**Dissecting Schizophrenia Biology Using Cognitive Genomics Pleiotropy**

Upasana Bhattacharyya\*1,2, Jibin John\*1,2, Todd Lencz§1,2,3, Max Lam§1,4,5

1Institute of Behavioral Science, Feinstein Institutes for Medical Research, Manhasset, NY

2Division of Psychiatry Research, Zucker Hillside Hospital, Glen Oaks, NY

3Departments of Psychiatry and Molecular Medicine, Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY

4Institute of Mental Health, Singapore

5Lee Kong Chian School of Medicine, Population and Global Health, Nanyang Technological University

\*These authors contributed equally

§These authors contributed equally

Corresponding Authors: tlencz@northwell.edu; max.lam@ntu.edu.sg

Supplementary Methods

The subsequent sections detail our methodological approach, encompassing data curation, pleiotropic meta-analysis, identification of independent genome-wide significant loci, labeling of loci in terms of the presence of concordant and discordant variants, and characterization of meta-analytic loci relative to input GWAS. Additionally, we outline our gene set analysis technique and examine these gene sets using temporal gene expression profiles in the brain.

## 1. Data Curation

We initiated the data curation process by collecting the most recent genome-wide association study (GWAS) summary statistics for the relevant traits. All GWAS summary statistics used in this study consisted of individuals with European ancestry1–3. Specifically, for schizophrenia, we obtained the GWAS summary statistics from the Psychiatric Genomics Consortium (<https://pgc.unc.edu/for-researchers/download-results/>; N= 53,386 cases and 77,258 controls; GWAS mean χ2 =2.0126; total Variants = 7,659,767)3. The GWAS summary statistics for educational attainment were acquired from the Social Science Genomics Association Consortium (SSGAC;https://thessgac.com) (N=766,345; mean GWAS χ²=2.647; total Variants=10,101,242)2. This GWAS was selected, rather than the most recent educational attainment GWAS4, to maintain a sample size comparable to the other GWASs included in the current report. We utilized the summary statistics from our previous meta-analytic publication for cognitive task performance, the largest available for the phenotype (N=373,617; GWAS mean χ2 =2.11; Total Variants = 8,050,310)1.

We implemented rigorous quality control (QC) protocols on the GWAS summary statistics using the MungeSumstat tool5. This process entailed a thorough suite of QC checks designed to rectify common genetic data inconsistencies, including allele discrepancies, strand ambiguity, and anomalies in chromosome nomenclature. We further enhanced the uniformity of GWAS summary statistics by standardizing strand designation, SNP ID, reference genome build, effect allele, p-value, and effect size formats. We ensured every variant displayed a nonzero effect size and/or standard error, with sample counts (N) maintained as integers across all datasets. Our QC measures also eliminated non-biallelic SNPs, indels, and SNPs with INFO scores below 0.3 or minor allele frequencies under 0.001. Following these rigorous quality control steps, we retained 6,803,445 SNPs consistent across the three phenotypes and available for subsequent analysis.

## 2. Meta-Analysis

We then utilized PLEIO to conduct a cross-phenotype pleiotropic meta-analysis to identify shared genetic loci between schizophrenia, cognition, and education. PLEIO has been demonstrated6 to have several advantages over the ASSET7 (Association analysis based on subsets) method, which we used in our prior report8. PLEIO overcomes limitations associated with fixed-effect model approaches by employing a random-effect model that captures genetic correlations and heritabilities across trait pairs, maximizing statistical power6. Additionally, PLEIO enhances computational efficiency relative to other approaches by employing spectral decomposition on the covariance matrix. PLEIO utilizes a variance component test to identify pleiotropic SNPs that influence multiple traits simultaneously, incorporating an importance sampling method to provide an accurate p-value estimate.

In this analysis, we have utilized the *ldsc\_preprocess.py* script implemented in PLEIO that calls the LDSC software6,8 (<https://github.com/bulik/ldsc>) to estimate both genetic covariance and environmental correlation matrices across traits - at the same time generating GWAS summary statistics with the standardized effect sizes, before carrying out the pleiotropic meta-analysis. We then used the *Pleio.py* script (<https://github.com/cuelee/pleio>) implemented in PLEIO to perform the pleiotropic meta-analysis. Specific codes and commands are provided in the GitHub repository (https://github.com/mlamcogent/cogent-pleiotropy-proj/tree/main)

## 3. Identification of Subsets of Variants with Homogenous Relationships across Schizophrenia, Educational Attainment, and Cognitive Ability

PLEIO employs a recalibration process to calculate trait-specific beta values and estimate their standard errors using the *--blup* option during the meta-analysis6. Consistent with our earlier report9, we utilized the directionality of the recalibrated beta effect sizes generated as part of the pleiotropic meta-analytic procedures to identify subsets of variants genome-wide across schizophrenia, cognitive task performance, and educational attainment. As with our earlier report, the direction of effect sizes for each variant was reversed within the schizophrenia GWAS summary statistics to align with the interpretation of cognition and education scores, such that a negative allele effect would indicate risk for schizophrenia and a positive allele effect would suggest a protective effect to schizophrenia.

Variants were then organized into subsets described in detail in our prior report9. Also consistent with our earlier report9, the symbol ∩ designates a variant with the same effect direction across any two phenotypes (after the reversal of schizophrenia). In contrast, the symbol | denotes variants in which one phenotype displays an effect direction that opposes those of the other two traits (again, after the reversal of schizophrenia). Variants with beta close to zero (between 0.001 and -0.001) for a given input trait were considered to have no direction of effect for that trait. Thus, we define one subgroup of variants as "Concordant" if allelic effects associated with increased cognitive ability and educational attainment are associated with a decrease in schizophrenia risk (scz ∩ edu ∩ cog). By contrast, "Discordant" variants, as in our prior report, are comprised of three subsets: (1) edu ∩ cog | scz (schizophrenia outliers, which have alleles associated with an increase in cognitive ability, educational attainment, and schizophrenia risk); (2) scz ∩ cog | edu (education outliers, where alleles associated with an increase in schizophrenia risk and reduced cognitive ability are also associated with increased educational attainment) and (3) scz ∩ edu | cog (cognition outliers, with alleles associated with an increase in schizophrenia risk and reduced educational attainment, but also increased cognitive performance). These subgroups are further represented visually in Supplementary Fig 1.

## 4. Identification and Consolidation of Independent Loci

Independent genome-wide-significant loci for each PLEIO meta-analysis subset (Concordant and DIscordant) were identified using the FUnctional Mapping and Annotation (FUMA; version 1.5.2) pipeline. We have used the default parameters of FUMA to perform double clumping of SNPs based on the European 1000 Genomes Project Phase 3 reference panels. The clumping of variants to identify significant genome-wide loci was done in two steps. First, the clumping parameter was set to clump SNPs with *p*-value < 5x10‐8 and an LD of r2 < 0.6 to identify “independent significant SNPs.” The second clumping step identified ‘lead SNPs” using r2 < 0.1 to clump independent significant SNPs with the strongest signal at a given locus. The MHC region (Chr6:25MB-35MB) was excluded from the clumping analysis; due to the highly complex LD structure within the region, only a single top SNP from this region was retained. We included SNPs from the 1000 Genomes reference panel represented within the given locus to maximize the potential for biological annotation within identified loci. LD blocks of independent significant variants 250 kb or less were merged into a single genomic locus. We identified 503 loci harboring concordant SNPs and 428 loci discordant SNPs. However, some overlap existed between these categories; following the procedure followed in our prior report10, we harmonized the loci obtained from the two subsets using default settings of the “*merge*” function implemented in the “bedtools” package (Version 2; <https://github.com/arq5x/bedtools2>)10 i.e, merging loci overlapping each other by at least 1 base pair. While harmonizing across significant genomic loci, we found that many contained both concordant and discordant variant subsets. For the current report, these are called "Dual"loci. For subsequent analysis, we subgrouped the pleiotropic meta-analysis summary statistics from PLEIO into three distinct groups of SNPs: (i) those within concordant loci, (ii) those solely within discordant loci, and (iii) those present in the dual loci.

##

## 5. Validation of Novel Loci

Consistent with our earlier work9, we sought to validate the novel loci against more recent summary statistics to demonstrate that an outcome of investigating pleiotropy is increased discovery power. As we are presently already using the latest and most well-powered GWAS summary statistics input for schizophrenia and cognitive task performance, it would be out of the scope of the current report to validate novel loci in those traits. Nonetheless, we were able to check if any of the novel loci appeared in the recent educational attainment GWAS20, which has almost three times the sample size of the current one2. However, in the recent educational GWAS, full locus data (including all SNPs at each significant locus) were not reported, and complete summary statistics, including 23andMe, were unavailable. Thus we downloaded the file EA4\_additive\_p1e-5\_clumped.txt from <https://www.thessgac.org/>, which contains additive GWAS meta-analysis of all discovery cohorts for approximately 8,618 independent (pairwise r2 < 0.1) autosomal variants with P<1x10-5 and N>500,000 (total sample size = 3,037,499). To conduct the look-up, we selected genome-wide significant variants (adjusted P<1.29x10-8). We checked if any loci identified from our analysis overlap within 20kb upstream or downstream of any reported SNPs.

## 6. MAGMA gene set analysis

Gene set analyses were performed for each of the three subsets (Concordant, Discordant, and Dual) separately using MAGMA v1.10, consisting of three steps: annotation, to map SNPs onto genes; gene analysis, to compute gene p-values; and finally, gene set analysis. We utilized the gene location file (build 37) of protein-coding genes obtained from https://ctg.cncr.nl/software/magma for annotation of SNPs to genes in MAGMA. We defined the boundaries of the genes as 35 kb upstream and 10 kb downstream to include likely regulatory regions associated with these genes11–14 Following annotation, gene-level analysis was conducted, and a snp-wise mean model was used to compute the p-value for each gene. The gene sets were obtained from the most recent GO annotations presented in the GO-basic obo file released on 2023-07-27 and NCBI gene2go annotations downloaded on 2023-09-01 by MSigdb. We used the C5 collection of MSigdb (MSigDB v2023.1.Hs). These annotations are based on specific references that describe the experiments or analyses supporting the association between a particular gene product and a corresponding Gene Ontology (GO) term, along with an evidence code to indicate the level of support for the association with the given GO term. In this subcollection, the gene sets are identified with prefixes such as "GOBP" (Biological Process), "GOMF" (Molecular Function), or "GOCC" (Cellular Component), indicating their respective sources in the Gene Ontology. We removed any genes belonging to chromosome X or Y or extended MHC region (Chr6:25000000-35000000), as these regions were not included in the meta-analysis. We then removed any gene sets that contained less than 10 genes. Finally, MAGMA competitive gene set analysis was conducted against 7,353 gene sets. Due to considerable redundancy across GO annotations, a strict Bonferroni correction would be overly conservative; consequently, we employed a false discovery rate threshold (FDR<0.05) using the Benjamini-Hochberg procedure15.

**7. Follow-up Temporal Gene Expression Analysis via BrainSpan**

Follow-up annotation of gene set enrichment results was conducted via publicly available BrainSpan data (<https://www.brainspan.org/static/download.html>). Detailed methodology describing the pre-processing and curation of the BrainSpan data has been described elsewhere16,17. MAGMA gene set enrichment outputs were post-processed by selecting only Gene Ontology (GO) gene sets that showed significant association at FDR < 0.05 separately for Concordant, Discordant, and Dual loci. Gene-based p-values for each gene list were vectorized and converted to -logP values for subsequent weighting of expression values. We recoded the developmental stage into “Weeks'' of development to obtain a higher spatial-temporal gene expression profile resolution, as we have done previously1. We aggregated the gene expression for a given gene list across brain regions to estimate mean gene expression over time, yielding a whole-brain gene expression estimate. In addition, gene expression for each gene list was weighted by the gene-based -log10 P-values and then normalized by the total gene expression across the BrainSpan data set. Final gene expression profiles were then expressed as gene expression Z-scores. To examine differences in gene expression for each Concordant, Discordant, and Dual gene lists, we carried out linear mixed model analysis for each pair of expression profiles over time (i) Concordant-Discordant (ii) Dual-Discordant (iii) Dual-Concordant. Based on our expectations, the gene expression profile for the Concordant genes, with neurodevelopmental enrichment, would follow a negative gradient; conversely, the gene expression profile for Discordant genes would be follow a positive gradient. To evaluate the hypothesis, we built an interaction model with trait pair (Concordant v Discordant; Dual v Concordant; Dual v Concordant) \* Time. We report both the main effect and interaction effects for the trait pairs indicated.

###

##

## Supplementary Table Captions

**Supplementary Table 1a**: GWAS significant Pleiotropic Genomic Risk Loci for Concordant Variants

*Note:* Genomic locus : Index of genomic rick loci; uniqID : Unique ID of SNPs consisting of chr:position:allele1:allele2 where alleles are alphabetically ordered; rsID : rsID of the top lead SNP based on dbSNP build 146; chr : chromosome of top lead SNP; pos : position of top lead SNP on hg19; P-value : P-value of top lead SNP (from the input file); start : Start position of the locus; end : End position of the locus; nSNPs : The number of unique candidate SNPs in the genomic locus, including non-GWAS-tagged SNPs (which are available in the user selected reference panel). Candidate SNPs are all SNPs that are in LD (give user-defined r2) with any of independent significant SNPs and either have a P-value below the user defined threshold or are only available in 1000G; nGWASSNPs : The number of unique GWAS-tagged candidate SNPs in the genomic locus which is available in the GWAS summary statistics input file. This is a subset of "nSNPs"; nIndSigSNPs : The number of the independent (at user defined r2) significant SNPs in the genomic locus; IndSigSNPs : rsID of the independent significant SNPs in the genomic locus; nLeadSNPs : The number of lead SNPs in the genomic locus. Lead SNPs are subset of independent significant SNPs at r2 0.1; LeadSNPs : rsID of lead SNPs in the genomic locus.

**Supplementary Table 1b**: Annotations for variants within GWAS significant Pleiotropic Genomic Risk Loci for Concordant Variants

*Note:* uniqID : Unique ID of SNPs consisting of chr:position:allele1:allele2 where alleles are alphabetically ordered; rsID : rsID based on dbSNP build 146; chr : chromosome; pos : position on hg19; effect\_allele : Effect/risk allele if it is provided in the input GWAS summary statistics file. If not, this is the alternative (minor) allele in 1000G; non\_effect\_allele : Non-effect/non-risk allele if it is provided in the input GWAS summary statistics file. If not, this is the reference (major) allele in 1000G; MAF : Minor allele frequency computed based on 1000G; gwasP : P-value provided in the input GWAS summary statistics file. Non-GWAS tagged SNPs (which do not exist in input file but are extracted from the reference panel) have "NA" instead; or : Odds ratio provided in the input GWAS summary statistics file if available. Non-GWAS tagged SNPs (which do not exist in input file but are extracted from the reference panel) have "NA" instead; beta : Beta provided in the input GWAS summary statistics file if available. Non-GWAS tagged SNPs (which do not exist in input file but are extracted from the reference panel) have "NA" instead; se : Standard error provided in the input GWAS summary statistics file if available. Non-GWAS tagged SNPs (which do not exist in input file but are extracted from the reference panel) have "NA" instead; r2 : The maximum r2 of the SNP with one of the independent significant SNPs; IndSigSNP : rsID of the independent significant SNP which has the maximum r2 with the SNP; Genomic locus : Index of the genomic risk loci matching with "Genomic risk loci" table; Novel : Whether the index SNP/locus is novel to the pleiotropic analysis, and if the variant/locus has been validated in recent education attainment GWAS (information from Supplementary Table 4); nearestGene : The nearest Gene of the SNP based on ANNOVAR annotations. Note that ANNOVAR annotates "consequence" function by prioritizing the most deleterious annotation for SNPs which are locating a genomic region where multiple genes are obverlapped. Genes are ecoded in symbol, if it is available otherwise; Ensembl ID. Genes include all transcripts from Ensembl gene build 85 including non-protein coding genes and RNAs; dist : Distance to the nearest gene. SNPs which are locating in the gene body or 1kb up- or down-stream of TSS or TES have 0; func : Functional consequence of the SNP on the gene obtained from ANNOVAR. For exonic SNPs, detailed annotation (e.g. non-synonymous, stop gain and so on) is available in the ANNOVAR table (annov.txt); CADD : CADD score which is computed based on 63 annotations. The higher the score, the more deleterious the SNP is. 12.37 is the suggested threshold by Kicher et al (2014); RDB : RegulomeDB score which is a categorical score (from 1a to 7). 1a is the highest score for SNPs with the most biological evidence to be a regulatory element; minChrState : The minimum 15-core chromatin state across 127 tissue/cell type; commonChrState : The most common 15-core chromatin state across 127 tissue/cell types; posMapFilt : Whether the SNP was used for eQTL mapping or not. 1 is used, otherwise 0. When eqtl mapping is not performed, all SNPs have 0.

**Supplementary Table 1c:** GWAS significant variants within r2 = 0.6 of Concordant sentinel variants

*Note:* No : Index of independent significant SNPs; Genomic Locus : Index of assigned genomic locus matched with "Genomic risk loci" table. Multiple independent lead SNPs can be assigned to the same genomic locus; uniqID : Unique ID of SNPs consisting of chr:position:allele1:allele2 where alleles are alphabetically ordered; rsID : rsID based on dbSNP build 146; chr : chromosome; pos : position on hg19; P-value : P-value (from the input file); nSNPs : The number of SNPs in LD with the lead SNP given r2, including non-GWAS-tagged SNPs (which are extracted from 1000G); nGWASSNPs : The number of GWAS-tagged SNPs in LD with the lead SNP given r2. This is a subset of "nSNPs".

**Supplementary Table 1d**: GWAS significant Pleiotropic Concordant sentinel variants within a given genomic locus

*Note:* No : Index of lead SNPs; Genomic Locus : Index of assigned genomic locus matched with "Genomic risk loci" table. Multiple lead SNPs can be assigned to the same genomic locus; uniqID : Unique ID of SNPs consisting of chr:position:allele1:allele2 where alleles are alphabetically ordered; rsID : rsID based on dbSNP build 146; chr : chromosome; pos : position on hg19; P-value : P-value (from the input file); nIndSigSNPs : Number of independent significant SNPs which are in LD with the lead SNP at r2 0.1; IndSigSNPs : rsID of independent significant SNPs which are in LD with the lead SNP at r2 0.1.

**Supplementary Table 2a**: GWAS significant Pleiotropic Genomic Risk Loci for Discordant Variants

*Note:* Genomic locus : Index of genomic rick loci; uniqID : Unique ID of SNPs consisting of chr:position:allele1:allele2 where alleles are alphabetically ordered; rsID : rsID of the top lead SNP based on dbSNP build 146; chr : chromosome of top lead SNP; pos : position of top lead SNP on hg19; P-value : P-value of top lead SNP (from the input file); start : Start position of the locus; end : End position of the locus; nSNPs : The number of unique candidate SNPs in the genomic locus, including non-GWAS-tagged SNPs (which are available in the user selected reference panel). Candidate SNPs are all SNPs that are in LD (give user-defined r2) with any of independent significant SNPs and either have a P-value below the user defined threshold or are only available in 1000G; nGWASSNPs : The number of unique GWAS-tagged candidate SNPs in the genomic locus which is available in the GWAS summary statistics input file. This is a subset of "nSNPs"; nIndSigSNPs : The number of the independent (at user defined r2) significant SNPs in the genomic locus; IndSigSNPs : rsID of the independent significant SNPs in the genomic locus; nLeadSNPs : The number of lead SNPs in the genomic locus. Lead SNPs are subset of independent significant SNPs at r2 0.1; LeadSNPs : rsID of lead SNPs in the genomic locus.

**Supplementary Table 2b**: Annotations for variants within GWAS significant Pleiotropic Genomic Risk Loci for Discordant Variants

*Note:* uniqID : Unique ID of SNPs consisting of chr:position:allele1:allele2 where alleles are alphabetically ordered; rsID : rsID based on dbSNP build 146; chr : chromosome; pos : position on hg19; effect\_allele : Effect/risk allele if it is provided in the input GWAS summary statistics file. If not, this is the alternative (minor) allele in 1000G; non\_effect\_allele : Non-effect/non-risk allele if it is provided in the input GWAS summary statistics file. If not, this is the reference (major) allele in 1000G; MAF : Minor allele frequency computed based on 1000G; gwasP : P-value provided in the input GWAS summary statistics file. Non-GWAS tagged SNPs (which do not exist in input file but are extracted from the reference panel) have "NA" instead; or : Odds ratio provided in the input GWAS summary statistics file if available. Non-GWAS tagged SNPs (which do not exist in input file but are extracted from the reference panel) have "NA" instead; beta : Beta provided in the input GWAS summary statistics file if available. Non-GWAS tagged SNPs (which do not exist in input file but are extracted from the reference panel) have "NA" instead; se : Standard error provided in the input GWAS summary statistics file if available. Non-GWAS tagged SNPs (which do not exist in input file but are extracted from the reference panel) have "NA" instead; r2 : The maximum r2 of the SNP with one of the independent significant SNPs; IndSigSNP : rsID of the independent significant SNP which has the maximum r2 with the SNP; Genomic locus : Index of the genomic risk loci matching with "Genomic risk loci" table; Novel : Whether the index SNP/locus is novel to the pleiotropic analysis, and if the variant/locus has been validated in recent education attainment GWAS (information from Supplementary Table 4); nearestGene : The nearest Gene of the SNP based on ANNOVAR annotations. Note that ANNOVAR annotates "consequence" function by prioritizing the most deleterious annotation for SNPs which are locating a genomic region where multiple genes are obverlapped. Genes are ecoded in symbol, if it is available otherwise; Ensembl ID. Genes include all transcripts from Ensembl gene build 85 including non-protein coding genes and RNAs; dist : Distance to the nearest gene. SNPs which are locating in the gene body or 1kb up- or down-stream of TSS or TES have 0; func : Functional consequence of the SNP on the gene obtained from ANNOVAR. For exonic SNPs, detailed annotation (e.g. non-synonymous, stop gain and so on) is available in the ANNOVAR table (annov.txt); CADD : CADD score which is computed based on 63 annotations. The higher the score, the more deleterious the SNP is. 12.37 is the suggested threshold by Kicher et al (2014); RDB : RegulomeDB score which is a categorical score (from 1a to 7). 1a is the highest score for SNPs with the most biological evidence to be a regulatory element; minChrState : The minimum 15-core chromatin state across 127 tissue/cell type; commonChrState : The most common 15-core chromatin state across 127 tissue/cell types; posMapFilt : Whether the SNP was used for eQTL mapping or not. 1 is used, otherwise 0. When eqtl mapping is not performed, all SNPs have 0.

**Supplementary Table 2c:** GWAS significant variants within r2 = 0.6 of Discordant sentinel variants

*Note:* No : Index of independent significant SNPs; Genomic Locus : Index of assigned genomic locus matched with "Genomic risk loci" table. Multiple independent lead SNPs can be assigned to the same genomic locus; uniqID : Unique ID of SNPs consisting of chr:position:allele1:allele2 where alleles are alphabetically ordered; rsID : rsID based on dbSNP build 146; chr : chromosome; pos : position on hg19; P-value : P-value (from the input file); nSNPs : The number of SNPs in LD with the lead SNP given r2, including non-GWAS-tagged SNPs (which are extracted from 1000G); nGWASSNPs : The number of GWAS-tagged SNPs in LD with the lead SNP given r2. This is a subset of "nSNPs".

**Supplementary Table 2d**: GWAS significant Pleiotropic Discordant sentinel variants within a given genomic locus

*Note:* No : Index of lead SNPs; Genomic Locus : Index of assigned genomic locus matched with "Genomic risk loci" table. Multiple lead SNPs can be assigned to the same genomic locus; uniqID : Unique ID of SNPs consisting of chr:position:allele1:allele2 where alleles are alphabetically ordered; rsID : rsID based on dbSNP build 146; chr : chromosome; pos : position on hg19; P-value : P-value (from the input file); nIndSigSNPs : Number of independent significant SNPs which are in LD with the lead SNP at r2 0.1; IndSigSNPs : rsID of independent significant SNPs which are in LD with the lead SNP at r2 0.1.

**Supplementary Table 3a**: GWAS significant Pleiotropic Genomic Risk Loci for “Dual” Variants

*Note:* Genomic locus : Index of genomic rick loci; uniqID : Unique ID of SNPs consisting of chr:position:allele1:allele2 where alleles are alphabetically ordered; rsID : rsID of the top lead SNP based on dbSNP build 146; chr : chromosome of top lead SNP; pos : position of top lead SNP on hg19; P-value : P-value of top lead SNP (from the input file); start : Start position of the locus; end : End position of the locus; nSNPs : The number of unique candidate SNPs in the genomic locus, including non-GWAS-tagged SNPs (which are available in the user selected reference panel). Candidate SNPs are all SNPs that are in LD (give user-defined r2) with any of independent significant SNPs and either have a P-value below the user defined threshold or are only available in 1000G; nGWASSNPs : The number of unique GWAS-tagged candidate SNPs in the genomic locus which is available in the GWAS summary statistics input file. This is a subset of "nSNPs"; nIndSigSNPs : The number of the independent (at user defined r2) significant SNPs in the genomic locus; IndSigSNPs : rsID of the independent significant SNPs in the genomic locus; nLeadSNPs : The number of lead SNPs in the genomic locus. Lead SNPs are subset of independent significant SNPs at r2 0.1; LeadSNPs : rsID of lead SNPs in the genomic locus.

**Supplementary Table 3b**: Annotations for variants within GWAS significant Pleiotropic Genomic Risk Loci for “Dual” Variants

*Note:* uniqID : Unique ID of SNPs consisting of chr:position:allele1:allele2 where alleles are alphabetically ordered; rsID : rsID based on dbSNP build 146; chr : chromosome; pos : position on hg19; effect\_allele : Effect/risk allele if it is provided in the input GWAS summary statistics file. If not, this is the alternative (minor) allele in 1000G; non\_effect\_allele : Non-effect/non-risk allele if it is provided in the input GWAS summary statistics file. If not, this is the reference (major) allele in 1000G; MAF : Minor allele frequency computed based on 1000G; gwasP : P-value provided in the input GWAS summary statistics file. Non-GWAS tagged SNPs (which do not exist in input file but are extracted from the reference panel) have "NA" instead; or : Odds ratio provided in the input GWAS summary statistics file if available. Non-GWAS tagged SNPs (which do not exist in input file but are extracted from the reference panel) have "NA" instead; beta : Beta provided in the input GWAS summary statistics file if available. Non-GWAS tagged SNPs (which do not exist in input file but are extracted from the reference panel) have "NA" instead; se : Standard error provided in the input GWAS summary statistics file if available. Non-GWAS tagged SNPs (which do not exist in input file but are extracted from the reference panel) have "NA" instead; r2 : The maximum r2 of the SNP with one of the independent significant SNPs; IndSigSNP : rsID of the independent significant SNP which has the maximum r2 with the SNP; Genomic locus : Index of the genomic risk loci matching with "Genomic risk loci" table; Novel : Whether the index SNP/locus is novel to the pleiotropic analysis, and if the variant/locus has been validated in recent education attainment GWAS (information from Supplementary Table 4); nearestGene : The nearest Gene of the SNP based on ANNOVAR annotations. Note that ANNOVAR annotates "consequence" function by prioritizing the most deleterious annotation for SNPs which are locating a genomic region where multiple genes are obverlapped. Genes are ecoded in symbol, if it is available otherwise; Ensembl ID. Genes include all transcripts from Ensembl gene build 85 including non-protein coding genes and RNAs; dist : Distance to the nearest gene. SNPs which are locating in the gene body or 1kb up- or down-stream of TSS or TES have 0; func : Functional consequence of the SNP on the gene obtained from ANNOVAR. For exonic SNPs, detailed annotation (e.g. non-synonymous, stop gain and so on) is available in the ANNOVAR table (annov.txt); CADD : CADD score which is computed based on 63 annotations. The higher the score, the more deleterious the SNP is. 12.37 is the suggested threshold by Kicher et al (2014); RDB : RegulomeDB score which is a categorical score (from 1a to 7). 1a is the highest score for SNPs with the most biological evidence to be a regulatory element; minChrState : The minimum 15-core chromatin state across 127 tissue/cell type; commonChrState : The most common 15-core chromatin state across 127 tissue/cell types; posMapFilt : Whether the SNP was used for eQTL mapping or not. 1 is used, otherwise 0. When eqtl mapping is not performed, all SNPs have 0.

**Supplementary Table 3c:** GWAS significant variants within r2 = 0.6 of “Dual” sentinel variants

*Note:* No : Index of independent significant SNPs; Genomic Locus : Index of assigned genomic locus matched with "Genomic risk loci" table. Multiple independent lead SNPs can be assigned to the same genomic locus; uniqID : Unique ID of SNPs consisting of chr:position:allele1:allele2 where alleles are alphabetically ordered; rsID : rsID based on dbSNP build 146; chr : chromosome; pos : position on hg19; P-value : P-value (from the input file); nSNPs : The number of SNPs in LD with the lead SNP given r2, including non-GWAS-tagged SNPs (which are extracted from 1000G); nGWASSNPs : The number of GWAS-tagged SNPs in LD with the lead SNP given r2. This is a subset of "nSNPs".

**Supplementary Table 3d**: GWAS significant Pleiotropic “Dual” sentinel variants within a given genomic locus

*Note:* No : Index of lead SNPs; Genomic Locus : Index of assigned genomic locus matched with "Genomic risk loci" table. Multiple lead SNPs can be assigned to the same genomic locus; uniqID : Unique ID of SNPs consisting of chr:position:allele1:allele2 where alleles are alphabetically ordered; rsID : rsID based on dbSNP build 146; chr : chromosome; pos : position on hg19; P-value : P-value (from the input file); nIndSigSNPs : Number of independent significant SNPs which are in LD with the lead SNP at r2 0.1; IndSigSNPs : rsID of independent significant SNPs which are in LD with the lead SNP at r2 0.1.

**Supplementary Table 4**: Novel Loci from PLEIO pleiotropic meta-analysis and Validation with recent Education Attainment GWAS

*Note:* Genomic Coordintes/Locus Information: lociID: LocusID in the format chr:start:stop; chrom : chromosome; start : start base pair coordinate of locus; end : end base pair coordinate of locus. Input GWAS: PGC\_SCZ: more recent schizophrenia GWAS accessed from PGC’s website, Edu\_Lee: 2019 data freeze of SSGAC GWAS accessed from SSGAC’s website, LAM\_CTP: cognitive data access via available via COGENT data links available from Lam et al., 2022 Nat Comm manuscript; Education (Validation): Most recent 2023 Education GWAS significant SNPs accessed from SSGAC’s website; further details in Data Availability section; Status: Novel: loci that are not previously reported on in the input GWAS; Novel replicated: loci that are note previously reported on in the input GWAS, but validated in most recent Education Attainment GWAS; Pleiotropy Meta-Analysis: Concordant: output of pleiotropic meta-analysis of concordant variant set; Discordant: output of pleiotropic meta-analysis of discordant variant set; Dual: output of pleiotropic meta-analysis of dual variant set; index variant: index variant within significant genomic locus; pval: p-value of index variant.

**Supplementary Table 5a**: MAGMA Gene Set Analysis for Concordant Variant Set

*Note:* FULL\_NAME: name of gene set; Total\_NumberGenes: total number of genes within each gene set; NGENES: number of genes included from MAGMA gene-based genome-wide analysis; BETA: effect size of MAGMA gene set analysis; BETA\_STD: standard deviation of effect size; SE: standard error of gene set analysis; P: uncorrected p-value; P\_bonferroni\_corr: Bonferroni corrected p-values; P\_sidak\_corr: Sidak corrected p-values; P\_holm-sidak\_corr: Holm-Sidak corrected p-values; P\_fdr\_bh\_corr: Benjamini-Hochberg FDR corrected p-values; P\_holm\_corr: Holm corrected p-values; CommonGenes: gene included within MAGMA gene set analysis results; CommonGenes\_Pvalues: MAGMA gene-based p-values

**Supplementary Table 5b**: MAGMA Gene Set Analysis for Discordant Variant Set

*Note:* FULL\_NAME: name of gene set; Total\_NumberGenes: total number of genes within each gene set; NGENES: number of genes included from MAGMA gene-based genome-wide analysis; BETA: effect size of MAGMA gene set analysis; BETA\_STD: standard deviation of effect size; SE: standard error of gene set analysis; P: uncorrected p-value; P\_bonferroni\_corr: Bonferroni corrected p-values; P\_sidak\_corr: Sidak corrected p-values; P\_holm-sidak\_corr: Holm-Sidak corrected p-values; P\_fdr\_bh\_corr: Benjamini-Hochberg FDR corrected p-values; P\_holm\_corr: Holm corrected p-values; CommonGenes: gene included within MAGMA gene set analysis results; CommonGenes\_Pvalues: MAGMA gene-based p-values

**Supplementary Table 5c**: MAGMA Gene Set Analysis for Dual Variant Set

*Note:* FULL\_NAME: name of gene set; Total\_NumberGenes: total number of genes within each gene set; NGENES: number of genes included from MAGMA gene-based genome-wide analysis; BETA: effect size of MAGMA gene set analysis; BETA\_STD: standard deviation of effect size; SE: standard error of gene set analysis; P: uncorrected p-value; P\_bonferroni\_corr: Bonferroni corrected p-values; P\_sidak\_corr: Sidak corrected p-values; P\_holm-sidak\_corr: Holm-Sidak corrected p-values; P\_fdr\_bh\_corr: Benjamini-Hochberg FDR corrected p-values; P\_holm\_corr: Holm corrected p-values; CommonGenes: gene included within MAGMA gene set analysis results; CommonGenes\_Pvalues: MAGMA gene-based p-values

## Supplementary Figures



**Supplementary Figure 1**. Concordant and Discordant Effect Allele Sets

Top panel: Possible allelic configurations that constitute the Concordant allele variant set. Bottom panel: Possible allele configurations that constitute the Discordant allele variant set.

**Supplementary References**

1. Lam, M., Chen, C.-Y., Hill, W.D., Xia, C., Tian, R., Levey, D.F., Gelernter, J., Stein, M.B., Hatoum, A.S., Huang, H., et al. (2022). Collective genomic segments with differential pleiotropic patterns between cognitive dimensions and psychopathology. Nat. Commun. *13*, 6868.

2. Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T.A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat. Genet. *50*, 1112–1121.

3. Trubetskoy, V., Pardiñas, A.F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T.B., Bryois, J., Chen, C.-Y., Dennison, C.A., Hall, L.S., et al. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature *604*, 502–508.

4. Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S.M., Sidorenko, J., Kweon, H., Goldman, G., Gjorgjieva, T., et al. (2022). Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. Nat. Genet. *54*, 437–449.

5. Skene, N.G., Bryois, J., Bakken, T.E., Breen, G., Crowley, J.J., Gaspar, H.A., Giusti-Rodriguez, P., Hodge, R.D., Miller, J.A., Muñoz-Manchado, A.B., et al. (2018). Genetic identification of brain cell types underlying schizophrenia. Nat. Genet. *50*, 825–833.

6. Lee, C.H., Shi, H., Pasaniuc, B., Eskin, E., and Han, B. (2021). PLEIO: a method to map and interpret pleiotropic loci with GWAS summary statistics. Am. J. Hum. Genet. *108*, 36–48.

7. Bhattacharjee, S., Rajaraman, P., Jacobs, K.B., Wheeler, W.A., Melin, B.S., Hartge, P., GliomaScan Consortium, Yeager, M., Chung, C.C., Chanock, S.J., et al. (2012). A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. Am. J. Hum. Genet. *90*, 821–835.

8. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan, L., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat. Genet. *47*, 1236–1241.

9. Lam, M., Hill, W.D., Trampush, J.W., Yu, J., Knowles, E., Davies, G., Stahl, E., Huckins, L., Liewald, D.C., Djurovic, S., et al. (2019). Pleiotropic Meta-Analysis of Cognition, Education, and Schizophrenia Differentiates Roles of Early Neurodevelopmental and Adult Synaptic Pathways. Am. J. Hum. Genet. *105*, 334–350.

10. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

11. Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium (2015). Corrigendum: Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. Nat. Neurosci. *18*, 926.

12. Sey, N.Y.A., Hu, B., Mah, W., Fauni, H., McAfee, J.C., Rajarajan, P., Brennand, K.J., Akbarian, S., and Won, H. (2020). A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. Nat. Neurosci. *23*, 583–593.

13. Pardiñas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L., et al. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat. Genet. *50*, 381–389.

14. Singh, T., Walters, J.T.R., Johnstone, M., Curtis, D., Suvisaari, J., Torniainen, M., Rees, E., Iyegbe, C., Blackwood, D., McIntosh, A.M., et al. (2017). The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. Nat. Genet. *49*, 1167–1173.

15. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. *57*, 289–300.

16. BrainSpan (2014). Brain Span.

17. Miller, J.A., Ding, S.-L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K., et al. (2014). Transcriptional landscape of the prenatal human brain. Nature *508*, 199–206.