Improved Detection of Common Variants Associated with Schizophrenia by Leveraging Pleiotropy with Cardiovascular-Disease Risk Factors

Ole A. Andreassen,[1,2,3,*] Srdjan Djurovic,[1,2] Wesley K. Thompson,[3] Andrew J. Schork,[4,5,6] Kenneth S. Kendler,[7] Michael C. O'Donovan,[8] Dan Rujescu,[9] Thomas Werge,[10] Martijn van de Bunt,[11] Andrew P. Morris,[11] Mark I. McCarthy,[11] International Consortium for Blood Pressure GWAS, Diabetes Genetics Replication and Meta-analysis Consortium, Psychiatric Genomics Consortium Schizophrenia Working Group, J. Cooper Roddey,[4,13] Linda K. McEvoy,[4,12] Rahul S. Desikan,[4,12] and Anders M. Dale[3,4,12,13,*]

Several lines of evidence suggest that genome-wide association studies (GWASs) have the potential to explain more of the "missing heritability" of common complex phenotypes. However, reliable methods for identifying a larger proportion of SNPs are currently lacking. Here, we present a genetic-pleiotropy-informed method for improving gene discovery with the use of GWAS summary-statistics data. We applied this methodology to identify additional loci associated with schizophrenia (SCZ), a highly heritable disorder with significant missing heritability. Epidemiological and clinical studies suggest comorbidity between SCZ and cardiovascular-disease (CVD) risk factors, including systolic blood pressure, triglycerides, low- and high-density lipoprotein, body mass index, waist-to-hip ratio, and type 2 diabetes. Using stratified quantile-quantile plots, we show enrichment of SNPs associated with SCZ as a function of the association with several CVD risk factors and a corresponding reduction in false discovery rate (FDR). We validate this "pleiotropic enrichment" by demonstrating increased replication rate across independent SCZ substudies. Applying the stratified FDR method, we identified 25 loci associated with SCZ at a conditional FDR level of 0.01. Of these, ten loci are associated with both SCZ and CVD risk factors, mainly triglycerides and low- and high-density lipoproteins but also waist-to-hip ratio, systolic blood pressure, and body mass index. Together, these findings suggest the feasibility of using genetic-pleiotropy-informed methods for improving gene discovery in SCZ and identifying potential mechanistic relationships with various CVD risk factors.

Introduction

Complex human traits and disorders are influenced by numerous genes that each have small individual effects,[1] and thousands of SNPs have been identified by genome-wide association studies (GWASs).[2,3] However, these SNPs fail to explain a substantial proportion of the heritability of the complex phenotypes studied;[4] this is often referred to as the "missing heritability." Recent results indicate that GWASs have the potential to explain a greater proportion of the heritability of common complex phenotypes,[5,6] and additional SNPs are likely to be identified in larger samples.[7] Because of the polygenic architecture of most complex traits and disorders, a large number of SNPs have associations too weak to be identified in the currently available sample sizes.[4] Cost-effective analytical methods are needed for reliably identifying a larger proportion of SNPs associated with complex diseases and phenotypes given that recruitment and genotyping of new participants is expensive. Here, we present a genetic-pleiotropy-informed approach for GWASs to capture more of the polygenic effects in complex disorders and traits. Given the high number of traits in humans and the relative small number of genes (~20,000), some genes have to affect multiple traits (genetic pleiotropy).[8] By combining independent GWASs from associated traits or comorbid disorders, we hypothesize that a genetic-pleiotropy-informed approach can significantly improve discovery of genes and help capture a greater proportion of the missing heritability.

Reports indicate overlapping SNPs between several human traits[9,10] and disorders.[11,12] To date, methods for assessing genetic pleiotropy have not taken full advantage of the existing GWAS data, and the majority of these studies have focused on the subset of SNPs exceeding a Bonferroni-corrected threshold of significance for each trait or disorder.[10,13] However, this approach cannot detect SNPs that reach genome-wide significance in the

[1]K.G. Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo 0407, Norway; [2]Division of Mental Health and Addiction, Oslo University Hospital, Oslo 0407, Norway; [3]Department of Psychiatry, University of California, San Diego, La Jolla, CA 92037, USA; [4]Multimodal Imaging Laboratory, University of California, San Diego, La Jolla, CA 92037, USA; [5]Cognitive Sciences Graduate Program, University of California, San Diego, La Jolla, CA 92037, USA; [6]Center for Human Development, University of California, San Diego, La Jolla, CA 92037, USA; [7]Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond, VA 23298, USA; [8]Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK; [9]Department of Psychiatry, University of Halle-Wittenberg, Halle 06112, Germany; [10]Institute of Biological Psychiatry, Mental Health Center Sct. Hans, University of Copenhagen, Copenhagen, Roskilde DK 4000, Denmark; [11]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7LJ, UK; [12]Department of Radiology, University of California, San Diego, La Jolla, CA 92037, USA; [13]Department of Neurosciences, University of California, San Diego, La Jolla, CA 92037, USA
*Correspondence: o.a.andreassen@medisin.uio.no (O.A.A.), amdale@ucsd.edu (A.M.D.)

combined analysis (hereafter referred to as polygenic pleiotropy) but do not meet Bonferroni-corrected significance in the individual phenotype. Combining GWASs from two traits or disorders also provides increased power to discover genes associated with common biological mechanisms and to potentially inform shared pathophysiological relationships between the phenotypes. In the current study, we use schizophrenia (SCZ [MIM 181500]) as an example of how a pleiotropy-informed analytical approach can improve gene discovery in a disorder with high heritability[14] and for which, despite recent discoveries,[13,15,16] most of the underlying genetic architecture remains unknown.[13]

SCZ, a debilitating mental health disorder, is among the leading global causes of disability[17] and constitutes a substantial portion of disease burden worldwide. Systematic reviews and meta-analyses indicate that individuals with SCZ have significantly higher mortality rates than the general population, and this corresponds to a 10–20 year reduction in life expectancy.[18–20] Although the mortality rate from suicide is high, lifestyle and cardiovascular-disease (CVD) risk factors contribute substantially to life-expectancy reduction in SCZ.[19–21] Epidemiological research has shown increased rates of dyslipidemia, type 2 diabetes (T2D [MIM 125853]), and obesity (MIM 601665) and a high prevalence of metabolic syndrome among people with SCZ.[22] This increase in CVD risk factors has been primarily attributed to lifestyle factors such as unhealthy diet, sedentary habits, excessive smoking, and the side effects of antipsychotic medication.[19,23–25] However, as suggested by studies predating the introduction of antipsychotics,[26] studies of untreated first-episode individuals and their healthy relatives,[27] and the identification of overlapping candidate genes,[28] shared genetics between SCZ and CVD risk factors might also be of importance.

Large GWASs have reported SNPs associated with a number of CVD risk factors, including systolic blood pressure (SBP), diastolic blood pressure (DBP),[29] low-density lipoprotein (LDL) cholesterol,[30] high-density lipoprotein (HDL) cholesterol,[30] triglycerides (TGs),[30] T2D,[31] body mass index (BMI),[32] and waist-to-hip ratio (WHR).[33] In the current study, we employed model-free strategies and leveraged the power of multiple large independent GWASs to identify SNPs exhibiting pleiotropy between SCZ and eight CVD risk factors by using summary statistics from six studies. After applying genomic inflation control, we computed the stratified empirical cumulative distribution functions (cdfs) of the nominal p values. Strata were determined by the relative enrichment of pleiotropic SNPs in SCZ as a function of increased nominal p values in the different CVD risk factors. For each nominal p value, an estimate of the stratum-specific true discovery rate (TDR = 1 – false discovery rate [FDR]) was obtained from the empirical cdfs.[34,35] We demonstrate that the stratified analysis improves power to detect SNPs by computing replication rates for nominal p value thresholds by using independent substudies for discovery and replication samples. We show that for a given replication rate, nominal p value thresholds are approximately 100 times larger for the most pleiotropic SNPs in SCZ than for all SNPs (hereafter referred to as genetic enrichment). Using this stratified methodology, we constructed a two-dimensional (2D) FDR "look-up" table in which the FDR in SCZ SNPs was computed conditionally on nominal CVD-risk-factor p values (this is referred to as *conditional* FDR). Using this table, we identified 25 loci that are significantly associated with SCZ at a conditional FDR level of 0.01. Finally, we constructed the *conjunction* FDR to investigate SNPs significantly associated with both SCZ and CVD risk factors. Specifically, we computed the conditional FDR for SCZ given CVD-risk-factor nominal p values, as well as conditional FDR for CVD risk factors given SCZ nominal p values, and we took the maximum of both values as the conjunction FDR. With this approach, we identified ten pleiotropic loci implicating overlapping genetic mechanisms between SCZ and blood lipids.

Material and Methods

Participant Samples

We obtained complete GWAS results in the form of summary-statistics p values from public-access websites or through collaboration with investigators (T2D cases and controls were from the Diabetes Genetics Replication and Meta-analysis [DIAGRAM] Consortium, and SCZ cases and controls were from the Psychiatric GWAS Consortium [PGC] [Table S1, available online]). There was no overlap between participants in the CVD GWAS and the SCZ case-control sample (n = 21,856), except for 2,974 of 12,462 (24%) controls.[13]

The SCZ GWAS summary-statistics results were obtained from the PGC,[13] which consists of 9,394 cases with SCZ or schizoaffective disorder and 12,462 controls (52% screened) from a total of 17 samples from 11 countries. The quality of phenotypic data was verified by a systematic review of data-collection methods and procedures at each site, and only studies that fulfilled these criteria were included. This involved the following nine key items: (1) the use of a structured psychiatric interview, (2) systematic training of interviewers in the use of the instrument, (3) systematic quality control of diagnostic accuracy, (4) reliability trials, (5) review of medical-record information, (6) best-estimate procedure employed, (7) specific inclusion and exclusion criteria developed and utilized, (8) final diagnostic determination made by MDs or PhDs, and (9) special additional training for the final SCZ PGC sample. One sample from Sweden used another approach, but further empirical support for the validity of this approach was provided. Controls consisted of 12,462 European-ancestry samples collected from the same countries. Because the prevalence of SCZ is low, a large control sample in which some controls were not screened for SCZ was utilized. For further details on sample characteristics and quality-control procedures applied, please see Ripke et al.[13] A total of 2,974 controls in the SCZ UK case-control sample[16] from the Wellcome Trust Case Control Consortium (WTCCC) were also included in several of the CVD-risk-factor GWASs. This constitutes 24% of the total number of controls (n = 12,462) in the SCZ PGC sample.[13]

More information about inclusion criteria and phenotype characteristics of the CVD-risk-factor samples of the different GWASs is included in the original publications.[29–33] The relevant institutional review boards or ethics committees approved the research protocol of the individual GWASs used in the current analysis, and all human participants gave written informed consent.

Statistical Analyses

Stratified Quantile-Quantile Plots

Quantile-quantile (Q-Q) plots compare a nominal probability distribution against an empirical distribution. In the presence of all null relationships, nominal p values form a straight line on a Q-Q plot when they are plotted against the empirical distribution. For each phenotype, for all SNPs, and for each categorical subset (strata), $-\log_{10}$ nominal p values were plotted against $-\log_{10}$ empirical p values (stratified Q-Q plots). Leftward deflections of the observed distribution from the projected null line reflect increased tail probabilities in the distribution of test statistics (Z scores) and, consequently, an overabundance (also termed "enrichment") of low p values compared to that expected by chance.

Under large-scale testing paradigms, such as GWASs, quantitative estimates of probably true associations can be estimated from the distributions of summary statistics.[36,37] A common method for visualizing the enrichment of statistical association relative to that expected under the global null hypothesis is through Q-Q plots of nominal p values obtained from GWAS summary statistics. The usual Q-Q curve has the nominal p value, denoted by "p," as the y ordinate and the corresponding value of the empirical cdf, denoted by "q," as the x ordinate. Under the global null hypothesis, the theoretical distribution is uniform on the interval [0, 1]. As is common in GWASs, we instead plot $-\log_{10}(p)$ against $-\log_{10}(q)$ to emphasize tail probabilities of the theoretical and empirical distributions. Therefore, genetic enrichment results in a leftward shift in the Q-Q curve, corresponding to a larger fraction of SNPs with a nominal $-\log_{10}$ p value greater than or equal to a given threshold. *Stratified* Q-Q plots are constructed by the creation of subsets of SNPs on the basis of levels of an auxiliary measure for each SNP and the computation of Q-Q plots separately for each level. If SNP enrichment is captured by variation in the auxiliary measure, this is expressed as successive leftward deflections in a stratified Q-Q plot as levels of the auxiliary measure increase.

Genomic Control

The empirical null distribution in GWASs is affected by global variance inflation due to population stratification and cryptic relatedness[38] and deflation due to overcorrection of test statistics for polygenic traits by standard genomic-control methods,[39] in addition to incorrect asymptotic approximation used for computing the p values. We applied a control method leveraging only intergenic SNPs, which are most likely depleted for true associations (unpublished data). First, we annotated the SNPs to genic (5′ UTR, exon, intron, and 3′ UTR) and intergenic regions by using information from the 1000 Genomes Project (1KGP). As illustrated in Figure S1, there are more functional genic regions than intergenic regions in SCZ. We used intergenic SNPs because their relative depletion of associations suggests that they provide a robust estimate of true null effects, and they thus seem to be a better category for genomic control than all SNPs. We converted all p values to Z scores, and for each phenotype, we estimated the genomic inflation factor λ_{GC} for intergenic SNPs. We computed the infla-

tion factor λ_{GC} as the median Z score squared divided by the expected median of a chi-square distribution with one degree of freedom, and we adjusted all test statistics by λ_{GC}. The stratified Q-Q plot for SCZ after control for genomic inflation is shown in Figure S1.

Stratified Q-Q Plots for Pleiotropic Enrichment

To assess pleiotropic enrichment, we used a Q-Q plot stratified by "pleiotropic" effects. For a given associated phenotype, enrichment for pleiotropic signals is present if the degree of deflection from the expected null line is dependent on SNP associations with the second phenotype. We constructed stratified Q-Q plots of empirical quantiles of nominal $-\log_{10}(p)$ values for SNP association with SCZ for all SNPs, as well as for subsets (strata) of SNPs determined by the nominal p values of their association with a given CVD risk factor. Specifically, we computed the empirical cumulative distribution of nominal p values for a given phenotype for all SNPs and for SNPs with significance levels below the indicated cutoffs for the other phenotype ($-\log_{10}(p) \geq 0$, $-\log_{10}(p) \geq 1$, $-\log_{10}(p) \geq 2$, and $-\log_{10}(p) \geq 3$ corresponding to p < 1, p < 0.1, p < 0.01, and p < 0.001, respectively). The nominal p values ($-\log_{10}(p)$) are plotted on the y axis, and the empirical quantiles ($-\log_{10}(q)$, where q = cdf(p)) are plotted on the x axis. To assess polygenic effects below the standard GWAS significance threshold, we focused the stratified Q-Q plots on SNPs with nominal $-\log_{10}(p) < 7.3$ (corresponding to p > 5 × 10^{-8}).

Significance of Enrichment

Using Q-Q plots (with 95% confidence intervals [CIs]) of empirical versus nominal $-\log_{10}(p)$ values in SCZ as a function of the significance of association with the CVD risk factors, we estimated the significance of the polygenic enrichment (Figures S2A–S2C). After using intergenic SNPs to estimate and control for genomic inflation (Figure S1), we pruned the SNPs by removing SNPs in linkage disequilibrium (LD) ($r^2 \geq 0.2$) and computed 95% CIs for the Q-Q plots. From these CIs, we calculated standard errors and used two sample t tests to estimate the difference (degree of departure) between the empirical distribution of SCZ (phenotype 1) SNPs that were above a given association threshold ($-\log_{10}(p) > 1$, $-\log_{10}(p) > 2$, $-\log_{10}(p) > 3$, and $-\log_{10}(p) > 4$; red lines) and the distribution of SNPs with $-\log_{10}(p) \leq 1$ for the CVD-risk-factor (phenotype 2) category (blue line). The p values listed in Table S5 indicate the most significant difference, as assessed by a two-sample t test, between the red ($-\log_{10}(p) > 1$, 2, 3, or 4) and blue ($-\log_{10}(p) \leq 1$) lines. This is reflected in the largest difference between the 95% CIs. The 95% CIs also illustrate the region containing significant differences in the Q-Q plot. For differences between the distributions, we only report p values appearing above the $-\log_{10}(p) > 2$ threshold on the Q-Q plots. This clearly shows significant enrichment conditioning SCZ on TG and WHR.

For the CVD risk factors with significant enrichment, we further calculated the significance of the enrichment for SNPs with a Fisher's combined p value below the genome-wide significance level of 5 × 10^{-8}, and we hereafter refer to this as "censoring." This made it possible to examine whether the enrichment was entirely explained by the most significant pleiotropic SNPs or whether it was due to a more general, polygenic effect (polygenic pleiotropy). As illustrated for SCZ and TG in Figure S2C, the polygenic pleiotropic enrichment is highly significant. See also Table S5.

Stratified TDR

Enrichment seen in the stratified Q-Q plots can be directly interpreted in terms of TDR (equivalent to 1 – FDR[40]). We applied the stratified FDR method,[35] previously used for GWAS

enrichment based on linkage information.[34] Specifically, for a given p value cutoff, the FDR is defined as

$$FDR(p) = \frac{\pi_0 F_0(p)}{F(p)}, \qquad \text{(Equation 1)}$$

where π_0 is the proportion of null SNPs, F_0 is the null cdf, and F is the cdf of all SNPs, both null and non-null; see below for details on this simple mixture-model formulation.[41] Under the null hypothesis, F_0 is the cdf of the uniform distribution on the unit interval [0, 1], so Equation 1 reduces to

$$FDR(p) = \frac{\pi_0 p}{F(p)}. \qquad \text{(Equation 2)}$$

The cdf F can be estimated by the empirical cdf $q = N_p / N$, where N_p is the number of SNPs with a p value less than or equal to p and N is the total number of SNPs. By replacing F with q in Equation 2, we get

$$\text{Estimated } FDR(p) = \frac{\pi_0 p}{q}, \qquad \text{(Equation 3)}$$

which is biased upward as an estimate of the FDR.[41] Replacing π_0 in Equation 3 with unity gives an estimated FDR that is further biased upward:

$$q^* = \frac{p}{q}. \qquad \text{(Equation 4)}$$

If π_0 is close to 1, as is probably true for most GWASs, the increase in bias from Equation 3 is minimal. The quantity $1 - p / q$ is therefore biased downward and hence is a conservative estimate of the TDR.

Referring to the formulation of the Q-Q plots, we see that q^* is equivalent to the nominal p value divided by the empirical quantile, as defined earlier. Given the $-\log_{10}$ of the Q-Q plots, we can easily obtain

$$-\log_{10}(q^*) = \log_{10}(q) - \log_{10}(p), \qquad \text{(Equation 5)}$$

demonstrating that the (conservatively) estimated FDR is directly related to the horizontal shift of the curves in the stratified Q-Q plots from the expected line $x = y$, i.e., a larger shift corresponds to a smaller FDR, as illustrated in Figure 1. As before, the estimated TDR can be obtained as $1 - FDR$. For each range of p values (stratum) in a pleiotropic trait, we calculated the TDR as a function of p values in SCZ (indicated by different colored curves) in Figure 1 by using each observed p value as a threshold according to Equation 5.

Estimates of Pleiotropy

Let z be the GWAS test statistic for a corresponding p value. The two-group mixture model for Z scores implicit in Equation 1 is given by

$$f(z) = \pi_0 f_0(z) + (1 - \pi_0) f_1(z), \qquad \text{(Equation 6)}$$

where f_0 is the null distribution (standard normal after appropriate genomic control), f_1 is the non-null distribution (which can be estimated parametrically or nonparametrically),[36] and π_0 is the proportion null, as before. We can easily generalize this model to two Z scores from phenotypes simultaneously (z_1 for phenotype 1 and z_2 for phenotype 2) by using a bivariate density from the four-group mixture model,

$$f(z_1, z_2) = \pi_0 f_0(z_1, z_2) + \pi_1 f_1(z_1, z_2) + \pi_2 f_2(z_1, z_2) + \pi_3 f_3(z_1, z_2), \qquad \text{(Equation 7)}$$

where π_0 is the proportion of SNPs for which both phenotypes are null, π_1 is the proportion of SNPs for which phenotype 1 is non-null and phenotype 2 is null, π_2 is the proportion of SNPs for which phenotype 1 is null and phenotype 2 is non-null, and π_3 is the proportion of SNPs for which both phenotypes are non-null (i.e., the pleiotropic SNPs). The mixture densities in Equation 7 are given by

$$\begin{aligned}
f_0(z_1, z_2) &= \varphi(z_1)\,\varphi(z_2) \\
f_1(z_1, z_2) &= g_1(z_1)\,\varphi(z_2) \\
f_2(z_1, z_2) &= \varphi(z_1)\,g_2(z_2) \\
f_3(z_1, z_2) &= g_1(z_1)\,g_2(z_2),
\end{aligned} \qquad \text{(Equation 8)}$$

where $\phi()$ denotes the standard normal density and g_1 and g_2 denote the non-null marginal densities of z_1 and z_2, respectively. We found that modeling the marginal non-null densities with normal Laplace densities or the Weibull distribution on the squared Z scores (z^2) (in which case the null densities are central chi-square with one degree of freedom) fits the data well. The proportions $\pi = (\pi_0, \pi_1, \pi_2, \pi_3)$ and the parameters of the non-null distributions can be estimated with maximum likelihood or Bayesian methods such as Markov Chain Monte Carlo. From the probability density function (pdf) (Equation 7), we can compute the joint and conditional cdfs, and hence the FDR (Equation 2), of one phenotype conditionally on tail probabilities of the second.

Figure S3 presents the observed and fitted Q-Q curves for SCZ, TG, and WHR; these curves are based on the marginal cdfs from the bivariate-mixture-model fits, indicating very good fit. The non-null distributions were modeled parametrically with Weibull distributions for non-null z^2 and with chi-square with one degree of freedom for null z^2. The estimated vector of probabilities π from these fits can also be used for testing whether the degree of pleiotropy is significantly higher than that expected by chance if both phenotypes are independent. Independence implies that the joint pdf of both phenotype Z scores is a product of two two-group mixtures (Equation 6). It is easy to show that demonstrating excess pleiotropy from that predicted by independence is equivalent to showing that $\pi_3 > \pi_1 \pi_2 / \pi_0$ in Equation 7 or that the log odds ratio (LOR)

$$\begin{aligned}
LOR(\text{phenotype 1, phenotype 2}) &= \log\left\{\frac{\pi 3}{1 - \pi 3}\right\} \\
&- \log\left\{\frac{(\pi_1 \pi_2 / \pi_0)}{(1 - \pi_1 \pi_2 / \pi_0)}\right\}
\end{aligned} \qquad \text{(Equation 9)}$$

is greater than zero. With a multivariate normal approximation to the maximum-likelihood estimates with covariance obtained from the inverse Fisher information matrix, estimates of LOR with 95% CIs are LOR(SCZ,TG) = 4.0 [3.8, 4.3] and LOR(SCZ,WHR) = 2.4 [2.1, 2.7], which are both highly significantly different from zero. These 95% CIs include an adjustment that assumes an effective degree of freedom of 500,000 independent SNPs to account for the correlation of SNPs (i.e., LD).

Stratified Replication Rate

For each of the 17 substudies contributing to the final meta-analysis in SCZ, we independently adjusted Z scores by using intergenic inflation control. For 1,000 of the possible combinations of the eight-study discovery sets and nine-study replication sets, we calculated the eight-study combined discovery Z score and eight- or nine-study combined replication Z score for each SNP as the average Z score across the eight or nine studies and multiplied these by the square root of the number of studies.
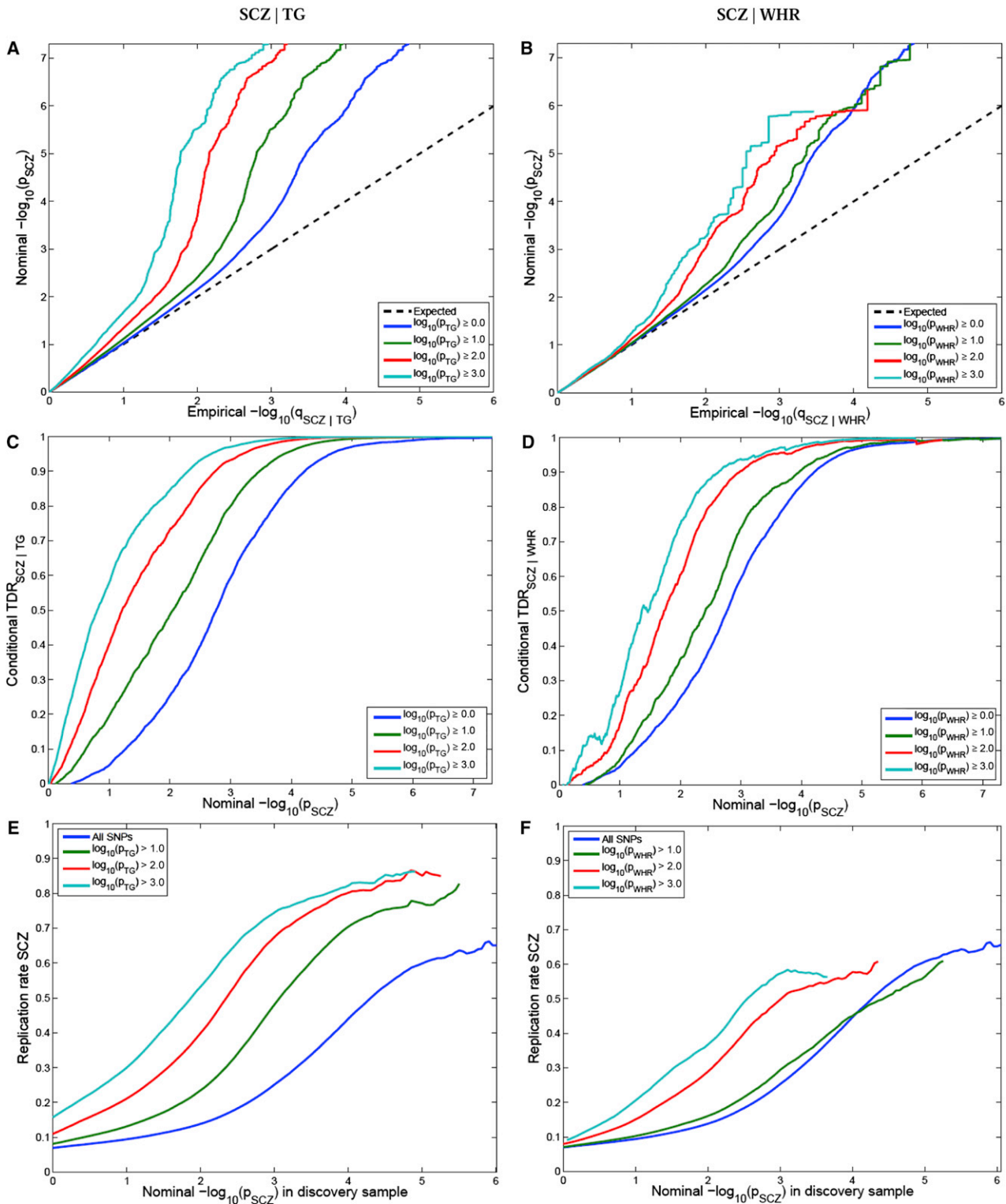
Figure 1. Enrichment and Replication

(A and B) Stratified Q-Q plot of nominal versus empirical $-\log_{10}$ p values (corrected for inflation) in SCZ below the standard GWAS threshold of $p < 5 \times 10^{-8}$ as a function of significance of association with (A) TGs and (B) WHR at the levels of $-\log_{10}(p) > 0$, $-\log_{10}(p) > 1$, $-\log_{10}(p) > 2$, and $-\log_{10}(p) > 3$, which correspond to $p < 1$, $p < 0.1$, $p < 0.01$, and $p < 0.001$, respectively. Dashed lines indicate the null hypothesis.

(C and D) Stratified TDR plots illustrating the TDR increase associated with increased pleiotropic enrichment in (C) SCZ conditioned on TG (SCZ|TG) and (D) SCZ conditioned on WHR (SCZ|WHR).

(E and F) Cumulative replication plot showing the average rate of replication ($p < 0.05$) within SCZ substudies for a given p value threshold demonstrates that pleiotropic enriched SNP categories replicate at a higher rate in independent SCZ samples for (E) SCZ conditioned on TG (SCZ|TG) and (F) SCZ conditioned on WHR (SCZ|WHR). The vertical intercept is the overall replication rate per category.

For discovery samples, the Z scores were converted to two-tailed p values, whereas replication samples were converted to one-tailed p values, preserving the direction of effect in the discovery sample. For each of the 1,000 discovery-replication pairs, cumulative rates of replication were calculated over 1,000 equally spaced bins spanning the range of negative $\log_{10}$(p values) observed in the discovery samples. The cumulative replication rate for any bin was calculated as the proportion of SNPs with a $-\log_{10}$(discovery p value) greater than the lower bound of the bin with a replication p value < 0.05. Cumulative replication rates were calculated independently for each of the four pleiotropic enrichment categories, as well as for intergenic SNPs and all SNPs. For each category, the cumulative replication rate for each bin was averaged across the 1,000 discovery-replication pairs, and the results are reported in Figure 1. The vertical intercept is the overall replication rate.

Stratified Replication Effect Sizes
Stratified TDR is directly related to stratified replication effect sizes and hence replication rates.

As before, for each of the 17 substudies contributing to the final meta-analysis in SCZ, we independently adjusted Z scores by using intergenic inflation control. For 1,000 of the possible combinations of the eight-study discovery sets and nine-study replication sets, we calculated the eight-study combined discovery Z score and eight- or nine-study combined replication Z score for each SNP. The effect sizes were stratified by levels of $\log_{10}$(p values) from the TG GWAS. As illustrated in Figure S4, we also calculated the cumulative replication rate without overlapping controls (we removed the UK sample that included the WTCCC controls).

For visualization, a cubic smoothing spline was fit for relating the discovery Z score bin midpoints to the corresponding average replication Z scores (see Figure S5). The nonlinear pattern of shrinkage is typical of that observed in mixture models, as in Equation 1. Importantly, the amount of shrinkage is highly dependent on enrichment stratum: replication effect sizes in more enriched strata exhibit more fidelity with discovery sample effect sizes. This directly relates to increased TDR and translates into increased replication rates for enriched strata.

Conditional Statistics—Test of Association with SCZ
To improve detection of SNPs associated with SCZ, we used a stratified FDR approach in which we leveraged associated phenotypes by using established stratified FDR methods.[34,35] Specifically, we stratified SNPs on the basis of p values in the pleiotropic phenotype (e.g., TGs). On the basis of the combination of p values for the SNP in SCZ and the pleiotropic trait, we then assigned a conditional FDR value (denoted as $FDR_{SCZ \mid TG}$) for SCZ to each SNP by interpolating into a 2D look-up table (Figure S6). All SNPs with FDR < 0.01 ($-\log_{10}$(FDR) > 2) in SCZ given the different CVD risk factors are listed in Table 1 after "pruning" (removing all SNPs with $r^2 > 0.2$ according to 1KGP LD structure). A significance threshold of FDR < 0.01 corresponds to 1 false positive per 100 reported associations. We also list all SNPs with FDR < 0.05 ($-\log_{10}$(FDR) > 1.3) in Table S2.

Conditional Manhattan Plots
To illustrate the localization of the genetic markers associated with SCZ given the CVD-risk-factor effect, we created a "conditional Manhattan plot" by plotting all SNPs within an LD block in relation to their chromosomal location. As illustrated in Figure 2, the large points represent the SNPs with FDR < 0.05, whereas the small points represent the nonsignificant SNPs. All SNPs without pruning are shown. The strongest signal in each LD block

is illustrated with a black line around the circles. We identified these signals by ranking all SNPs in increasing order on the basis of the conditional FDR value for SCZ and then by removing SNPs in LD $r^2 > 0.2$ with any higher ranked SNP. Thus, the selected locus was the most significantly associated with SCZ in each LD block (Figure 2).

Conjunction Statistics—Test of Association with Both Phenotypes
In order to identify which of the SNPs associated with SCZ given the CVD risk factor (SCZ|CVD, Table 1) were also associated with CVD risk factors given SCZ (opposite direction), we calculated the conditional FDR in the other direction (CVD|SCZ). This is reported in Table 2. The corresponding Z scores are listed in Table S3. The Z scores were calculated from the p values, and the direction of effect was determined by the risk allele.

In addition, to make a comprehensive, unselected map of pleiotropic signals, we used a conjunction testing procedure, as outlined for p value statistics in Nichols et al.,[42] and adapted this method for FDR statistics on the basis of the conditional-FDR approach.[34,35] On the basis of the combination of p values for the SNP in SCZ and the pleiotropic trait, we defined the conjunction statistics (denoted as $FDR_{SCZ \& TG}$) as the maximum conditional FDR in both directions, i.e.,

$$FDR_{SCZ \& TG} = \max(FDR_{SCZ \mid TG}, FDR_{TG \mid SCZ}),$$

by interpolating into a bidirectional 2D look-up table (Figure S7). The conjunction statistic allows for identification of SNPs that are associated with both phenotypes, which minimizes the effect of a single phenotype driving the common association signal. All SNPs with conjunction FDR < 0.05 ($-\log_{10}$(FDR) > 1.3) with SCZ and any of the CVD risk factors considered are listed in Table S4 (after pruning).

Conjunction Manhattan Plots
To illustrate the localization of the pleiotropic genetic markers in association with both SCZ and CVD risk factors, we used a conjunction Manhattan plot, for which we plotted all SNPs with a significant conjunction FDR within an LD block in relation to their chromosomal location. As illustrated in Figure S8, the large points represent the significant SNPs (FDR < 0.05), whereas the small points represent the nonsignificant SNPs. All SNPs without pruning are shown, and the strongest signal in each LD block is illustrated with a black line around the circles. First, we ranked all SNPs on the basis of the conjunction FDR and removed SNPs in LD $r^2 > 0.2$ with any higher ranked SNP (Figure S8).

Results

Q-Q Plots of SCZ SNPs Stratified by Association with Pleiotropic CVD Risk Factors
Stratified Q-Q plots for SCZ conditioned on nominal p values of association with TGs showed enrichment across different levels of significance for TGs (Figure 1A). The earlier departure from the null line (leftward shift) suggests a greater proportion of true associations for a given nominal SCZ p value. Successive leftward shifts for decreasing nominal TG p values indicate that the proportion of non-null effects varies considerably across different levels of association with CVD risk factors. For example, in the $-\log_{10}(p_{TG}) \geq 3$ category, the proportion

Table 1. Conditional FDR: SCZ Loci Given CVD Risk Factors

Locus	SNP	Gene Region	MIM	Chr	SCZ p Value	SCZ FDR	Minimum Conditional FDR	CVD Risk Factor
4	rs1625579	*AK094607*[a]	614304	1p21.3	5.52×10^{-6}	0.02105	0.00420	TG
9	rs2272417	*IFT172*	607386	2p23.3	4.47×10^{-5}	0.07516	0.00193	TG
17	rs17180327	*CWC22*	-	2q31.3	6.37×10^{-6}	0.02332	0.00780	HDL
20	rs13025591	*AGAP1*	608651	2q37	9.26×10^{-6}	0.02953	0.00131	TG
22	rs2239547	*ITIH4*[a]	600564	3p21.1	1.73×10^{-5}	0.03920	0.00400	HDL
23	rs11715438	*PTPRG*	176886	3p21-p14	2.47×10^{-6}	0.01601	0.00222	HDL
26	rs9838229	*DKFZp434A128*	-	3q27.2	1.11×10^{-5}	0.02953	0.00825	HDL
37[b]	rs2021722	*TRIM26*[a]	600830	6p21.3	**2.08×10^{-9}**	**0.00046**	0.00001	TG
38	rs7383287	*HLA-DOB*	600629	6p21.3	3.44×10^{-5}	0.06382	0.00748	HDL
39	rs1480380	*HLA-DMA*	142855	6p21.3	3.05×10^{-6}	0.01746	0.00028	TG
40	rs9462875	*CUL9*	607489	6p21.1	1.20×10^{-5}	0.03383	0.00739	WHR
42	rs1107592	*MAD1L1*	602686	7p22	7.63×10^{-7}	**0.00919**	0.00493	HDL
48	rs10503253	*CSMD1*[a]	608397	8p23.2	3.96×10^{-6}	0.01912	0.00432	TG
51	rs12234997	*AK055863*	-	8p23.1	2.23×10^{-5}	0.04590	0.00347	TG
55	rs755223	*BC037345*	-	8q12.3	6.91×10^{-5}	0.10338	0.00895	HDL
56	rs7004633	*MMP16*[a]	602262	8q21.3	2.60×10^{-7}	**0.00504**	0.00141	HDL
65	rs11191580	*NT5C2*[a]	600417	10q24.32	3.73×10^{-7}	**0.00625**	0.00013	SBP
	rs7914558	*CNNM2*[a]	607803	10q24.32	1.90×10^{-6}	0.01464	0.00101	HDL
	rs2296569	*CNNM2*	607803	10q24.32	3.78×10^{-6}	0.01912	0.00127	TG
	rs10748835	*AS3MT*	611806	10q24.32	2.21×10^{-6}	0.01464	0.00274	HDL
67	rs11191732	*NEURL*	603804	10q25.1	2.55×10^{-6}	0.01601	0.00160	HDL
71	rs2172225	*METT5D1*	-	11p14.1	4.88×10^{-5}	0.08828	0.00238	TG
	rs7938219	*CR618717*	-	11p14.1	3.75×10^{-5}	0.07516	0.00331	TG
78	rs548181	*STT3A*	601134	11q23.3	4.65×10^{-7}	**0.00707**	0.00044	WHR
	rs11220082	*FEZ1*	604825	11q24.2	2.84×10^{-6}	0.01746	0.00279	TG
	rs671789	*PKNOX2*	613066	11q24.2	1.46×10^{-5}	0.03920	0.00695	WHR
80	rs7972947	*CACNA1C*[a]	114205	12p13.2	7.12×10^{-6}	0.02609	0.00415	TG
81	rs4765905	*CACNA1C*[a]	114205	12p13.3	7.99×10^{-6}	0.02609	0.00758	TG
84	rs8003074	*KIAA0391*	609947	14q13.2	7.23×10^{-6}	0.02609	0.00484	HDL
	rs10135277	*KIAA0391*	609947	14q13.1	5.02×10^{-6}	0.02105	0.00491	TG
87	rs1869901	*PLCB2*	604114	15q15	3.66×10^{-6}	0.01912	0.00203	TG
101	rs17597926	*TCF4*[a]	602272	18q21.1	6.49×10^{-7}	**0.00805**	0.00216	TG

Independent complex or single-gene loci ($r^2 < 0.2$) with SNP(s) with a conditional FDR < 0.01 in SCZ given the association in CVD risk factors. We defined the most significant SCZ SNP in each LD block on the basis of the minimum conditional FDR for each phenotype. Listed are the most significant SNPs in each gene of the LD block, as well as the CVD risk factor that provided the signal. All loci with SNPs with conditional FDR < 0.05 were used for defining the number of the loci (Table S2). This and the respective FDR values in each phenotype are listed in Table S2. SCZ FDR values < 0.01 are in bold. The following abbreviations are used: chr, chromosomal region; SCZ, schizophrenia; FDR, false-discovery rate; CVD, cardiovascular disease; TG, triglyceride; HDL, high-density lipoprotein; WHR, waist-to-hip ratio; and SBP, systolic blood pressure.
[a]Same locus identified in previous SCZ GWASs. All data were first corrected for genomic inflation.
[b]There are additional independent SNPs in the HLA region on chromosome 6 (locus 37). The complete SNP list is shown in Table S6.

of SNPs reaching a given significance level (e.g., $-\log_{10}(p_{SCZ}) > 6$) is roughly 100 times greater than that for the $-\log_{10}(p_{TG}) \geq 0$ category (all SNPs), indicating a very high level of enrichment. Similarly, a clear pleio-tropic enrichment was also seen for HDL and LDL cholesterol. A less clear pleiotropic enrichment was seen for WHR (Figure 1B), BMI, and SBP, but there was no evidence for enrichment in T2D (data not shown).
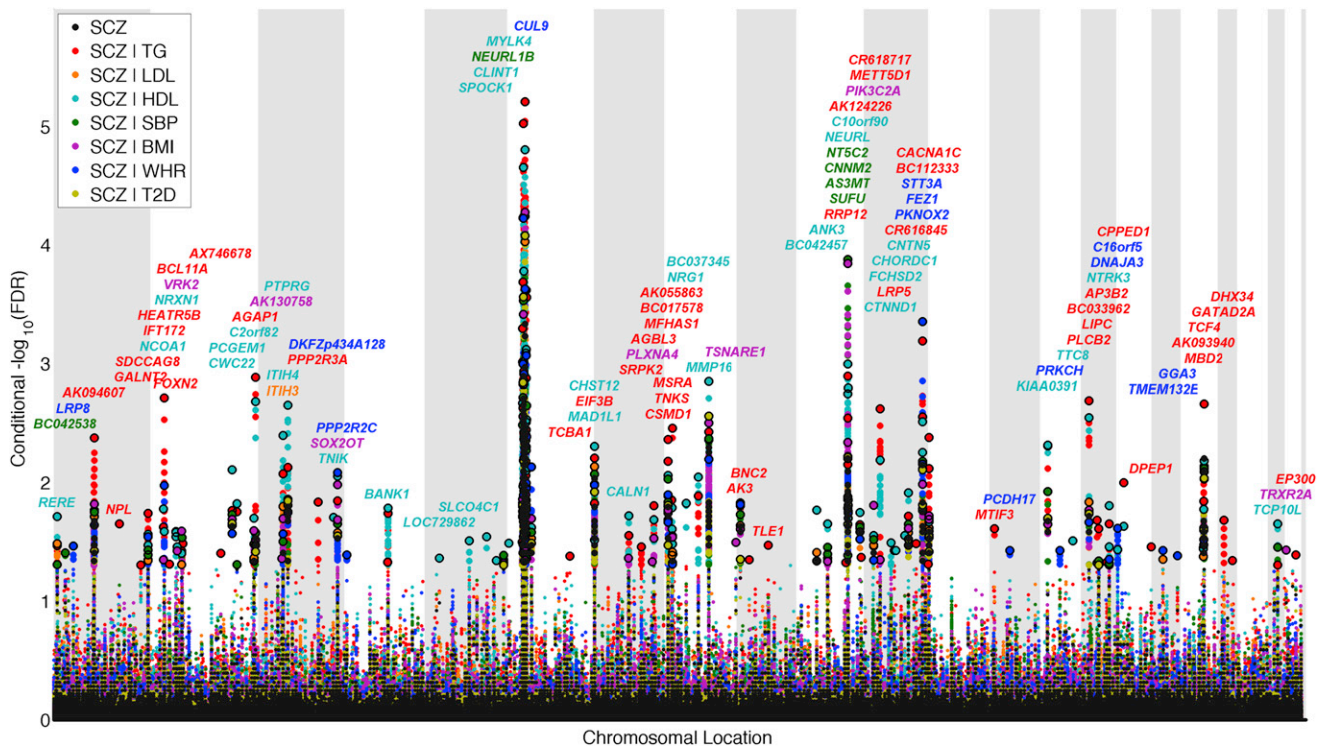
Figure 2. Conditional Manhattan Plot

Conditional Manhattan plot of conditional $-\log_{10}$ (FDR) values for SCZ alone (black) and SCZ given the following CVD risk factors: TGs (SCZ|TG, red), LDL (SCZ|LDL, orange), HDL (SCZ|HDL, cyan), SBP (SCZ|SBP, green), BMI (SCZ|BMI, purple), WHR (SCZ|WHR, blue), and T2D (SCZ|T2D, chartreuse). SNPs with conditional $-\log_{10}$ FDR > 1.3 (i.e., FDR < 0.05) are shown with large points. A black line around the large points indicates the most significant SNP in each LD block, and this SNP was annotated with the closest gene, which is listed above the symbols in each locus (except for the HLA region on chromosome 6) and in Table S2. The figure shows the localization of 106 loci on a total of 21 chromosomes (1–19, 21, and 22). Details for the loci with $-\log_{10}$ FDR > 2 (i.e., FDR < 0.01) are shown in Table 1.

Conditional TDR in SCZ Is Increased by CVD Risk Factors

Because categories of SNPs with stronger pleiotropic enrichment are more likely to be associated with SCZ, all tag SNPs should not be treated exchangeably so that power for discovery can be maximized. Specifically, variation in enrichment across pleiotropic categories is expected to be associated with corresponding variation in the TDR (equivalent to 1 – FDR)[40] for association of SNPs with SCZ. A conservative estimate of the TDR for each nominal p value is equivalent to 1 – (p / q), easily obtained from the stratified Q-Q plots. This relationship is shown for SCZ conditioned on TG (Figure 1C) and WHR (Figure 1D). For a given conditional TDR, the corresponding estimated nominal p value threshold varies by a factor of 100 from the most to the least enriched SNP category (strata) for SCZ conditioned on TG (SCZ|TG) and approximately by a factor of 40 for SCZ conditioned on WHR (SCZ|WHR). Phenotypes with weaker pleiotropy with SCZ showed smaller increases in conditional TDR (data not shown). Because TDR is strongly related to predicted replication rate, it is expected that the replication rate will increase for a given nominal p value for SNPs in categories with higher conditional TDR.

Replication Rate in SCZ Is Increased by Pleiotropic CVD Risk Factors

To demonstrate that the observed pattern of differential enrichment does not result from spurious (i.e., nongeneralizable) associations due to category-specific stratification or errors in statistical modeling, we also studied the empirical replication rate across independent substudies of SCZ. Figures 1E and 1F show the empirical cumulative replication-rate plots as a function of nominal p value for the same categories as for the conditional stratified TDR plots in Figures 1C and 1D. Consistent with the conditional TDR pattern, we found that the nominal p value corresponding to a wide range of replication rates was 100 times higher for the $-\log_{10}(p_{TG}) \geq 3$ category than for the $-\log_{10}(p_{TG}) \geq 0$ category (Figure 1E). Similarly, SNPs from pleiotropic SNP categories showing the greatest enrichments ($-\log_{10}(p_{TG}) \geq 3$) replicated at the highest rates—up to five times higher than all SNPs ($-\log_{10}(p_{TG}) \geq 0$)—for a wide range of p value thresholds. This suggests that adjusting p value thresholds according to the estimated category-specific conditional TDR could improve the discovery of replicating SNP associations. The same relationship between conditional TDR and replication rate was shown for SCZ|WHR (Figure 1F), but here,

Table 2. Conditional FDR: CVD-Risk-Factor Loci Given SCZ

Locus	SNP	Gene	MIM	Chr	TG|SCZ	LDL|SCZ	HDL|SCZ	SBP|SCZ	BMI|SCZ	WHR|SCZ	T2D|SCZ
9	rs780110	*IFT172*	607386	2p23.3	**0.00000**	0.73578	0.66350	0.88851	0.57686	**0.01079**	1.00000
	rs2272417	*IFT172*	607386	2p23.3	**0.00000**	0.86268	0.55896	0.83749	0.70039	0.06244	1.00000
20	rs6759206	*AGAP1*	608651	2q37	**0.01764**	0.89696	0.25333	1.00000	1.00000	0.95347	1.00000
22	rs3617	*ITIH3*	146650	3p21.1	0.69128	0.84071	0.37022	0.97795	0.45287	**0.00942**	1.00000
	rs2276817	*ITIH4*	600564	3p21.1	0.28255	**0.04717**	0.25333	0.61208	0.45287	1.00000	1.00000
37	rs2328893	*SLC17A4*	604216	6p22.2	**0.03788**	0.34581	**0.00396**	0.83749	0.65586	1.00000	1.00000
	rs1324082	*SLC17A1*	182308	6p22.2	**0.03113**	0.63999	**0.00465**	0.65717	0.78940	0.95347	1.00000
	rs13198474	*SLC17A3*	611034	6p22.2	0.69128	0.73578	**0.00289**	0.80634	1.00000	0.93285	1.00000
	rs16891235	*HIST1H1A*	142709	6p22.2	0.95191	**0.02569**	**0.00213**	0.70268	1.00000	0.93285	1.00000
	rs13194781	*HIST1H2BN*	602801	6p22.2	**0.00239**	0.97314	0.14244	0.88851	1.00000	0.93285	1.00000
	rs1235162	*GABBR1*	603540	6p22.1	**0.00117**	0.73578	0.10885	0.70268	0.82974	1.00000	1.00000
	rs2844762	*HLA-B*	142830	6p22.1	**0.00491**	0.53895	0.78537	0.61208	NA	0.93285	1.00000
	rs3130380	*HCG18*	-	6p22.1	**0.00708**	0.73578	**0.01852**	0.77857	0.70039	0.81643	1.00000
	rs2524222	*GNL1*	143024	6p22.1	0.28255	**0.02945**	0.41447	0.80634	1.00000	0.93285	1.00000
	rs9262143	*KIAA1949*	610990	6p22.1	**0.00004**	0.26238	0.05759	0.77857	0.92201	0.52829	1.00000
	rs3095326	*IER3*	602996	6p22.1	**0.00003**	**0.04717**	**0.04502**	0.74450	0.92201	0.42354	1.00000
	rs3099840	*HCP5*	604676	6p21.3	**0.00000**	0.39032	**0.02988**	0.28698	1.00000	0.37454	1.00000
	rs2284178	*HCP5*	604676	6p21.3	**0.01764**	0.48709	0.25333	0.18351	0.74603	0.87368	1.00000
	rs805294	*LY6G6C*	610435	6p21.33	1.00000	0.97314	0.12393	**0.00248**	0.61339	0.75370	1.00000
	rs3117577	*MSH5*	603382	6p21.3	**0.00000**	**0.02164**	0.41447	0.61208	0.87106	0.42354	1.00000
	rs3130679	*C6orf48*	605447	6p21.33	**0.00000**	0.07243	0.14244	0.41364	0.70039	0.13758	1.00000
	rs412657	*AK123889*	-	6p21.33	0.69128	0.97314	**0.03447**	0.65717	0.65586	0.37454	1.00000
	rs9268219	*C6orf10*	606766	6p21.33	**0.00000**	**0.04220**	0.12393	0.38400	0.65586	**0.03366**	1.00000
	rs3129963	*BTNL2*	606000	6p21.33	0.59071	0.77938	**0.00548**	0.52604	0.92201	**0.04119**	1.00000
	rs9268853	*HLA-DRA*	142860	6p21.3	0.69128	0.81421	**0.03447**	0.41364	0.61339	**0.02983**	1.00000
	rs9275524	*HLA-DQA2*	613503	6p21.32	**0.00409**	**0.03128**	**0.00548**	0.33310	0.27214	0.05832	1.00000
39	rs1480380	*HLA-DMA*	142855	6p21.3	**0.00708**	0.86268	0.41447	0.18351	0.78940	0.10401	NA
40	rs7832	*C6orf108*	-	6p21.1	**0.03399**	0.97057	0.10762	NA	NA	NA	NA
51	rs983309	*AK055863*	-	8p23.1	0.48760	**0.00000**	**0.00000**	0.80634	0.78940	0.47533	1.00000
	rs17660635	*AK055863*	-	8p23.1	0.69128	**0.00080**	**0.00010**	0.74450	0.92201	0.81643	1.00000
65	rs4919666	*SUFU*	607035	10q24.32	0.85168	0.86268	0.78537	**0.04405**	0.40025	0.87368	1.00000
	rs2296569	*CNNM2*	607803	10q24.32	0.15574	0.59079	**0.03950**	1.00000	1.00000	1.00000	1.00000
	rs11191560	*NT5C2*	600417	10q24.32	0.69128	0.97314	0.72193	**0.00000**	**0.02776**	0.47533	1.00000
	rs11191580	*NT5C2*	600417	10q24.32	0.78905	1.00000	0.61021	**0.00000**	**0.02897**	0.52829	1.00000
71	rs2958625	*METT5D1*	-	11p14.1	**0.00491**	0.89696	**0.02569**	0.88851	0.52128	0.52829	1.00000
	rs10835491	*METT5D1*	-	11p14.1	**0.00409**	0.89696	**0.03950**	0.88851	0.52128	0.52829	1.00000
78	rs10790734	*PKNOX2*	613066	11q24.2	0.37774	0.89696	1.00000	0.80634	0.65586	**0.04476**	1.00000

For the independent complex or single-gene loci ($r^2 < 0.2$) with SNP(s) with a conditional FDR < 0.01 in SCZ given associated CVD risk factors (Table 1), the conditional FDR in the other direction is provided, i.e., FDR CVD risk factors given association in SCZ. All independent loci are listed consecutively, and the same locus numbering is used as in Table 1. All data were first corrected for genomic inflation. FDR values < 0.05 are in bold. The following abbreviations are used: chr, chromosomal region; TG, triglyceride; SCZ, schizophrenia; LDL, low-density lipoprotein; HDL, high-density lipoprotein; SBP, systolic blood pressure; BMI, body mass index; WHR, waist-to-hip ratio; T2D, type 2 diabetes; and NA, not available.

the increase in enrichment, and thus the increase in replication rate, was weaker than that for SCZ|TG.

SCZ Gene Loci Identified with Conditional FDR

To identify SNPs associated with SCZ, we constructed a conditional Manhattan plot showing the FDR conditional on each of the CVD risk factors (Figure 2). We identified significant loci located on a total of 21 chromosomes (1–19, 21, and 22) associated with SCZ, leveraging the reduced FDR obtained by the associated CVD risk factor. To estimate the number of independent loci, we pruned the associated SNPs (i.e., removed SNPs with LD > 0.2) and identified a total of 106 independent loci with a significance threshold of conditional FDR < 0.05 (Table S2). With the more conservative conditional-FDR threshold of 0.01, there remained 25 significant independent loci, of which 4 were complex and 21 were single genes (Table 1 and black line around large circles in Figure 2). The largest locus was on chromosome 6 in the human-leukocyte-antigen (HLA) region. This is the only locus that would have been discovered by standard methods based on p values (Bonferroni correction), and the 6p21.3 region (close to TRIM26 [MIM 600830]) was significantly associated with SCZ in the primary analysis of the current sample.[13] With the FDR method in SCZ alone, six loci were identified. Of these, the regions close to TRIM26 (6p21.3), MMP16 (8q21.3 [MIM 602262]), CNNM2/NT5C2 (10q24.32 [MIM 607803 and 600417]), and TCF4 (18q21.1 [MIM 602272]) were identified in earlier GWASs only after large replication samples were included,[13,15] except for 6p21.3. The remaining 19 loci would not have been identified in the current sample without the use of the pleiotropy-informed stratified FDR method. Of interest, the AK094607/MIR137 region (1p21.3 [MIM 614304]) and the CSMD1 region (8p23.2 [MIM 608397]) were identified in the primary analysis of the current SCZ sample after the inclusion of a large replication sample,[13] and the ITIH4 (3p21.1 [MIM 600564]) and CACNA1C (12p13.3, locus 81 [MIM 114205]) regions were identified in the primary analysis after combination with a large bipolar-disorder sample.[12,13] Thus, the current pleiotropy-informed FDR method validated nine loci discovered in considerably larger samples and discovered 16 additional loci. Furthermore, several of these additional loci are located in regions with borderline significance association with SCZ in previous studies: AGAP1 (2q37; CENTG2 [MIM 608651]),[13] PTPRG (3p21 [MIM 176886]),[13] MAD1L1 (7p22 [MIM 602686]),[43] STT3A (11q23.3 [MIM 601134]),[13] and PLCB2 (15q15 [MIM 604114]).[13]

Pleiotropic Gene Loci in SCZ and CVD Risk Factors Identified with Conjunction FDR

As a secondary analysis, we investigated whether any of the SNPs associated with SCZ conditioned on CVD (SCZ|CVD) were also significantly associated with CVD risk factors conditioned on SCZ (CVD|SCZ), i.e., the conditional FDR in the opposite direction. We identified ten independent loci (pruned on the basis of LD > 0.2) with a significant association also with the CVD risk factor (conditional FDR < 0.05); these included three complex loci and seven single-gene loci (Table 2). Of these, the ITIH4 region (3p21.1) and the CNNM2/NT5C2 region (10q24.32), in addition to the HLA region on chromosome 6, have been identified in previous SCZ studies after the inclusion of large replication samples.[13] The significant loci were found in the analyses of TG|SCZ (six loci), LDL|SCZ (three loci), HDL|SCZ (four loci), SBP|SCZ (two loci), BMI|SCZ (one locus), and WHR|SCZ (four loci), and six loci were jointly associated with SCZ and more than one CVD risk factor (Table 2). This suggests that overlapping genetic pathways are involved in SCZ and CVD risk factors. The direction of the different SNP associations (Z scores) is shown in Table S3. There was no clear evidence for systematic directions across any of the SNPs in the different phenotypes, probably as a result of complex LD structures, especially on chromosome 6.

Further, to provide a comprehensive, unselected map of pleiotropic loci between SCZ and CVD risk factors, in addition to those primarily associated with SCZ, we performed a conjunction-FDR analysis and constructed a conjunction Manhattan plot (Figure S8). We detected 26 independent pleiotropic loci (pruned on the basis of LD > 0.2; black line around large circles) with a significance threshold of conjunction FDR < 0.05 on a total of 14 chromosomes. See Table S4 for more details.

Discussion

Here, leveraging the power of GWAS data from over 250,000 individuals, we demonstrate that GWASs from associated CVD risk factors can improve discovery of SCZ susceptibility loci. By using the stratified conditional-FDR approach[34,35] in the combined analyses of the SCZ and CVD-risk-factor GWASs, we identified a total of 25 significant loci. By analyzing the SCZ GWAS alone, we identified five loci. In contrast, with standard GWAS methods, one locus was significant in the SCZ sample after genomic-control correction.[13] The identified pleiotropic loci are associated with overlapping biological processes, and nine of them have been identified in previous SCZ GWASs after the inclusion of large additional samples. This shows the feasibility of using a pleiotropy-informed stratified FDR approach in SCZ in combination with associated phenotypes; it is much more cost efficient than increasing the sample size of SCZ individuals.[44]

To date, it has been difficult to use GWASs to discover a significant proportion of the missing heritability of complex human traits and disorders. Our statistical framework is based on the fact that SNPs are not exchangeable. Rather, SNPs with effects in pleiotropic phenotypes have a higher probability of being true non-nulls and hence also a higher probability of being replicated in independent studies. We therefore developed a conditional-FDR

approach for GWAS summary statistics by adapting stratification methods originally used for linkage analysis and microarray expression data.[34,35] Decreased conditional FDR (equivalently, increased conditional TDR) for a given nominal p value increases power to detect true non-null effects. Increased conditional TDR is directly related to increased replication effect sizes and replication rates in de novo samples. Importantly, we validated the conditional-FDR approach by demonstrating increased replication rates in independent SCZ substudies for given nominal p value cutoffs. Equivalently, conditional FDR can be used for controlling FDR at a given level while increasing power to discover non-null SNPs over the usual unconditional approaches that treat all SNPs as exchangeable.[45]

We also developed a conjunction-FDR approach to identify SNPs that are highly pleiotropic with SCZ and one or more CVD risk factors. The conjunction FDR is the minimum of the conditional FDR for SCZ given a CVD risk factor and vice versa. SNPs that exceed a stringent conjunction-FDR threshold are highly probable to be non-null in two phenotypes simultaneously. Of note, conjunction FDR is different from the Fisher combined probability test, for which the alternative hypothesis is that the SNP has a significant effect on at least one (but not necessarily both) phenotypes. We validated our approach by applying a bivariate model to estimate the covariation between SCZ and the CVD risk factors. This showed that for the pleiotropic phenotypes, the degree of pleiotropy is highly significantly different from zero (Figure S3). Further confidence in the significance of the current findings comes from the CI Q-Q plots (Figure S2), which show significant pleiotropic enrichment. Given that the current analyses are based on GWAS summary statistics, the findings depend on correctly computed p values in the original studies.

The current findings of difference in magnitude of enrichment and variation in the pleiotropic loci across the three lipid phenotypes show that the results are not driven by genetic stratification, given that the lipid phenotypes were all obtained from the same individuals[30] and each had approximately the same sample size. Moreover, the improved replication rate with increasing pleiotropic enrichment further argues against nonspecific genetic stratification. As such, polygenic pleiotropy could potentially be a nonspecific phenomenon related to heritable human phenotypes, but the lack of polygenic enrichment and of significant loci between SCZ and T2D (an example of a successful GWAS[31]) suggests that the current results are phenotype specific. It is unlikely that nongenetic correlations explain the observed pleiotropy given that only a fraction of control participants in the CVD-risk-factor GWAS samples overlapped with the SCZ GWAS samples (WTCCC controls). The replication rate based on substudies was not driven by the UK sample (which included the overlapping WTCCC controls), as shown in Figure S4. The current threshold for significant

association of SCZ|CVD was set at FDR < 0.01 as a result of the seven CVD risk factors tested. However, the CVD risk factors are highly correlated, and thus the 0.01 level is conservative despite the number of CVD phenotypes tested and is comparable to the standard FDR threshold of 0.05, which translates to 5 false positives per 100 findings.

In the current study, we defined pleiotropy as the association between a single gene or variant and more than one distinct phenotype (diseases or traits).[9] It is possible that some of the loci identified in the current study might not be pleiotropic but rather underlie common aspects of the SCZ and CVD-risk-factor phenotypes.[9] This can be investigated in samples with more detailed phenotypes. In the present study, we focused on SNPs, but gene-based pleiotropy is also interesting;[8] however, this requires raw data from individual participants.

Our results implicate potential shared pathological mechanisms between SCZ and CVD risk factors. The ten pleiotropic loci were associated with multiple CVD phenotypes, supporting the hypothesis that the pathobiology of SCZ is heterogeneous and has numerous underlying mechanisms. The majority of the pleiotropic signal was found with lipid levels, suggesting that lipid biology might be involved in SCZ pathophysiology. As such, genetically determined dyslipidemia in SCZ is in line with evidence for white-matter abnormalities and myelin dysfunction[46,47] and supports the neurodevelopmental hypothesis.[48] However, the lack of consistent directionality suggests the need for further experimental studies for determining the mechanistic relationship between dyslipidemia and SCZ.

Our results show that a "model-free," empirical, FDR framework that uses unthresholded summary-statistics data from independent GWASs can provide insights into relationships between risk factors and diseases. This approach can be used for examining the shared genetic basis between a number of diseases and traits. With the recent discovery of many common genetic variants influencing diseases and traits, there is increasing interest in pleiotropy. One recent review suggests that pleiotropy is common and associated with ~17% of genes and ~5% of SNPs associated with complex humans diseases and traits.[9] In addition to identifying potential targets for drug development, gaining insight into the degree of genetic "connectivity" between diseases and traits provides an opportunity to ascertain whether current diagnoses and classifications are consistent with genetic architecture or whether genetic similarities traverse clinical conditions. Examining overlap in common variants can elucidate important pathobiology and might identify potential therapeutic targets for common diseases.

In conclusion, the current findings demonstrate that in SCZ, the pleiotropy-informed stratified FDR method can improve the statistical power for detecting "polygenic" effects and can offer insights into mechanistic relationships between lipid biology and SCZ pathogenesis.

Supplemental Data

Supplemental Data include eight figures, six tables, and a list of Schizophrenia Psychiatric GWAS Consortium members and can be found with this article online at http://www.cell.com/AJHG.

Web Resources

The URLs for data presented here are as follows:

1000 Genomes Project, http://www.1000genomes.org/
DAVID Bioinformatics Resources 6.7, http://david.abcc.ncifcrf.gov/content.jsp?file=functional_annotation.html
dbSNP, http://www.ncbi.nlm.nih.gov/projects/SNP/
DIAGRAM Consortium, www.well.ox.ac.uk/DIAGRAM/
International HapMap Project, http://hapmap.ncbi.nlm.nih.gov/
Wellcome Trust Case Control Consortium (WTCCC), http://www.wtccc.org.uk/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/
Psychiatric Genomics Consortium, https://pgc.unc.edu/index.php
SNP Annotation and Proxy Search (SNAP), http://www.broadinstitute.org/mpg/snap

References

1. Glazier, A.M., Nadeau, J.H., and Aitman, T.J. (2002). Finding genes that underlie complex traits. Science 298, 2345–2349.

2. Hirschhorn, J.N., and Daly, M.J. (2005). Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. 6, 95–108.

3. Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S., and Manolio, T.A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc. Natl. Acad. Sci. USA 106, 9362–9367.

4. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature 461, 747–753.

5. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42, 565–569.

6. Yang, J., Manolio, T.A., Pasquale, L.R., Boerwinkle, E., Caporaso, N., Cunningham, J.M., de Andrade, M., Feenstra, B., Feingold, E., Hayes, M.G., et al. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. Nat. Genet. 43, 519–525.

7. Stahl, E.A., Wegmann, D., Trynka, G., Gutierrez-Achury, J., Do, R., Voight, B.F., Kraft, P., Chen, R., Kallberg, H.J., Kurreeman, F.A., et al.; Diabetes Genetics Replication and Meta-analysis Consortium; Myocardial Infarction Genetics Consortium. (2012). Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nat. Genet. 44, 483–489.

8. Wagner, G.P., and Zhang, J. (2011). The pleiotropic structure of the genotype-phenotype map: The evolvability of complex organisms. Nat. Rev. Genet. 12, 204–213.

9. Sivakumaran, S., Agakov, F., Theodoratou, E., Prendergast, J.G., Zgaga, L., Manolio, T., Rudan, I., McKeigue, P., Wilson, J.F., and Campbell, H. (2011). Abundant pleiotropy in human complex diseases and traits. Am. J. Hum. Genet. 89, 607–618.

10. Chambers, J.C., Zhang, W., Sehmi, J., Li, X., Wass, M.N., Van der Harst, P., Holm, H., Sanna, S., Kavousi, M., Baumeister, S.E., et al.; Alcohol Genome-wide Association (AlcGen) Consortium; Diabetes Genetics Replication and Meta-analyses (DIAGRAM+) Study; Genetic Investigation of Anthropometric Traits (GIANT) Consortium; Global Lipids Genetics Consortium; Genetics of Liver Disease (GOLD) Consortium; International Consortium for Blood Pressure (ICBP-GWAS); Meta-analyses of Glucose and Insulin-Related Traits Consortium (MAGIC). (2011). Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. Nat. Genet. 43, 1131–1138.

11. Cotsapas, C., Voight, B.F., Rossin, E., Lage, K., Neale, B.M., Wallace, C., Abecasis, G.R., Barrett, J.C., Behrens, T., Cho, J., et al.; FOCiS Network of Consortia. (2011). Pervasive sharing of genetic effects in autoimmune disease. PLoS Genet. 7, e1002254.

12. Sklar, P., Ripke, S., Scott, L.J., Andreassen, O.A., Cichon, S., Craddock, N., Edenberg, H.J., Nurnberger, J.I., Jr., Rietschel, M., Blackwood, D., et al.; Psychiatric GWAS Consortium Bipolar Disorder Working Group. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat. Genet. 43, 977–983.

13. Ripke, S., Sanders, A.R., Kendler, K.S., Levinson, D.F., Sklar, P., Holmans, P.A., Lin, D.Y., Duan, J., Ophoff, R.A., Andreassen, O.A., et al.; Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. (2011). Genome-wide association study identifies five new schizophrenia loci. Nat. Genet. 43, 969–976.

14. Lichtenstein, P., Yip, B.H., Björk, C., Pawitan, Y., Cannon, T.D., Sullivan, P.F., and Hultman, C.M. (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: A population-based study. Lancet 373, 234–239.

15. Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., Werge, T., Pietiläinen, O.P., Mors, O., Mortensen, P.B., et al.; Genetic Risk and Outcome in Psychosis (GROUP). (2009). Common variants conferring risk of schizophrenia. Nature 460, 744–747.

16. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P.; International Schizophrenia Consortium. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460, 748–752.

17. C.J. Murray, and A.D. Lopez, eds. (1996). The Global Burden of Disease: A comprehensive assessment of mortality, injuries, and risk factors in 1990 and projected to 2020. In Global Burden of disease and Injury Series, Vol. 1. (Cambridge, MA: Harvard University Press).

18. Colton, C.W., and Manderscheid, R.W. (2006). Congruencies in increased mortality rates, years of potential life lost, and causes of death among public mental health clients in eight states. Prev. Chronic Dis. 3, A42.

19. Laursen, T.M., Munk-Olsen, T., and Vestergaard, M. (2012). Life expectancy and cardiovascular mortality in persons with schizophrenia. Curr. Opin. Psychiatry 25, 83–88.

20. Saha, S., Chant, D., and McGrath, J. (2007). A systematic review of mortality in schizophrenia: Is the differential mortality gap worsening over time? Arch. Gen. Psychiatry 64, 1123–1131.

21. Marder, S.R., Essock, S.M., Miller, A.L., Buchanan, R.W., Casey, D.E., Davis, J.M., Kane, J.M., Lieberman, J.A., Schooler, N.R., Covell, N., et al. (2004). Physical health monitoring of patients with schizophrenia. Am. J. Psychiatry 161, 1334–1349.

22. Mitchell, A.J., Vancampfort, D., Sweers, K., van Winkel, R., Yu, W., and De Hert, M. (2011). Prevalence of Metabolic Syndrome and Metabolic Abnormalities in Schizophrenia and Related Disorders—A Systematic Review and Meta-Analysis. Schizophr. Bull.

23. American Diabetes Association; American Psychiatric Association; American Association of Clinical Endocrinologists; North American Association for the Study of Obesity. (2004). Consensus development conference on antipsychotic drugs and obesity and diabetes. Diabetes Care 27, 596–601.

24. De Hert, M.A., van Winkel, R., Van Eyck, D., Hanssens, L., Wampers, M., Scheen, A., and Peuskens, J. (2006). Prevalence of the metabolic syndrome in patients with schizophrenia treated with antipsychotic medication. Schizophr. Res. 83, 87–93.

25. Kaddurah-Daouk, R., McEvoy, J., Baillie, R.A., Lee, D., Yao, J.K., Doraiswamy, P.M., and Krishnan, K.R. (2007). Metabolomic mapping of atypical antipsychotic effects in schizophrenia. Mol. Psychiatry 12, 934–945.

26. Raphael, T.P., and Parsons, J.P. (1921). Blood sugar studies in dementia praecox and manic-depressive insanity. Arch. Neurol. Psychiatry 5, 687–709.

27. Ryan, M.C., Collins, P., and Thakore, J.H. (2003). Impaired fasting glucose tolerance in first-episode, drug-naive patients with schizophrenia. Am. J. Psychiatry 160, 284–289.

28. Hansen, T., Ingason, A., Djurovic, S., Melle, I., Fenger, M., Gustafsson, O., Jakobsen, K.D., Rasmussen, H.B., Tosato, S., Rietschel, M., et al. (2011). At-risk variant in TCF7L2 for type II diabetes increases risk of schizophrenia. Biol. Psychiatry 70, 59–63.

29. Ehret, G.B., Munroe, P.B., Rice, K.M., Bochud, M., Johnson, A.D., Chasman, D.I., Smith, A.V., Tobin, M.D., Verwoert, G.C., Hwang, S.J., et al.; International Consortium for Blood Pressure Genome-Wide Association Studies; CARDIoGRAM consortium; CKDGen Consortium; KidneyGen Consortium; EchoGen consortium; CHARGE-HF consortium. (2011). Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 478, 103–109.

30. Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466, 707–713.

31. Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G., et al.; MAGIC investigators; GIANT Consortium. (2010). Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat. Genet. 42, 579–589.

32. Speliotes, E.K., Willer, C.J., Berndt, S.I., Monda, K.L., Thorleifsson, G., Jackson, A.U., Lango Allen, H., Lindgren, C.M., Luan, J., Mägi, R., et al.; MAGIC; Procardis Consortium. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat. Genet. 42, 937–948.

33. Heid, I.M., Jackson, A.U., Randall, J.C., Winkler, T.W., Qi, L., Steinthorsdottir, V., Thorleifsson, G., Zillikens, M.C., Speliotes, E.K., Mägi, R., et al.; MAGIC. (2010). Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. Nat. Genet. 42, 949–960.

34. Yoo, Y.J., Pinnaduwage, D., Waggott, D., Bull, S.B., and Sun, L. (2009). Genome-wide association analyses of North American Rheumatoid Arthritis Consortium and Framingham Heart Study data utilizing genome-wide linkage results. BMC Proc 3(Suppl 7), S103.

35. Sun, L., Craiu, R.V., Paterson, A.D., and Bull, S.B. (2006). Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies. Genet. Epidemiol. 30, 519–530.

36. Efron, B. (2010). Large-scale inference: Empirical Bayes methods for estimation, testing, and prediction (New York: Cambridge University Press).

37. Schweder, T., and Spjotvoll, E. (1982). Plots of P-Values to Evaluate Many Tests Simultaneously. Biometrika 69, 493–502.

38. Devlin, B., and Roeder, K. (1999). Genomic control for association studies. Biometrics 55, 997–1004.

39. Yang, J., Weedon, M.N., Purcell, S., Lettre, G., Estrada, K., Willer, C.J., Smith, A.V., Ingelsson, E., O'Connell, J.R., Mangino, M., et al.; GIANT Consortium. (2011). Genomic inflation factors under polygenic inheritance. Eur. J. Hum. Genet. 19, 807–812.

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

41. Efron, B. (2007). Size, power and false discovery rates. Ann. Stat. 35, 1351–1377.

42. Nichols, T., Brett, M., Andersson, J., Wager, T., and Poline, J.B. (2005). Valid conjunction inference with the minimum statistic. Neuroimage 25, 653–660.

43. Wang, K.S., Liu, X.F., and Aragam, N. (2010). A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. Schizophr. Res. 124, 192–199.

44. Sullivan, P.F. (2012). Puzzling over schizophrenia: Schizophrenia as a pathway disease. Nat. Med. 18, 210–211.

45. Craiu, R.V., and Sun, L. (2008). Choosing the lesser evil: Trade-off between false discovery rate and non-discovery rate. Stat. Sin. 18, 861–879.

46. Davis, K.L., Stewart, D.G., Friedman, J.I., Buchsbaum, M., Harvey, P.D., Hof, P.R., Buxbaum, J., and Haroutunian, V. (2003). White matter changes in schizophrenia: Evidence for myelin-related dysfunction. Arch. Gen. Psychiatry 60, 443–456.

47. Karoutzou, G., Emrich, H.M., and Dietrich, D.E. (2008). The myelin-pathogenesis puzzle in schizophrenia: A literature review. Mol. Psychiatry 13, 245–260.

48. Marenco, S., and Weinberger, D.R. (2000). The neurodevelopmental hypothesis of schizophrenia: Following a trail of evidence from cradle to grave. Dev. Psychopathol. 12, 501–527.