each causal variant were drawn from the standard Normal and applied to the normalized SNP (such that all SNPs explain the same variance in expectation) with random noise added to yield an $h^2_g = 0.5$. We then analyzed the performance of a single variance component which accounts for all SNPs versus two jointly modeled components [9] corresponding to rare (MAF < 0.01) and common (MAF ≥ 0.01) SNPs, where we compute the h^2_g estimate as the sum of the corresponding $h^2_{g,rare}$ and $h^2_{g,common}$. In both scenarios we also consider the impact of LD adjusting the GRMs internally using the LD-residual method [1] (h^2_{gLD}), resulting in a total of four inference models.

Figure S12 shows the distribution of inferred h^2_g over the 10 disease architectures and 4 inference models. Under the un-adjusted single-component model - corresponding to the typical variance-components strategy - we observe both kinds of bias depending on the causal allele frequency cutoff. When causal variants are primarily rare (MAF ≤ 0.02) the mean estimate is significantly deflated down to 0.45, whereas when causal variants are more common (MAF ≤ 0.1) the mean estimate is significantly deflated up to 0.59. LD adjustment of the single component appears to fix the downwards bias, with mean estimate no lower than 0.49 (not significantly different from 0.50) but does not completely mitigate the upwards bias, with a mean estimate up to 0.57. On the other hand, splitting the data into two components for rare and common SNPs entirely removes the upwards bias but introduces downwards bias in most instances where causal variants can be common. Combining the two strategies and using two internally LD-adjusted components yields completely unbiased estimates with no disease architecture exhibiting h^2_g significantly different from 0.5. A simulation where effect-sizes were applied to the un-normalized SNP directly (such that rare SNPs explained less variance in expectation) showed similar patterns (Fig. S13), with the two-component, LD-adjusted strategy resulting in highest overall accuracy but slight downwards bias when common variants were primarily causal.

Exome heritability "contaminated" by non-coding SNPs

Another potential source of confounding when estimating exome h^2_g is heritability from nearby non-coding variants that is tagged by exonic variants due to LD. Because our interest is in identifying the purely exonic

contribution to phenotype, we consider the heritability from these non-coding variants to "contaminate" our estimates. Using the GWAS chip data from this cohort allows us to quantify the amount of contamination expected due to common non-coding SNPs.

We simulated a standard polygenic phenotype with $h^2 = 0.50$ coming exclusively from 5,000 randomly selected GWAS chip non-coding SNPs and then inferred h^2_g using variance-components constructed from coding SNPs. No coding SNPs were used to generate the phenotypes, and if no contamination was present we expect the inferred h^2_g to equal zero. However, we found that all coding variants together accounted for an average of 17.4% of the non-coding heritability (Table S22), significantly different from zero. This further broke down to slight but non-significant contamination of 2.7% at rare coding variants (MAF < 0.01) and a highly significant average of 11.8% from common coding variants (MAF ≥ 0.01), consistent with common variants being generally better tags of nearby common variation. Given the small physical size of the exome, contamination of 11.8% of the non-coding heritability could substantially bias the estimates from coding variants when estimated directly from exome chip data. To account for this contamination, we model an additional component consisting of the non-coding GWAS variants. When we conditioned in this way and estimate using a three variance-component model, we see statistically zero heritability attributed to the rare and common coding components. Because we only have genome-wide GWAS chip data available, which does not include rare variants and these variants are notoriously difficult to impute, the non-coding component is unlikely to account for contamination from rare non-coding variants. However, our simulations show that rare variants not significantly contaminated by common variants, and therefore even less likely to be contaminated by other rare variants.

Exome heritability tagged by non-coding SNPs

We set out to estimate the fraction of exome h^2 that is tagged by non-coding SNPs from the GWAS chip and 1,000 Genomes imputation. We simulate two groups of standard additive phenotypes from the rare and common exome variants, respectively, and infer $h^2_{\text{g,non-coding}}$ of these phenotypes from the non-coding SNPs. The ratio of $\hat{h}^2_{\text{g,non-coding}}$ to simulated $h^2_{\text{g,exome}}$ gives us an estimate of the fraction of exome heritability tagged by non-coding variants. In 10 simulations from chromosome 22 with $h^2_{\text{g,exome}} = 0.5$ the average ratio is 0.85 for common coding variants and 0.11 for rare coding variants (Table S23). However, the tagging between components is fully accounted for by a joint, three component model (Table S24).

h^2_g of known or candidate variants

PolyPhen2 functional prediction

We observed a significant enrichment in h^2_g at 6,600 loss of function variants, which collectively account for 5.3% of the exonic SNP variance but explain 24.3% of the exonic h^2_g (permuted $P = 0.02$). We saw no significant enrichment of h^2_g at coding sites that were predicted to be functionally important by PolyPhen2 [10]. Comparing likelihoods between a model where variants were split into probably/damaging, benign/other, and non-coding components to the model with only coding and non-coding components yielded no significant difference by 1df LRT ($P = 0.13$).

Known schizophrenia GWAS loci

Having identified no significant rare-variant h^2_{gLD} at all coding regions, we are interested in quantifying this phenomena at the set of loci known to be associated with schizophrenia. To do so, we construct variance-components only from SNPs at the 22 identified by the PGC in a large meta-analysis [11] and estimate them jointly with a component for the remaining non-coding variants genome-wide. As expected, we find the union of all non-coding GWAS variants at these loci to harbor significant heritability of 0.018 (0.004) (Table S29). However, we do not see any significant heritability from the coding variants at these classes when modeled

jointly with the other component. This is consistent with our genome-wide finding that common non-coding variants explain a substantial fraction of trait heritability and tag nearly half of the common coding variation.

Known psychiatric disease genes

We partitioned h^2_g at the set of 1,796 "composite genes reported by Purcell et. al.[12] to exhibit enrichment of rare disruptive mutations, modeled jointly with exome chip variants in the remaining genes and non-coding GWAS chip variants as separate components. However, no significant h^2_g was observed at either the entire set of composite variants ($h^2_g = 0.014$ s.e. 0.012) or the rare composite variants ($h^2_g = 0.008$ s.e. 0.012).

Estimating collapsed-variant heritability

For a given cohort, the variance of the heritability estimate tends to grow with the number of markers analyzed. Borrowing from gene-based burden association tests[13,14], we considered a strategy for reducing the variance of this estimate by collapsing rare variants in a gene into a single polymorphic site when computing the GRM. Over the full data-set, this procedure collapses the 60,000 effective SNPs into approximately 16,000 genes that contain polymorphic SNPs. This technique also has the benefit of incorporating singleton variants that violate the traditional variance-components model normality assumptions. However, as with burden-tests, the model assumes that all SNPs have identical normalized effect-sizes and will exhibit downwards bias when this assumption is violated.

Formally, the method recodes each gene as a multi-allelic "pseudo-SNP" where samples that carry a minor allele below frequency threshold $f_{\max}$ are considered carriers of the pseudo-SNP allele equal to the number of such variants they carry. The pseudo-SNPs are then normalized to have mean=0 and variance=1 and a new GRM is computed over the normalized pseudo-SNPs as in the standard model. The corresponding measure of $h^2_{g,\text{collapsed}}$ is estimated from this collapsed variance-component, jointly with a single non-coding component, which fully accounts for the minimal tagging of h^2_g from non-coding regions by collapsed variants (Table S25). Our simulations show that disease architectures with $> 50\%$ non-causal (or non-deleterious) variants capture substantially less heritability as to make this approach underpowered compared to the standard model considering all SNPs (Table S26, S27). The exome chip was designed with primarily non-synonymous variants, and we did not assess differences according to variant class.

In the collapsed analysis of schizophrenia, we observe a substantial reduction in standard error but do not find any allele-frequency threshold that yields a result significantly different from zero (Table S28). This does not invalidate the use of collapsed-gene burden tests for association and genetic mapping because the individual collapsed gene is still a fundamentally informative unit of association. It does, however, demonstrate that the maximum variance that can be explained by such methods is guaranteed to be substantially lower than that of association with the full model, as has been shown in previous analyses of burden tests[15]. For singleton variants, we can place a 95% upper bound on collapsed h^2_g at 0.014. This is consistent with the recent observation from exome sequencing that the burden of rare coding variants in a subset of 1,796 enriched genes explains 0.4%-0.6% of the variance in schizophrenia[12], and indicates that additional h^2_g could be identified by considering all rare variants in a variance-components model.

Functional enrichment from GWAS summary statistics

To emulate a single large GWAS study, we merged all of the imputed WTCCC2 traits into a single cohort of 32,752 samples and an intersection of 4,594,547 imputed variants. As in previous simulations, we generated 50 quantitative phenotypes with total $h^2 = 0.50$ by sampling causal variants from imputed DHS and coding categories with 79% and 8% (respectively) and all other categories uniformly, for a total of 8,300 causal SNPs. For each simulated phenotype, we then computed standard χ^2 statistics overall the imputed SNPs.

We followed the method described in Schork et. al.[16] to construct stratified QQ-plots. Using the European 1000 Genomes samples as a reference, we computed the sum of r^2 correlations between each GWAS SNP and

any neighboring variant (within 1Mbp, including the SNP itself) belonging to a non-intergenic functional category, so that every GWAS SNP had five LD-based scores. A variant was then considered part of a category if the corresponding score was ≥ 1. As in [16], intergenic variants were defined as those having a score of zero to every other category. Association statistics for each category were divided by the λ_{GC} observed in intergenic variants and QQ-lines computed. Similarly, we followed the method described in Maurano et. al.[17] to quantify p-value enrichment. Over increasingly restrictive p-value thresholds, we computed the fraction of SNPs passing a given threshold that belong to each category, divided by the genome-wide fraction of SNPs in that category. We note that [17] only considered non-coding variants, but examined all markers. For both algorithms, the mean and standard error were computed separately for each p-value bin over 50 simulations.

To ensure that the enrichment at significant loci was consistent with the genome-wide estimates, we partitioned h^2_{g} from SNPs lying within 1Mb of published GWAS loci for each trait (see Web Resources) (Fig. 8). Due to a small number of loci for some traits, the DHS component was jointly analyzed with only a single component including all non-DHS SNPs. We note that the choice of region size may impact the absolute enrichment, with larger regions expected to appear more like the genome-wide enrichment and yield a conservative estimate of the difference. We again observed a highly significant DHS enrichment in imputed data as well as a significant difference between the genotyped and imputed results ($P = 7.9 \times 10^{-18}$). Indeed, the DHS enrichment of genotyped SNPs at known loci was not statistically significant (1.1x, P=0.46) and we observed a marginally significant difference between the DHS enrichment at known loci versus genome-wide in the imputed data (3.6x versus 5.5x, P=0.004). This difference suggests that loci harboring large-effect, genome-wide significant SNPs may be less enriched for DHS variants than the rest of the genome. We stress that this p-value does not survive adjustment for multiple testing and the previously described biases in h^2_{g} and is only suggestive.

Expected risk prediction accuracy

We computed the expected GBLUP prediction accuracy using the previously derived [18,19] relationship that M effective SNPs, N training samples, and h^2_{g} are expected to yield prediction $r^2 = (h^2_g h^2_g)/(h^2_g + M/N)$. We did not account for ascertainment because prediction was assessed by cross-validation. For the PGC analysis, the observed-scale $h^2_g = 0.49$, $N = 10000$ and we assumed $M = 60000$, which is expected to yield genome-wide $r^2 = 0.037$. Assuming independent variance-components, we similarly estimated expected r^2 of the functionally stratified predictor by evaluating (jointly estimated) component-specific h^2_{g} directly in the data, estimating M from the fraction of SNPs in each component, and summing all of the functional expected r^2 to compute the genome-wide prediction. For the PGC analysis, this yielded an expected genome-wide $r^2 = 0.077$, or a 2.08× increase over the standard predictor. However, in the actual data we observed a genome-wide $r^2 = 0.043$ and a stratified $r^2 = 0.046$ (OLS R^2 reported here, Nagalkerke R^2 reported in main text for consistency with published estimates). This increase of only 1.07× is much lower than expected, indicating that the assumption of component independence is strongly violated and significant enrichments in component h^2_{g} do not necessarily translate into increased prediction accuracy.

References

[1] A. Gusev, G. Bhatia, N. Zaitlen, B. Vihjalmsson, D. Diogo, E. Stahl, P. Gregersen, J. Worthington, L. Klareskog, S. Raychaudhuri, R. M. Plenge, B Pasaniuc, and AL. Price. Quantifying missing heritability at known gwas loci. *PLoS Genetics*, 2013.

[2] Doug Speed, Gibran Hemani, Michael Johnson, and David Balding. Improved heritability estimation from genome-wide snps. *Am J Hum Genet*, 91(6):1011–1021, 2012. ISSN 00029297. doi: 10.1016/j.ajhg.2012.10.010.

[3] Frank MTA Busing, Erik Meijer, and Rien Van Der Leeden. Delete-m jackknife for unequal m. *Statistics and Computing*, 9(1):3–8, 1999.

[4] AL Price, N Zaitlen, B Vilhjalmsson, T Hayeck, S Pollack, J Yang, GB Chen, M Goddard, P Visscher, and N. Patterson. Enabling mixed model association analysis for large case-control studies. In *62nd Annual Meeting of the American Society of Human Genetics*, Nov 2012.

[5] D. Golan and S. Rosset. Narrowing the gap on heritability of common disease by direct estimation in case-control GWAS. *ArXiv e-prints*, May 2013.

[6] J Yang, N Zaitlen, M Goddard, P Visscher, and A. Price. Mixed model association methods: advatnages and pitfalls. *Nat Genet*, 2014.

[7] J.K. Haseman and R.C. Elston. The investigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics*, 2(1):3–19, 1972. ISSN 0001-8244. doi: 10.1007/BF01066731.

[8] Luc Janss, Gustavo de Los Campos, Nuala Sheehan, and Daniel Sorensen. Inferences from genomic models in stratified populations. *Genetics*, 192(2):693–704, October 2012. ISSN 1943-2631 (Electronic); 0016-6731 (Linking). doi: 10.1534/genetics.112.141143.

[9] S Hong Lee, Jian Yang, Guo-Bo Chen, Stephan Ripke, Eli A Stahl, Christina M Hultman, Pamela Sklar, Peter M Visscher, Patrick F Sullivan, Michael E Goddard, and Naomi R Wray. Estimation of snp heritability from dense genotype data. *Am J Hum Genet*, 93(6):1151–5, Dec 2013. doi: 10.1016/j.ajhg.2013.10.015.

[10] Ivan A. Adzhubei, Steffen Schmidt, Leonid Peshkin, Vasily E. Ramensky, Anna Gerasimova, Peer Bork, Alexey S. Kondrashov, and Shamil R. Sunyaev. A method and server for predicting damaging missense mutations. *Nat Methods*, 7(4):248–249, Apr 2010. doi: 10.1038/nmeth0410-248. URL `http://dx.doi.org/10.1038/nmeth0410-248`.

[11] Stephan Ripke, Colm O'Dushlaine, Kimberly Chambert, Jennifer L Moran, Anna K Kahler, Susanne Akterin, Sarah E Bergen, Ann L Collins, James J Crowley, Menachem Fromer, Yunjung Kim, Sang Hong Lee, Patrik K E Magnusson, Nick Sanchez, Eli A Stahl, Stephanie Williams, Naomi R Wray, Kai Xia, Francesco Bettella, Anders D Borglum, Brendan K Bulik-Sullivan, Paul Cormican, Nick Craddock, Christiaan de Leeuw, Naser Durmishi, Michael Gill, Vera Golimbet, Marian L Hamshere, Peter Holmans, David M Hougaard, Kenneth S Kendler, Kuang Lin, Derek W Morris, Ole Mors, Preben B Mortensen, Benjamin M Neale, Francis A O'Neill, Michael J Owen, Milica Pejovic Milovancevic, Danielle Posthuma, John Powell, Alexander L Richards, Brien P Riley, Douglas Ruderfer, Dan Rujescu, Engilbert Sigurdsson, Teimuraz Silagadze, August B Smit, Hreinn Stefansson, Stacy Steinberg, Jaana Suvisaari, Sarah Tosato, Matthijs Verhage, James T Walters, Multicenter Genetic Studies of Schizophrenia Consortium, Psychosis Endophenotypes International Consortium, Wellcome Trust Case Control Consortium 2, Elvira Bramon, Aiden P Corvin, Michael C O'Donovan, Kari Stefansson, Edward Scolnick, Shaun Purcell, Steven A McCarroll, Pamela Sklar, Christina M Hultman, and Patrick F Sullivan. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet*, 45(10):1150–1159, October 2013. ISSN 1061-4036. doi: 10.1038/ng.2742.

[12] A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*, 506(7487):185–190, February 2014. doi: 10.1038/nature12975.

[13] Bingshan Li and Suzanne M. Leal. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet*, 83(3):311–321, Sep 2008. doi: 10.1016/j.ajhg.2008.06.024. URL `http://dx.doi.org/10.1016/j.ajhg.2008.06.024`.

[14] Alkes L. Price, Gregory V. Kryukov, Paul I W. de Bakker, Shaun M. Purcell, Jeff Staples, Lee-Jen Wei, and Shamil R. Sunyaev. Pooled association tests for rare variants in exon-resequencing studies. *Am J Hum Genet*, 86(6):832–838, Jun 2010. doi: 10.1016/j.ajhg.2010.04.005. URL `http://dx.doi.org/10.1016/j.ajhg.2010.04.005`.

[15] Dajiang J Liu and Suzanne M Leal. Estimating genetic effects and quantifying missing heritability explained by identified rare-variant associations. *Am J Hum Genet*, 91(4):585–596, October 2012. ISSN 1537-6605 (Electronic); 0002-9297 (Linking). doi: 10.1016/j.ajhg.2012.08.008.

[16] Andrew J. Schork, Wesley K. Thompson, Phillip Pham, Ali Torkamani, J. Cooper Roddey, Patrick F. Sullivan, John R. Kelsoe, Michael C. O'Donovan, Helena Furberg, Nicholas J. Schork, Ole A. Andreassen, Anders M. Dale, The Tobacco, and The Schizophrenia Psychiatric Genomics Consortium Genetics Consortium, The Bipolar Disorder Psychiatric Genomics Consortium. All snps are not created equal: Genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated snps. *PLoS Genet*, 9(4):e1003449, April 2013. doi: 10.1371/journal.pgen.1003449.

[17] Matthew T. Maurano, Richard Humbert, Eric Rynes, Robert E. Thurman, Eric Haugen, Hao Wang, Alex P. Reynolds, Richard Sandstrom, Hongzhu Qu, Jennifer Brody, Anthony Shafer, Fidencio Neri, Kristen Lee, Tanya Kutyavin, Sandra Stehling-Sun, Audra K. Johnson, Theresa K. Canfield, Erika Giste, Morgan Diegel, Daniel Bates, R. Scott Hansen, Shane Neph, Peter J. Sabo, Shelly Heimfeld, Antony Raubitschek, Steven Ziegler, Chris Cotsapas, Nona Sotoodehnia, Ian Glass, Shamil R. Sunyaev, Rajinder Kaul, and John A. Stamatoyannopoulos. Systematic localization of common disease-associated variation in regulatory DNA. *Science*, 337(6099):1190–1195, 2012. doi: 10.1126/science.1222794. URL `http://www.sciencemag.org/content/337/6099/1190.abstract`.

[18] Hans D Daetwyler, Beatriz Villanueva, and John A Woolliams. Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One*, 3(10):e3395, 2008.

[19] Sang Hong Lee and Naomi R Wray. Novel genetic analysis for case-control genome-wide association studies: Quantification of power and genomic prediction accuracy. *PloS one*, 8(8):e71494, 2013.