# Genetic Architecture of Educational Attainment is Represented by Executive Function in Children

Joanne S. M. Chia[1], Evelyn C. Law[1,2,3], Chong Yap Seng[1,4], Peter D. Gluckman[1, 5], Johan G. Eriksson[1,4,6,7,8], Lourdes M. Daniel[9], Patrícia P. Silveira[2,10,11], Michael J. Meaney[11], Michelle Z. L. Kee[1,6*]

[*]**Corresponding author**

## Content of Online Supplementary Materials

**Supplementary Methods**

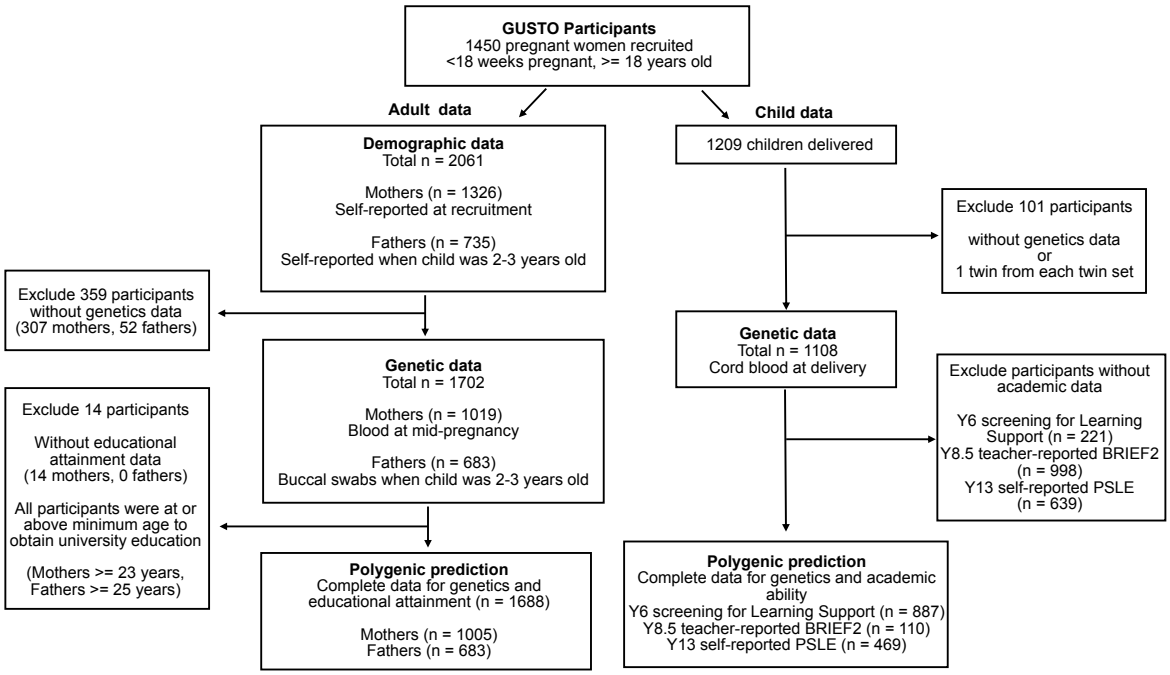
Calculation of LDpred2 polygenic scores

We derived Bayesian polygenic scores using LDpred2[1,2] for both GUSTO parents and children, based on the summary statistics for educational attainment[3], as well as the cognitive and non-cognitive components[4] of educational attainment. Linkage disequilibrium (LD) information is estimated from an independent reference data, specifically the HapMap3+[5] which includes 1,444,196 SNPs from participants of European ancestry in the UK Biobank. Genotype data were mean imputed using the "snp_fastImputeSimple" function of bigsnpr v1.12.18[6] to ensure good overlap with the LD reference, and duplicated variants were removed.

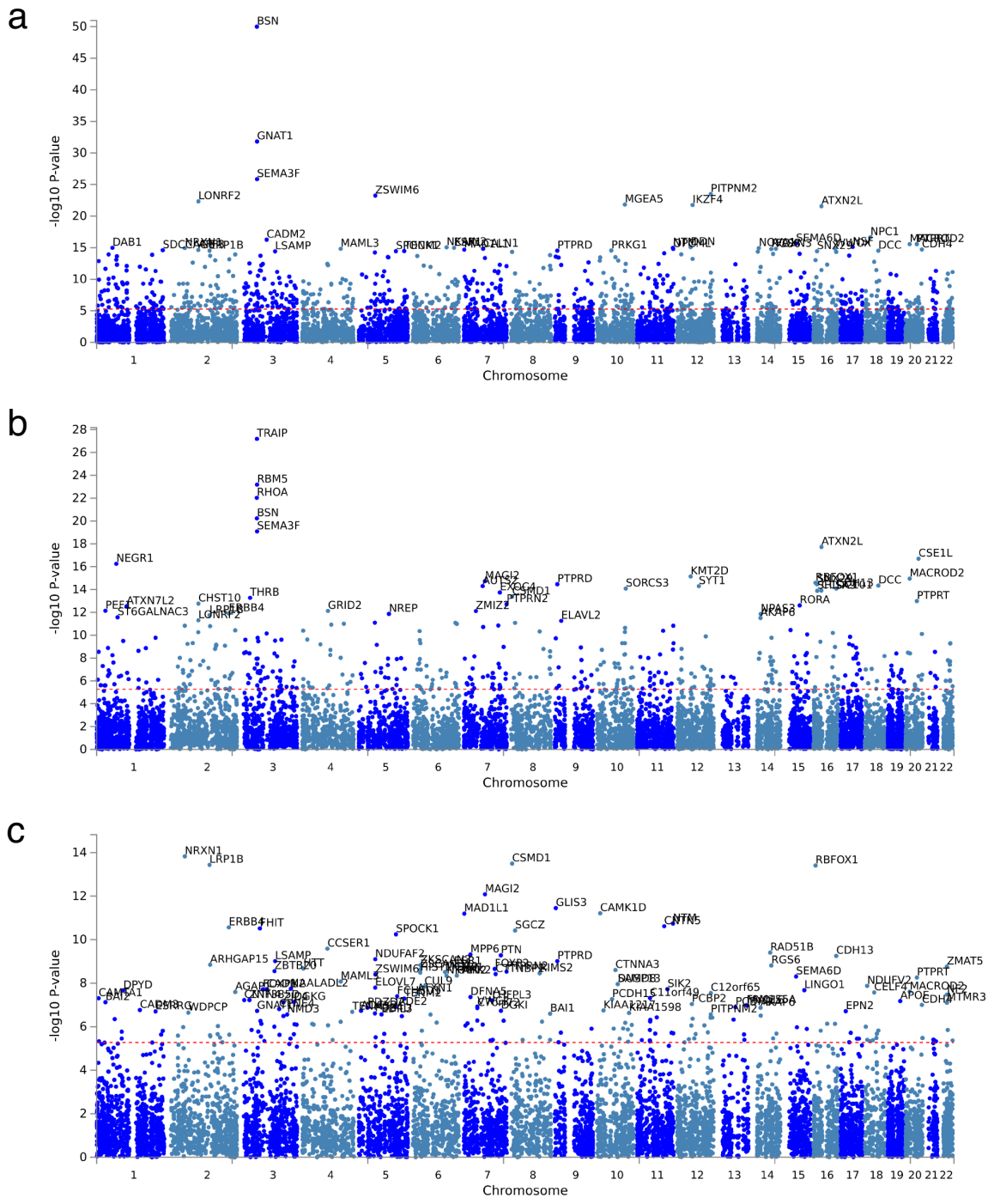
We followed the recommended quality control measures for the summary statistics, as misspecifications with the LD reference can lead to inaccurate predictive power[7]. Summary statistics were munged using the MungeSumstats package[8] to ensure standardized GWAS formatting (e.g., format used, missing data/columns, essential headers present, removal of ambiguous or biallelic alleles, allele flipping where necessary). Other quality control steps performed include discarding variants with MAF < 0.01, verifying alleles orientation, and ensuring that standard deviations of genotypes between summary statistics and LD reference were consistent.

Calculation of the child Primary School Leaving Examination scores

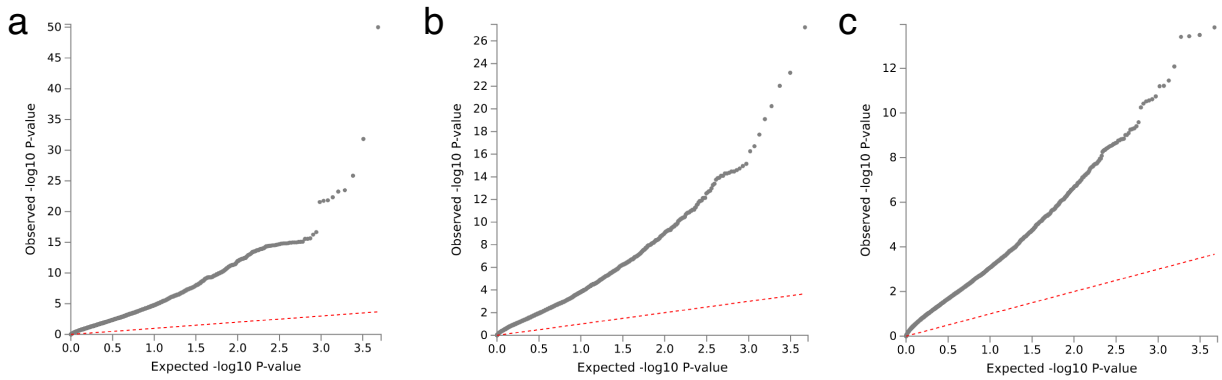
We use the child-reported Primary School Leaving Examination (PSLE) scores ($n$ = 469) as a measure of child academic achievement. The PSLE is a standardized examination taken by Singaporean children at the end of Primary School when they were 12 years old. Scores are calculated by summing 4 subject scores (English, Mathematics, Science and Mother Tongue Language), where lower scores represent a better grade, ranging from 4 - 32.



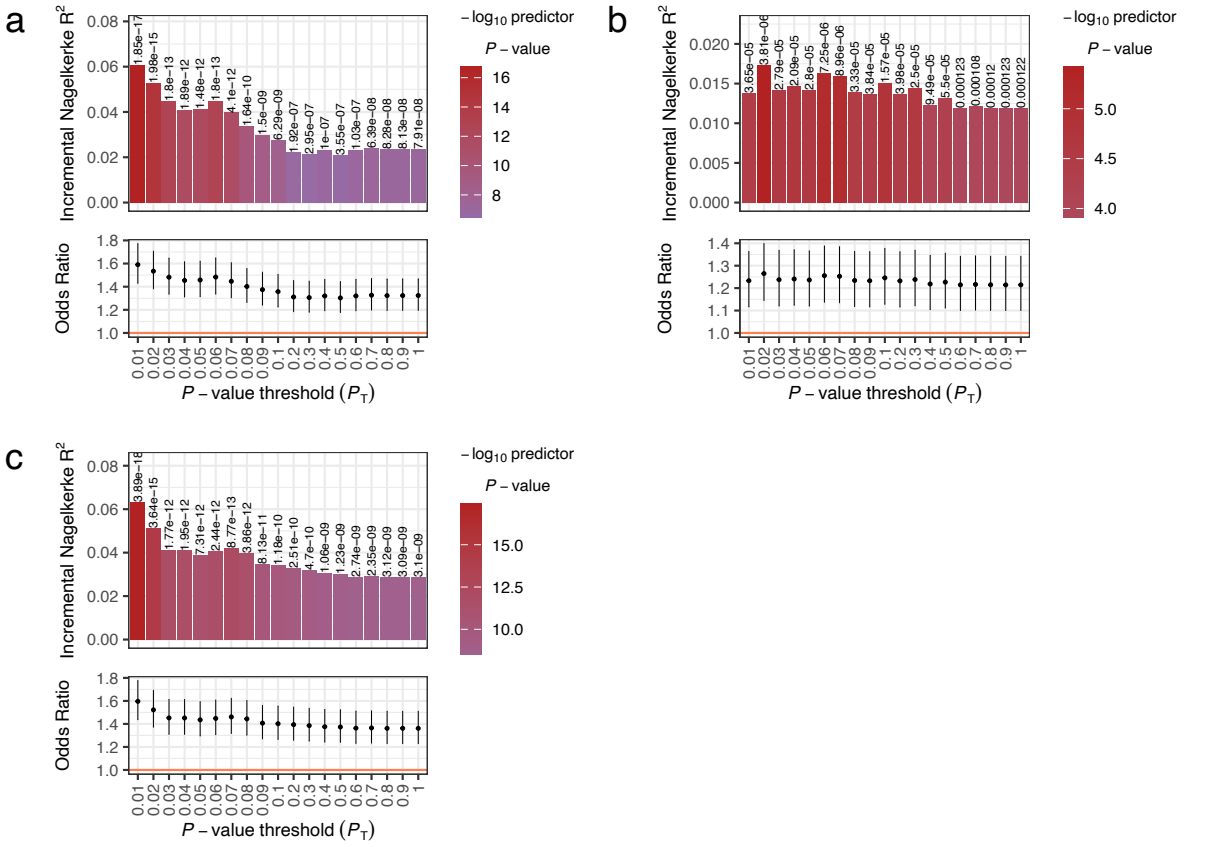

**Supplementary Figure S1. Study flowchart and analysis numbers.** Flow chart representing the number of adult and child participants from the GUSTO cohort with both genetic and academic measures used for analyses in our study. Y6, Y8.5 and Y13 refer to data collected when the child turned 6, 8.5, and 13 years old. GUSTO refers to Growing Up in Singapore Towards healthy Outcomes; T-BRIEF2 refers to Teacher-reported Behaviour Rating Inventory of Executive Function 2, and PSLE refers to the national Primary School Leaving Examination in Singapore.



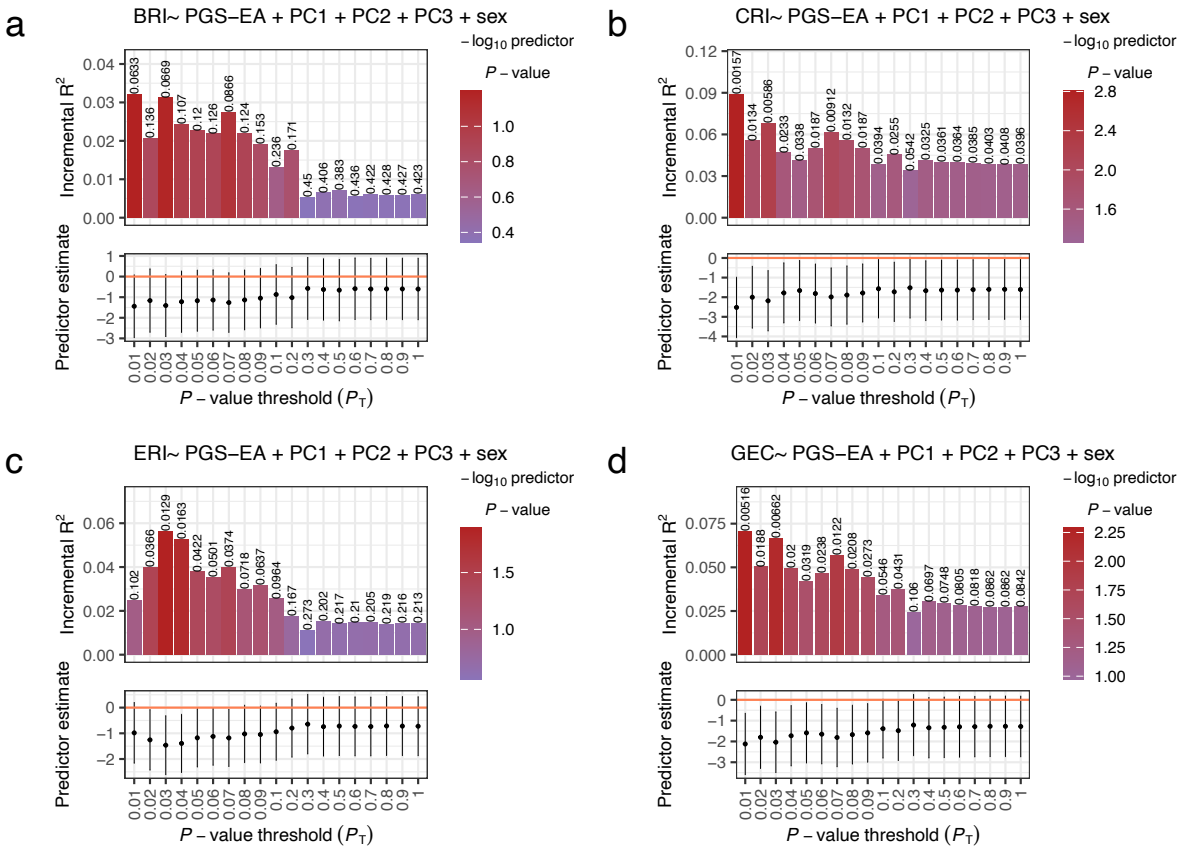

**Supplementary Figure S2**. **Manhattan plots of SNPs found both in the GUSTO cohort and the GWAS** for (a) educational attainment[3], (b) cognitive component of educational attainment[4], and (c) non-cognitive component of educational attainment[4]. The red dashed line represents the gene-based significance threshold (0.05/number of genes) based on the number of genes ($PGS_{EA}$: 9649; $PGS_{Cog}$: 9378; $PGS_{nonCog}$: 9373) that were mapped from the input SNPs.



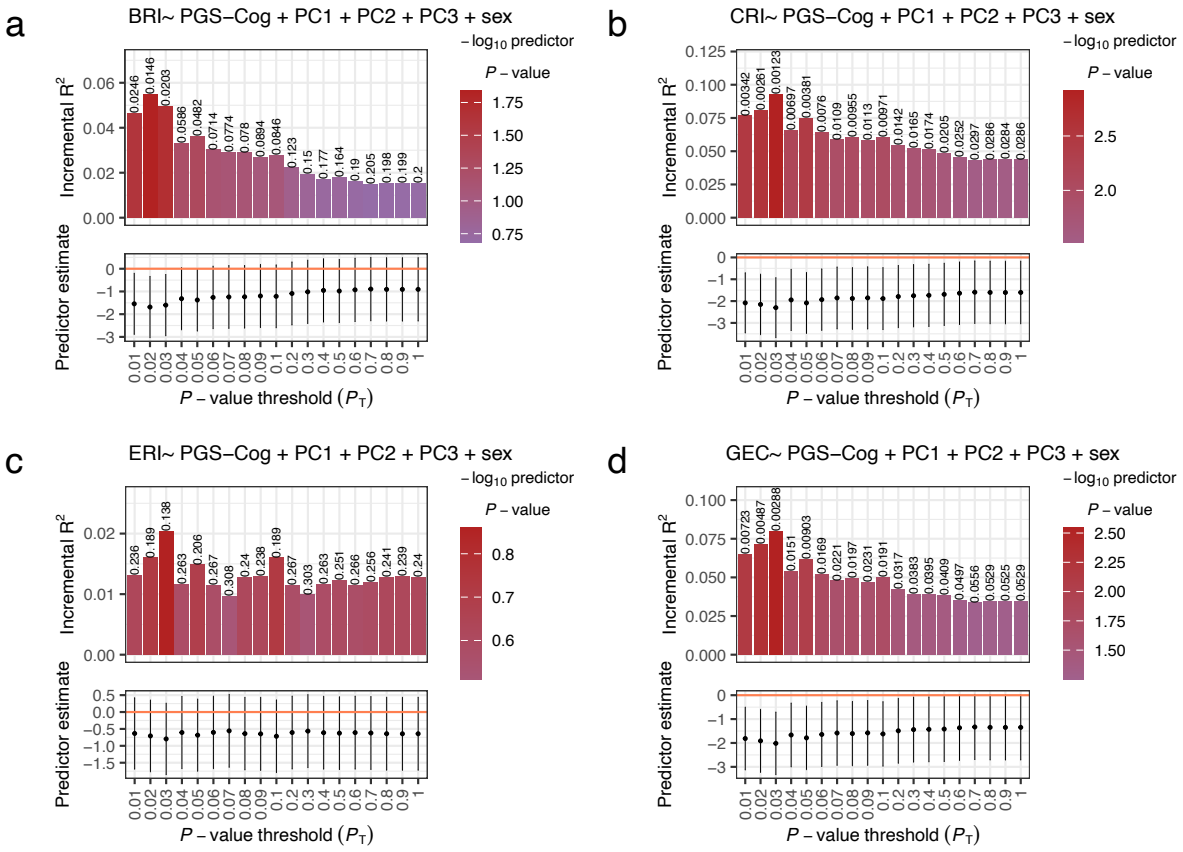

**Supplementary Figure S3**. **Q-Q plots of SNPs present both in the GUSTO cohort and the GWAS** for (a) educational attainment[3], (b) cognitive[4], and (c) non-cognitive components of educational attainment[4].



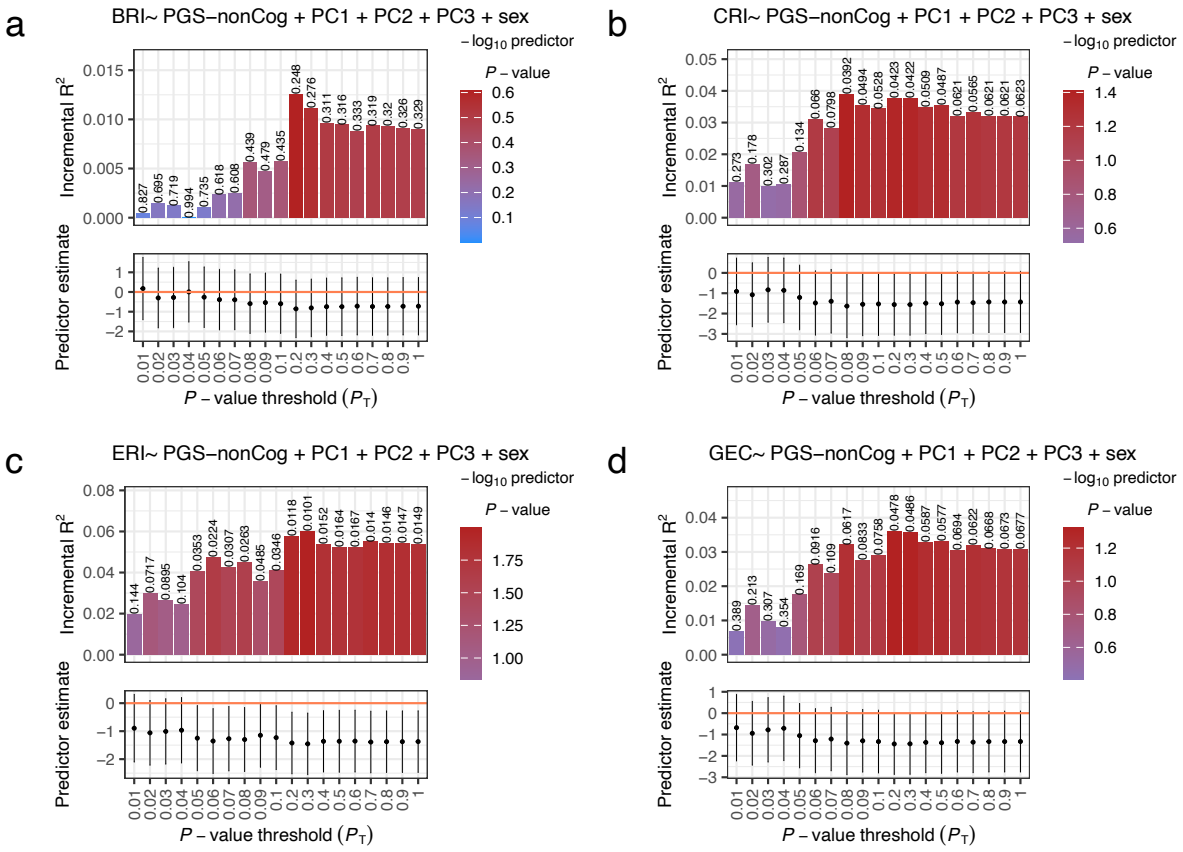

**Supplementary Figure S4. PGS prediction of university completion in adults.** GUSTO parents (a) $PGS_{EA}$, (b) $PGS_{Cog}$, and (c) $PGS_{nonCog}$ were used to predict university completion, using the first three genetic PCs and sex as covariates. Bars were plotted to represent incremental $R^2$, labelled with predictor *p*-value and coloured according to the -$\log_{10}$ *p*-value of the predictor across pTs (top). The odds ratio and 95% confidence intervals were plotted across pTs (bottom). The PGS at the pT yielding the highest incremental $R^2$ value was selected for subsequent analysis.



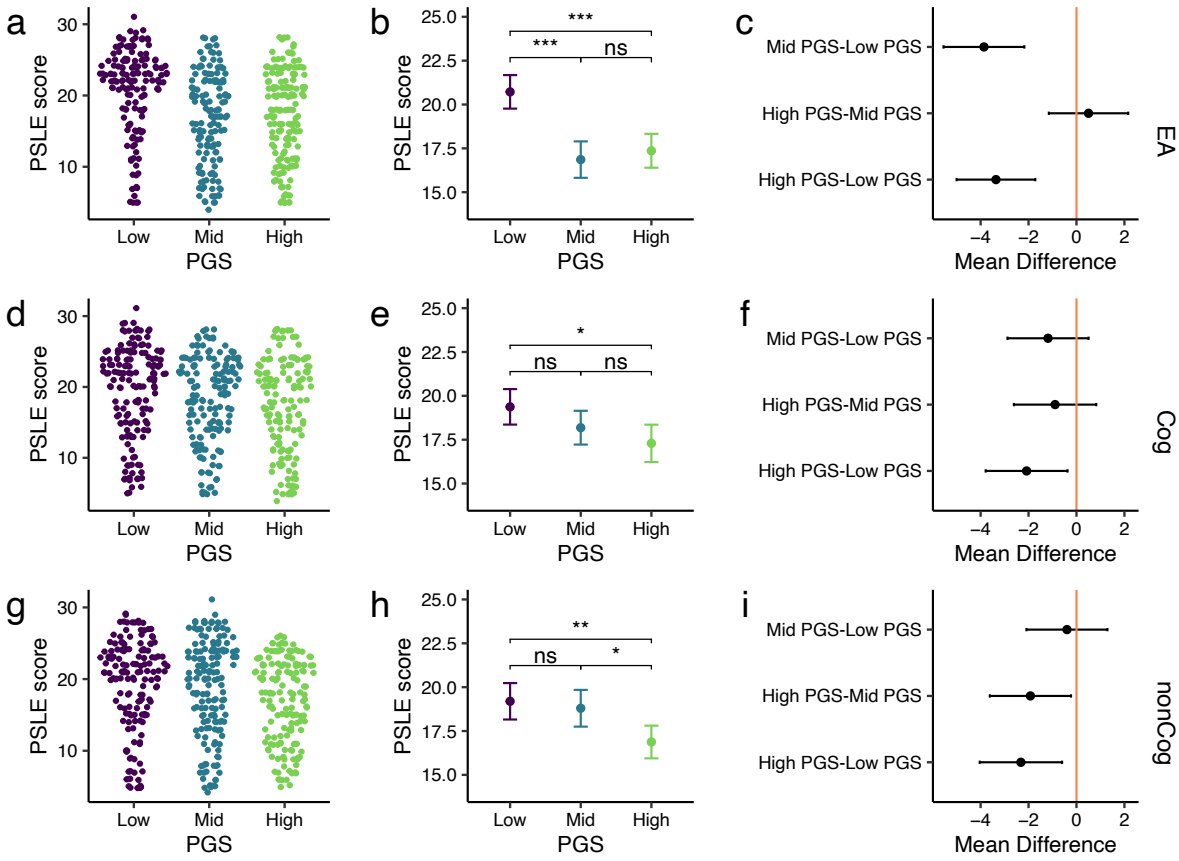

**Supplementary Figure S5. Child $PGS_{EA}$ predicting executive function outcomes of T-BRIEF2.** Child $PGS_{EA}$ was used to predict the (a) behavioral regulation (BRI), (b) cognitive regulation (CRI), (c) emotional regulation (ERI), and (d) general executive composite (GEC) indices of the T-BRIEF2, with the first three genetic PCs and child sex as covariates. Bars were plotted to represent incremental $R^2$, labelled with predictor *p*-value and coloured according to the $-\log_{10}$ *p*-value of the predictor across pTs (top). The odds ratio and 95% confidence intervals were plotted across pTs (bottom). The PGS at the pT = 0.03 yielded the highest incremental $R^2$ value and was selected for subsequent analysis. A higher T-BRIEF2 score is indicative of more problems with executive functioning.



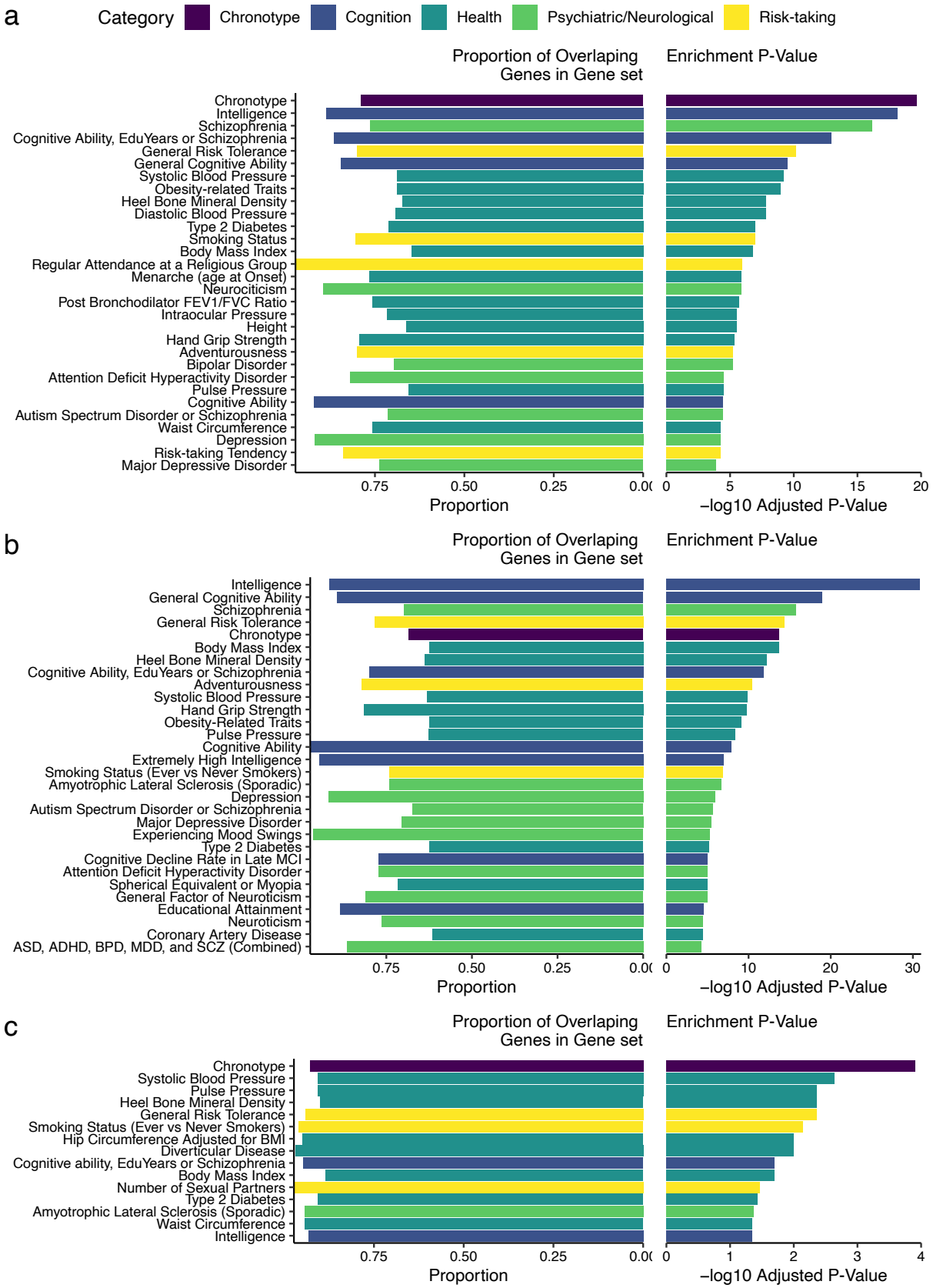

**Supplementary Figure S6. Child $PGS_{Cog}$ predicting executive function outcomes of T-BRIEF2.** Child $PGS_{Cog}$ was used to predict (a) behavioral regulation (BRI), (b) cognitive regulation (CRI), (c) emotional regulation (ERI), and (d) general executive composite (GEC) indices of the T-BRIEF2, with the first three genetic PCs and child sex as covariates. Bars were plotted to represent incremental $R^2$, labelled with predictor *p*-value and coloured according to the $-\log_{10}$ *p*-value of the predictor across pTs (top). The odds ratio and 95% confidence intervals were plotted across pTs (bottom). The PGS at the pT = 0.03 yielded the highest $R^2$ and the lowest *p*-value across indices and was used for downstream gene set analyses. A higher T-BRIEF2 score is indicative of more problems with executive functioning.




**Supplementary Figure S7. Child $PGS_{nonCog}$ predicting executive function outcomes of T-BRIEF2.** Child $PGS_{nonCog}$ was used to predict (a) behavioral regulation (BRI), (b) cognitive regulation (CRI), (c) emotional regulation (ERI), and (d) general executive composite (GEC) indices of the T-BRIEF2, with the first three genetic PCs and child sex as covariates. Bars were plotted to represent incremental $R^2$, labelled with predictor *p*-value and coloured according to the $-\log_{10}$ *p*-value of the predictor across pTs (top). The odds ratio and 95% confidence intervals were plotted across pTs (bottom). The PGS at the pT = 0.3 had the highest $R^2$ and lowest *p*-value across indices and was used for downstream gene set analyses. A higher T-BRIEF2 score is indicative of more problems with executive functioning.

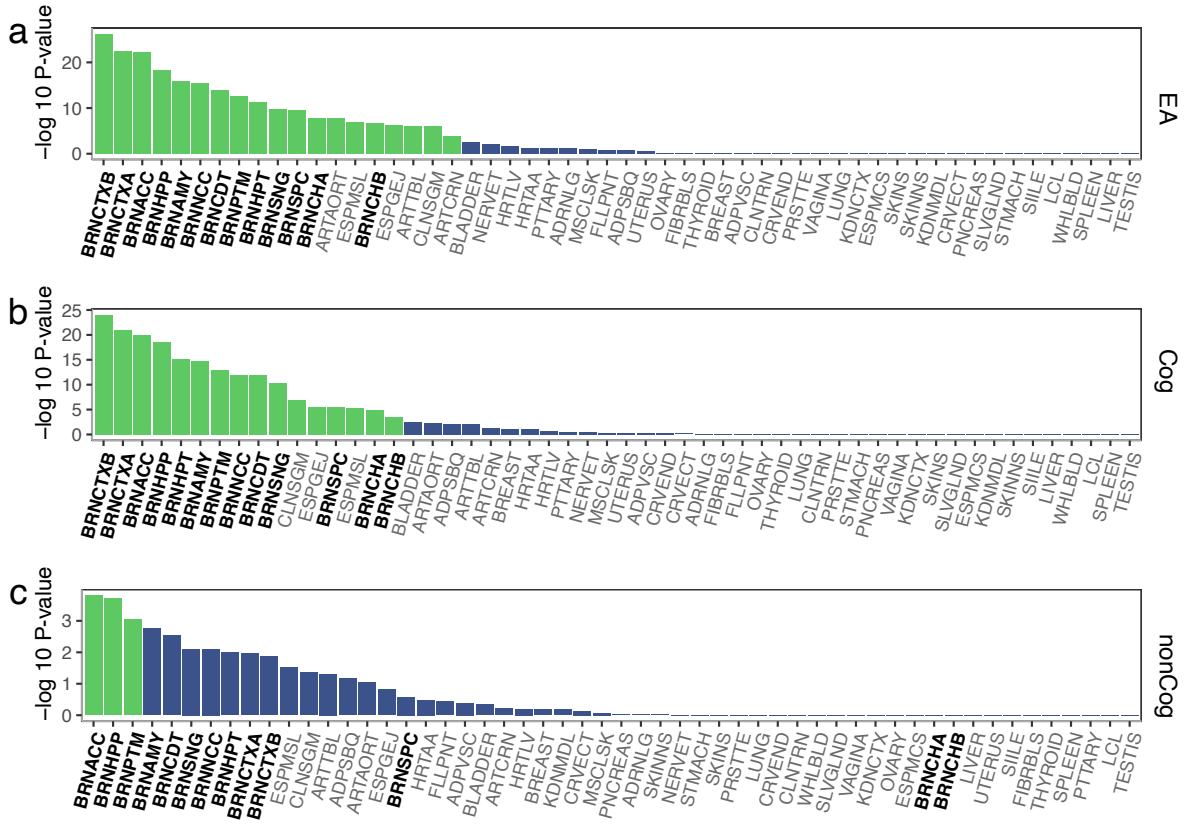



**Supplementary Figure S8. Child PGS predicting Primary School Leaving Examination scores.** Child $PGS_{EA}$, $PGS_{Cog}$, and $PGS_{nonCog}$ was used as a predictor for Primary School Leaving Examination (PSLE) scores as an outcome, with the first three genetic PCs and sex as covariates. Lower PSLE scores indicate better academic performance. PGS scores were divided into the groups "High", "Mid" and "Low" based on tertiles and plotted against PSLE scores (a, d, g). The mean PSLE score of the PGS groups were plotted (with 95% confidence intervals) and an analysis of variance followed by Tukey post-hoc test was conducted to determine if the means were statistically significantly different (b, e, h). The mean difference (with 95% confidence intervals) were plotted against the contrasts used in the Tukey post-hoc tests (c, f, i).



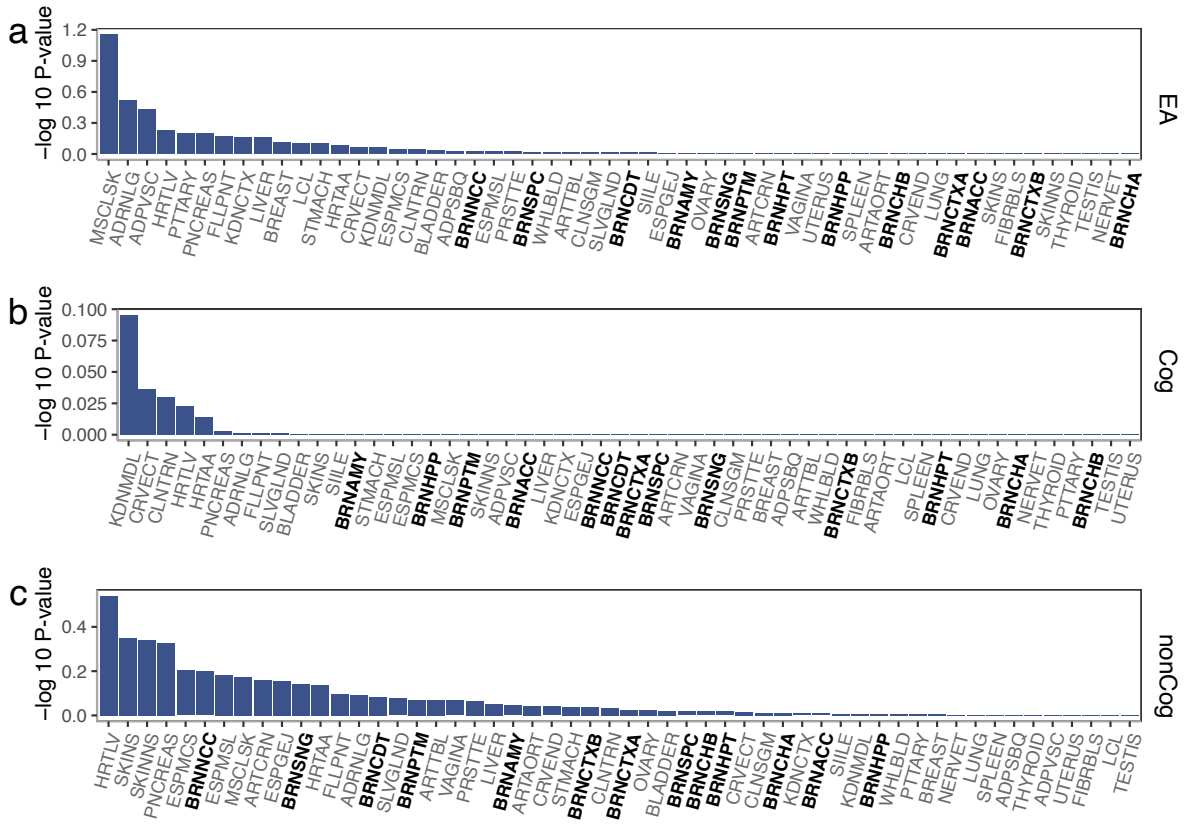

**Supplementary Figure S9. Overlap of gene sets from $PGS_{EA}$, $PGS_{Cog}$ and $PGS_{nonCog}$ with GWAS of other traits.** Gene sets derived from the (a) $PGS_{EA}$, (b) $PGS_{Cog}$, and (c) $PGS_{nonCog}$ were analysed for

the proportion of overlap with genes from GWASs for other common traits. Traits were classified under the categories of chronotype, cognition, health, psychological/neurological and risk-taking.



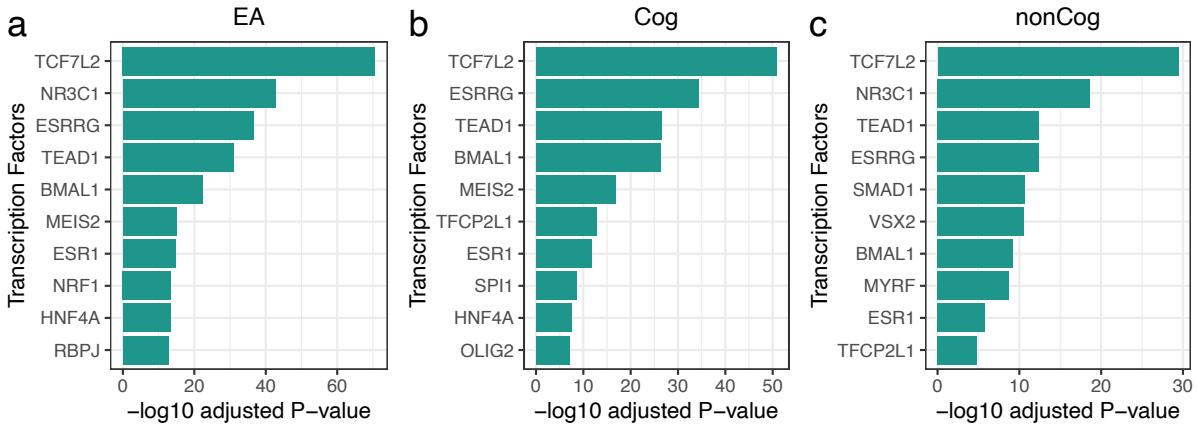

| GTEx Official Tissue Name | Abbreviation | GTEx Official Tissue Name | Abbreviation |
|---|---|---|---|
| Adipose – Subcutaneous | ADPSBQ | Esophagus – Gastroesophageal junction | ESPGEJ |
| Adipose – Visceral (Omentum) | ADPVSC | Esophagus – Mucosa | ESPMCS |
| Adrenal Gland | ADRNLG | Esophagus – Muscularis | ESPMSL |
| Artery – Aorta | ARTAORT | Fallopian Tube | FLLPNT |
| Artery – Coronary | ARTCRN | Heart – Atrial appendage | HRTAA |
| Artery – Tibial | ARTTBL | Heart – Left ventricle | HRTLV |
| Bladder | BLADDER | Kidney – Cortex | KDNCTX |
| Brain – Amygdala | BRNAMY | Kidney – Medulla | KDNMDL |
| Brain – Anterior cingulate cortex (BA24) | BRNACC | Liver | LIVER |
| Brain – Caudate (basal ganglia) | BRNCDT | Lung | LUNG |
| Brain – Cerebellar (Hemisphere) | BRNCHB | Minor salivary gland | SLVGLND |
| Brain – Cerebellum | BRNCHA | Muscle – Skeletal | MSCLSK |
| Brain – Cortex | BRNCTXA | Nerve – Tibial | NERVET |
| Brain – Fronal cortex (BA9) | BRNCTXB | Ovary | OVARY |
| Brain – Hippocampus | BRNHPP | Pancreas | PNCREAS |
| Brain – Hypothalamus | BRNHPT | Pituitary | PTTARY |
| Brain – Nucleus accumbens (basal ganglia) | BRNNCC | Prostate | PRSTTE |
| Brain – Putamen (basal ganlia) | BRNPTM | Skin – Not sun exposed (subrapubic) | SKINNS |
| Brain – Spinal cord (cervical c–1) | BRNSPC | Skin – Sun exposed (lower leg) | SKINS |
| Brain – Substantia nigra | BRNSNG | Small intestine – Terminal ileum | SIILE |
| Breast – Mammary tissue | BREAST | Spleen | SPLEEN |
| Cells – Cultured fibroblasts | FIBRBLS | Stomach | STMACH |
| Cells – EBV–transformed lymphocytes | LCL | Testis | TESTIS |
| Cervix – Ectocervix | CRVECT | Thyroid | THYROID |
| Cervix – Endocervix | CRVEND | Uterus | UTERUS |
| Colon – Sigmoid | CLNSGM | Vagina | VAGINA |
| Colon – Transverse | CLNTRN | Whole blood | WHLBLD |

**Supplementary Figure S10. Tissue specificity of genes associated with PGSs.** The genes derived from (a) $PGS_{EA}$, (b) $PGS_{Cog}$, and (c) $PGS_{nonCog}$ were analysed for enriched expression in a particular tissue compared to other tissue types based on RNA-seq data from the Genotype-Tissue Expression project (GTEx) and were plotted using FUMA software. Bold labels on the x-axis represent brain tissue. Green bars represent enriched tissue expressions that passed Bonferroni correction.

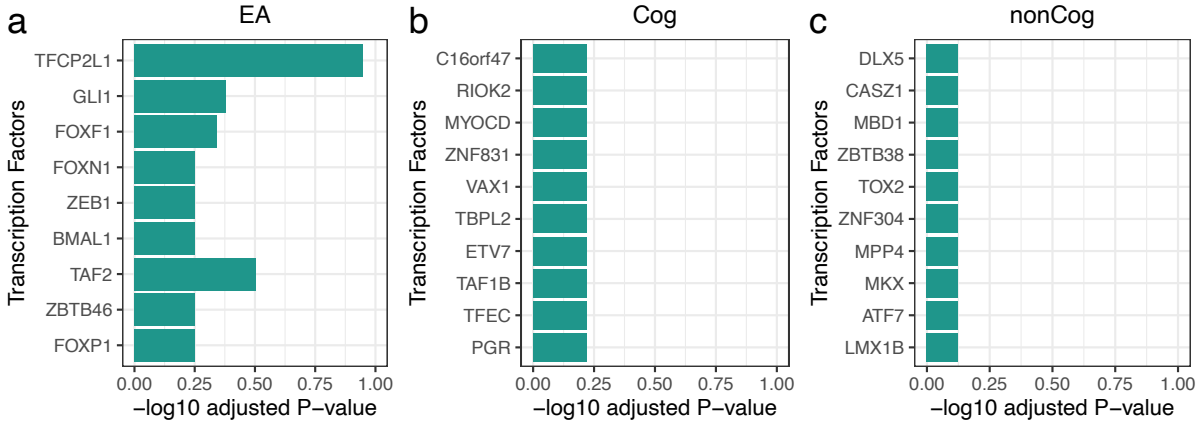



| GTEx Official Tissue Name | Abbreviation | GTEx Official Tissue Name | Abbreviation |
|---|---|---|---|
| Adipose – Subcutaneous | ADPSBQ | Esophagus – Gastroesophageal junction | ESPGEJ |
| Adipose – Visceral (Omentum) | ADPVSC | Esophagus – Mucosa | ESPMCS |
| Adrenal Gland | ADRNLG | Esophagus – Muscularis | ESPMSL |
| Artery – Aorta | ARTAORT | Fallopian Tube | FLLPNT |
| Artery – Coronary | ARTCRN | Heart – Atrial appendage | HRTAA |
| Artery – Tibial | ARTTBL | Heart – Left ventricle | HRTLV |
| Bladder | BLADDER | Kidney – Cortex | KDNCTX |
| Brain – Amygdala | BRNAMY | Kidney – Medulla | KDNMDL |
| Brain – Anterior cingulate cortex (BA24) | BRNACC | Liver | LIVER |
| Brain – Caudate (basal ganglia) | BRNCDT | Lung | LUNG |
| Brain – Cerebellar (Hemisphere) | BRNCHB | Minor salivary gland | SLVGLND |
| Brain – Cerebellum | BRNCHA | Muscle – Skeletal | MSCLSK |
| Brain – Cortex | BRNCTXA | Nerve – Tibial | NERVET |
| Brain – Fronal cortex (BA9) | BRNCTXB | Ovary | OVARY |
| Brain – Hippocampus | BRNHPP | Pancreas | PNCREAS |
| Brain – Hypothalamus | BRNHPT | Pituitary | PTTARY |
| Brain – Nucleus accumbens (basal ganglia) | BRNNCC | Prostate | PRSTTE |
| Brain – Putamen (basal ganlia) | BRNPTM | Skin – Not sun exposed (subrapubic) | SKINNS |
| Brain – Spinal cord (cervical c–1) | BRNSPC | Skin – Sun exposed (lower leg) | SKINS |
| Brain – Substantia nigra | BRNSNG | Small intestine – Terminal ileum | SIILE |
| Breast – Mammary tissue | BREAST | Spleen | SPLEEN |
| Cells – Cultured fibroblasts | FIBRBLS | Stomach | STMACH |
| Cells – EBV–transformed lymphocytes | LCL | Testis | TESTIS |
| Cervix – Ectocervix | CRVECT | Thyroid | THYROID |
| Cervix – Endocervix | CRVEND | Uterus | UTERUS |
| Colon – Sigmoid | CLNSGM | Vagina | VAGINA |
| Colon – Transverse | CLNTRN | Whole blood | WHLBLD |

**Supplementary Figure S11. Tissue specificity of control gene sets.** For each PGS, a control gene set was created by randomly selecting an equivalent number of genes at the most significant pT from the full set of GWAS genes (pT = 1). Differentially upregulated genes were identified in specific tissues compared to others using GTEx RNA-seq data and visualized using FUMA software. Bold labels on the x-axis represent brain tissue. None of the associations passed the Bonferroni correction.




**Supplementary Figure S12. Top 10 transcription factors enriched in gene sets** from (a) $PGS_{EA}$, (b) $PGS_{Cog}$, and (c) $PGS_{nonCog}$ ranked by FDR-adjusted *p*-value in each set.




**Supplementary Figure S13. Top 10 transcription factors enriched in control gene sets** from (a) $PGS_{EA}$, (b) $PGS_{Cog}$, and (c) $PGS_{nonCog}$. An equivalent number of genes at the most significant pTs in each PGS score were randomly selected from the full set of genes in the GWAS (pT = 1) as control gene sets. None of the top 10 transcription factors passes FDR correction.

**Supplementary Table 1. Comparison of polygenic scores obtained using LDpred2 and the optimal p-value threshold from PRSice2**

| | Effect size (95% CI) | *p*-value | Incremental $R^2$ | Correlation (95% CI) |
|---|---|---|---|---|
| **$PGS_{EA}$ (Adult)** | | | | |
| LDpred2 | 1.9 (1.69, 2.12) | 1.16E-28 | 0.108 | 0.28 (0.237, 0.325) |
| PRSice pT 0.01 | 1.6 (1.43, 1.77) | 1.85E-17 | 0.061 | 0.21 (0.165, 0.256) |
| **$PGS_{EA}$ (Child)** | | | | |
| LDpred2 | -1.63 (-2.21, -1.05) | 5.76E-08 | 0.054 | -0.0074 (-0.101, 0.086) |
| PRSice2 pT 0.01 | -0.985 (-1.56, -0.413) | 7.79E-04 | 0.021 | -0.0591 (-0.152, 0.344) |
| **$PGS_{Cog}$ (Adult)** | | | | |
| LDpred2 | 1.4 (1.28, 1.57) | 1.46E-11 | 0.038 | 0.17 (0.12, 0.213) |
| PRSice2 pT 0.02 | 1.3 (1.15, 1.4) | 3.81E-06 | 0.017 | 0.11 (0.065, 0.159) |
| **$PGS_{Cog}$ (Child)** | | | | |
| LDpred2 | -0.894 (-1.45, -0.338) | 0.0017 | 0.019 | 0.051 (-0.0428, 0.143) |
| PRSice2 pT 0.02 | -0.809 (-1.36, -0.261) | 0.0039 | 0.016 | 0.061 (-0.033, 0.153) |
| **$PGS_{nonCog}$ (Adult)** | | | | |
| LDpred2 | 1.6 (1.43, 1.77) | 2.09E-17 | 0.061 | 0.21 (0.165, 0.256) |
| PRSice2 pT 0.01 | 1.6 (1.44, 1.78) | 3.89E-18 | 0.063 | 0.22 (0.171, 0.262) |
| **$PGS_{nonCog}$ (Child)** | | | | |
| LDpred2 | -0.665 (-1.27, -0.0607) | 0.031 | 0.0088 | -0.0499 (-0.143, 0.0437) |
| PRSice2 pT 0.01 | -0.666 (-1.24, -0.096) | 0.022 | 0.0099 | -0.0627 (-0.155, 0.0308) |

**Supplementary Table 2. Associations of child PGSs with T-BRIEF2**

| T-BRIEF2 | $PGS_{EA}$ | | | $PGS_{Cog}$ | | | $PGS_{nonCog}$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | β | 95% CI | *p* | β | 95% CI | *p* | β | 95% CI | *p* |
| **Indices** | | | | | | | | | |
| GEC | **−2.04** | **−3.50, −0.58** | **.013** | **−2.02** | **−3.32, −0.71** | **.006** | −1.44 | −2.86, −0.01 | 0.065 |
| CRI | **−2.18** | **−3.72 , −0.64** | **.013** | **−2.30** | **−3.67, −0.93** | **.005** | −1.56 | −3.07, −0.06 | 0.065 |
| ERI | **−1.46** | **−2.61, −0.32** | **.017** | −0.80 | −0.85, 0.26 | .138 | **−1.45** | **−2.55, −035** | **0.040** |
| BRI | −1.40 | −2.91, 0.10 | .067 | **−1.60** | **−2.95, −0.26** | **.027** | −0.81 | −2.27, 0.65 | 0.276 |
| **CRI** | | | | | | | | | |
| Working memory | **−2.51** | **−4.21, −0.81** | **0.022** | **−2.90** | **−4.40, −1.39** | **0.0002** | - | - | - |
| Task Monitor | **−2.40** | **−4.22, −0.58** | **0.026** | **−2.17** | **−3.82, −0.53** | **0.015** | - | - | - |
| Organization of materials | **−1.53** | **−2.77, −0.29** | **0.027** | **−1.41** | **−2.53, −0.29** | **0.015** | - | - | - |
| Plan/organize | −1.32 | −2.74, 0.01 | 0.067 | **−1.72** | **−2.98, −0.46** | **0.015** | - | - | - |
| Initiate | −1.42 | −2.85, 0.003 | 0.063 | **−1.59** | **−2.87, −0.31** | **0.015** | - | - | - |
| **ERI** | | | | | | | | | |
| Shift | **−1.65** | **−3.05, −0.26** | **0.041** | - | - | - | **−1.59** | **−2.93, −0.25** | **0.040** |
| Emotional control | −1.05 | −2.17, 0.06 | 0.065 | - | - | - | −1.05 | −2.12, 0.02 | 0.056 |
| **BRI** | | | | | | | | | |
| Self-monitor | - | - | - | **−1.80** | **−3.16, −0.44** | **0.020** | - | - | - |
| Inhibit | - | - | - | −1.22 | −2.51, 0.07 | 0.063 | - | - | - |

*p*-values reported have been corrected for multiple testing using the Benjamini-Hochberg method. Bold numbers represent multiple testing corrected models that pass the significance level of .05.

**Supplementary Table S3. Top 15 enriched process networks from $PGS_{EA}$**

| **Process network** | ***p*-value** | **FDR** |
|---|---|---|
| Cell adhesion_Synaptic contact | 1.09E-10 | 1.73E-08 |
| Cell adhesion_Cadherins | 1.15E-06 | 9.14E-05 |
| Development_Neurogenesis_Synaptogenesis | 2.52E-06 | 1.33E-04 |
| Development_Neurogenesis_Axonal guidance | 1.18E-05 | 4.69E-04 |
| Neurophysiological process_Transmission of nerve impulse | 4.61E-05 | 1.47E-03 |
| Cell adhesion_Attractive and repulsive receptors | 2.81E-04 | 7.44E-03 |
| Cytoskeleton_Regulation of cytoskeleton rearrangement | 7.21E-04 | 1.64E-02 |
| Cell adhesion_Cell junctions | 8.37E-04 | 1.66E-02 |
| Transport_Potassium transport | 1.44E-03 | 2.54E-02 |
| Reproduction_GnRH signaling pathway | 2.33E-03 | 3.70E-02 |
| Cytoskeleton_Actin filaments | 2.94E-03 | 4.25E-02 |
| Neurophysiological process_Long-term potentiation | 4.13E-03 | 5.47E-02 |
| Transport_Calcium transport | 5.65E-03 | 6.12E-02 |
| Inflammation_IgE signaling | 5.68E-03 | 6.12E-02 |
| Muscle contraction_Nitric oxide signaling in the cardiovascular system | 5.77E-03 | 6.12E-02 |

**Supplementary Table S4. Top 15 enriched process networks from $PGS_{Cog}$**

| **Process network** | ***p*-value** | **FDR** |
|---|---|---|
| Cell adhesion_Synaptic contact | 4.66E-06 | 7.42E-04 |
| Development_Neurogenesis_Axonal guidance | 3.35E-05 | 2.66E-03 |
| Development_Neurogenesis_Synaptogenesis | 7.47E-05 | 3.57E-03 |
| Cell adhesion_Cadherins | 8.99E-05 | 3.57E-03 |
| Neurophysiological process_GABAergic neurotransmission | 3.36E-04 | 1.07E-02 |
| Cell adhesion_Attractive and repulsive receptors | 6.18E-04 | 1.64E-02 |
| Development_Neurogenesis in general | 4.17E-03 | 9.47E-02 |
| Muscle contraction_Nitric oside signaling in the cardiovascular system | 5.18E-03 | 9.49E-02 |
| Signal transduction_Oxytocin signaling | 5.37E-03 | 9.49E-02 |
| Reproduction_Gonadotropin regulation | 7.53E-03 | 1.14E-01 |
| Reproduction_GnRH signaling pathway | 7.88E-03 | 1.14E-01 |
| Reproduction_Progesterone signaling | 9.12E-03 | 1.21E-01 |
| Transport_Synaptic vesicle exocytosis | 1.27E-02 | 1.51E-01 |
| Cardiac development_Wnt_beta-catenin, Notch, VEGF, IP3 and integrin signaling | 1.33E-02 | 1.51E-01 |
| Transport_Potassium transport | 1.58E-02 | 1.68E-01 |

**Supplementary Table S5. Top 15 enriched process networks from $PGS_{nonCog}$**

| **Process Networks** | ***p*-value** | **FDR** |
|---|---|---|
| Cell adhesion_Synaptic contact | 1.82E-04 | 2.90E-02 |
| Cell adhesion_Cadherins | 6.43E-04 | 5.11E-02 |
| Transport_Calcium transport | 2.59E-03 | 1.37E-01 |
| Neurophysiological process_GABAergic neurotransmission | 4.82E-03 | 1.92E-01 |
| Development_Neurogenesis in general | 6.29E-03 | 2.00E-01 |
| Development_Neurogenesis_Synaptogenesis | 1.14E-02 | 2.58E-01 |
| Development_Neurogenesis_Axonal guidance | 1.16E-02 | 2.58E-01 |
| Transport_Synaptic vesicle exocytosis | 1.37E-02 | 2.58E-01 |
| Cell adhesion_Attractive and repulsive receptors | 1.46E-02 | 2.58E-01 |
| Neurophysiological process_Transmission of nerve impulse | 2.36E-02 | 3.75E-01 |
| Reproduction_GnRH signaling pathway | 2.71E-02 | 3.92E-01 |
| Transport_Sodium transport | 4.36E-02 | 5.78E-01 |
| Cell adhesion_Amyloid proteins | 5.05E-02 | 5.96E-01 |
| Cell adhesion_Cell junctions | 5.25E-02 | 5.96E-01 |
| Cytoskeleton_Regulation of cytoskeleton rearrangement | 6.06E-02 | 6.43E-01 |

**Supplementary Table S6: Top 15 enriched process networks from $PGS_{EA}$ (controls)**

| **Process network** | ***p*-value** | **FDR** |
|---|---|---|
| Transcription_Transcription by RNA polymerase II | 5.31E-03 | 5.16E-01 |
| Cell cycle_G1-S | 7.80E-03 | 5.16E-01 |
| Reproduction_FSH-beta signaling pathway | 1.09E-02 | 5.16E-01 |
| Cell adhesion_Leucocyte chemotaxis | 1.55E-02 | 5.16E-01 |
| Muscle contraction | 1.71E-02 | 5.16E-01 |
| Development_Skeletal muscle development | 2.31E-02 | 5.16E-01 |
| Development_Melanocyte development and pigmentation | 2.65E-02 | 5.16E-01 |
| Inflammation_Histamine signaling | 2.77E-02 | 5.16E-01 |
| Neurophysiological process_Transmission of nerve impulse | 3.23E-02 | 5.16E-01 |
| Immune response_BCR pathway | 3.25E-02 | 5.16E-01 |
| Protein folding_ER and cytoplasm | 3.61E-02 | 5.22E-01 |
| Signal Transduction_Cholecystokinin signaling | 4.46E-02 | 5.73E-01 |
| Reproduction_GnRH signaling pathway | 5.30E-02 | 5.73E-01 |
| Reproduction_Gonadotropin regulation | 5.57E-02 | 5.73E-01 |
| Inflammation_IgE signaling | 5.76E-02 | 5.73E-01 |

**Supplementary Table S7. Top 15 enriched process networks from $PGS_{Cog}$ (controls)**

| **Process Network** | ***p*-value** | **FDR** |
|---|---|---|
| Cardiac development_Wnt_beta-catenin, Notch, VEGF, IP3 and integrin signaling | 4.89E-04 | 7.77E-02 |
| Cytoskeleton_Cytoplasmic microtubules | 1.44E-03 | 1.15E-01 |
| Translation_Translation initiation | 3.34E-03 | 1.77E-01 |
| Reproduction_Progesterone signaling | 7.05E-03 | 2.80E-01 |
| Apoptosis_Anti-apoptosis mediated by external signals via NF-kB | 1.37E-02 | 4.34E-01 |
| Development_Cartilage development | 2.48E-02 | 6.13E-01 |
| Apoptosis_Endoplasmic reticulum stress pathway | 2.92E-02 | 6.13E-01 |
| Proteolysis_Ubiquitin-proteasomal proteolysis | 3.35E-02 | 6.13E-01 |
| Neurophysiological process_GABAergic neurotransmission | 3.47E-02 | 6.13E-01 |
| Transport_Sodium transport | 4.11E-02 | 6.54E-01 |
| Transcription_Nuclear receptors transcriptional regulation | 5.70E-02 | 7.91E-01 |
| Protein folding_ER and cytoplasm | 6.77E-02 | 7.91E-01 |
| Transport_Calcium transport | 7.50E-02 | 7.91E-01 |
| Development_Regulation of angiogenesis | 7.62E-02 | 7.91E-01 |
| Protein folding_Protein folding nucleus | 8.69E-02 | 7.91E-01 |

**Supplementary Table S8. Top 15 enriched process networks from $PGS_{nonCog}$ (controls)**

| **Process Networks** | ***p*-value** | **FDR** |
|---|---|---|
| Transport_Bile acids transport and its regulation | 1.42E-02 | 8.90E-01 |
| DNA damage_BER-NER repair | 1.54E-02 | 8.90E-01 |
| Protein folding_Protein folding nucleus | 1.68E-02 | 8.90E-01 |
| Transport_Potassium transport | 3.58E-02 | 1.00E+00 |
| Neurophysiological process_Corticoliberin signaling | 5.38E-02 | 1.00E+00 |
| Inflammation_TREM1 signaling | 5.86E-02 | 1.00E+00 |
| Inflammation_Innate inflammatory response | 6.06E-02 | 1.00E+00 |
| Neurophysiological process_Transmission of nerve impulse | 7.28E-02 | 1.00E+00 |
| Reproduction_Gonadotropin regulation | 9.82E-02 | 1.00E+00 |
| Transport_Calcium transport | 1.03E-01 | 1.00E+00 |
| Signal transduction_Neuropeptide signaling pathways | 1.03E-01 | 1.00E+00 |
| Apoptosis_Death Domain receptors & caspases in apoptosis | 1.22E-01 | 1.00E+00 |
| Inflammation_Complement system | 1.41E-01 | 1.00E+00 |
| Neurophysiological process_GABAergic neurotransmission | 1.55E-01 | 1.00E+00 |
| Neurophysiological process_Circadian rhythm | 1.71E-01 | 1.00E+00 |

**Supplementary Table S9. Top 15 Transcription Factors from $PGS_{EA}$**

| Gene IDs | Transcription Factors | z-score | *p*-value |
|---|---|---|---|
| TCF4 | TCF7L2 (TCF4) | 1.79E+01 | 1.07E-73 |
| NR3C1 | GCR | 1.41E+01 | 1.41E-45 |
| ESRRG | Esrrg | 1.30E+01 | 2.74E-39 |
| TEAD1 | TEF-1 | 1.19E+01 | 1.45E-33 |
| ARNTL | BMAL1 | 1.01E+01 | 8.54E-25 |
| MEIS2 | MEIS2 | 8.20E+00 | 1.87E-17 |
| ESR1 | ESR1 (nuclear) | 8.26E+00 | 6.13E-17 |
| NRF1 | NRF1 | 7.93E+00 | 9.69E-16 |
| HNF4A | HNF4-alpha | 7.92E+00 | 1.32E-15 |
| RBPJ | RBP-J kappa (CBF1) | 7.73E+00 | 5.61E-15 |
| MEIS1 | MEIS1 | 7.29E+00 | 1.09E-13 |
| HMGB1 | HMGB1 | 6.63E+00 | 1.53E-11 |
| NFE2L2 | NRF2 | 5.86E+00 | 2.16E-09 |
| PBX1 | PBX1 | 5.74E+00 | 3.73E-09 |
| LEF1 | Lef-1 | 5.71E+00 | 4.53E-09 |

**Supplementary Table S10. Top 15 Transcription Factors from $PGS_{Cog}$**

| Gene IDs | Transcription Factors | z-score | *p*-value |
|---|---|---|---|
| TCF7L2 | TCF7L2 (TCF4) | 1.54E+01 | 4.99E-54 |
| ESRRG | Esrrg | 1.27E+01 | 4.63E-37 |
| TEAD1 | TEF-1 | 1.12E+01 | 3.36E-29 |
| ARNTL | BMAL1 | 1.11E+01 | 1.02E-28 |
| MEIS2 | MEIS2 | 8.82E+00 | 3.40E-19 |
| TFCP2L1 | LBP9 | 7.79E+00 | 3.65E-15 |
| ESR1 | ESR1 (nuclear) | 7.45E+00 | 5.49E-14 |
| SPI1 | PU.1 | 6.38E+00 | 1.06E-10 |
| HNF4A | HNF4-alpha | 6.01E+00 | 1.03E-09 |
| OLIG2 | OLIG2 | 5.80E+00 | 3.63E-09 |
| PBX1 | PBX1 | 5.52E+00 | 2.09E-08 |
| ZBTB20 | ZNF288 | 4.24E+00 | 1.46E-05 |
| ASCL1 | hASH1 | 3.94E+00 | 5.44E-05 |
| TAL1 | TAL1 | 3.70E+00 | 1.18E-04 |
| PAX5 | PAX5 | 3.19E+00 | 9.45E-04 |

**Supplementary Table S11. Top 15 Transcription Factors from $PGS_{nonCog}$**

| Gene IDs | Transcription Factors | *z*-score | *p*-value |
|---|---|---|---|
| TCF4 | TCF7L2 (TCF4) | 1.14E+01 | 8.49E-33 |
| NR3C1 | GCR | 9.46E+00 | 1.00E-21 |
| TEAD1 | TEF-1 | 7.55E+00 | 3.16E-15 |
| ESRRG | Esrrg | 7.70E+00 | 4.04E-15 |
| SMAD1 | SMAD1 | 7.18E+00 | 2.50E-13 |
| VSX2 | CHX10 | 6.72E+00 | 3.74E-13 |
| ARNTL | BMAL1 | 6.45E+00 | 9.40E-12 |
| MYRF | C11orf9 | 6.21E+00 | 3.71E-11 |
| ESR1 | ESR1 (nuclear) | 5.33E+00 | 3.08E-08 |
| TFCP2L1 | LBP9 | 5.02E+00 | 3.87E-07 |
| NRF1 | NRF1 | 4.79E+00 | 5.72E-07 |
| MEIS1 | MEIS1 | 4.41E+00 | 2.50E-06 |
| SPI1 | PU.1 | 4.30E+00 | 4.92E-06 |
| MEIS2 | MEIS2 | 4.14E+00 | 5.15E-06 |
| PBX1 | PBX1 | 3.84E+00 | 3.05E-05 |

**Supplementary Table S12. Top 15 Transcription Factors from $PGS_{EA}$ (controls)**

| Gene IDs | Transcription Factors | *z*-score | *p*-value |
|---|---|---|---|
| TFCP2L1 | LBP9 | 3.35E+00 | 4.38E-04 |
| GLI1 | GLI-1 | 2.78E+00 | 3.22E-03 |
| FOXF1 | FOXF1 | -2.73E+00 | 5.27E-03 |
| FOXN1 | FoxN1 | 2.46E+00 | 9.37E-03 |
| ZEB1 | TCF8 | 2.12E+00 | 1.80E-02 |
| TAF2 | TFIID 30 kDa subunit | -2.36E+00 | 1.83E-02 |
| ARNTL | BMAL1 | 1.98E+00 | 2.55E-02 |
| ZBTB46 | Zbtb46 | 2.09E+00 | 2.62E-02 |
| TAF2 | TAFII70 | 2.19E+00 | 2.95E-02 |
| FOXP1 | FOXP1 | 1.94E+00 | 3.32E-02 |
| SOX5 | SOX5 | -2.00E+00 | 3.33E-02 |
| NR3C2 | MCR | 1.96E+00 | 3.44E-02 |
| GLIS3 | GLIS3 | 1.77E+00 | 4.03E-02 |
| TAF2 | TAF7L | -1.94E+00 | 4.05E-02 |
| SREBF2 | SREBP2 precursor | 1.85E+00 | 4.15E-02 |

**Supplementary Table S13. Top 15 Transcription Factors from $PGS_{Cog}$ (controls)**

| Gene IDs | Transcription Factors | *z*-score | *p*-value |
|---|---|---|---|
| C16orf47 | ATBF1 | 2.54E+00 | 1.25E-02 |
| RIOK2 | RIOK2 | 1.80E+00 | 1.12E-01 |
| MYOCD | Myocardin | 1.80E+00 | 1.12E-01 |
| ZNF831 | ZNF831 | 1.47E+00 | 2.32E-01 |
| VAX1 | Vax1 | 1.47E+00 | 2.32E-01 |
| TBPL2 | TRF3 | 1.47E+00 | 2.32E-01 |
| ETV7 | TEL2 | 1.47E+00 | 2.32E-01 |
| TAF1B | TAF(I)63 | 1.04E+00 | 4.82E-01 |
| TFEC | TFEC | 1.04E+00 | 4.82E-01 |
| PGR | PR (membrane) | 1.04E+00 | 4.82E-01 |
| SMARCA5 | SMARCA5 | 1.04E+00 | 4.82E-01 |
| ARID5B | ARID5B | 1.04E+00 | 4.82E-01 |
| MET | ETV3 | 1.04E+00 | 4.82E-01 |
| HIVEP1 | HIVEP1 (PRDII-BF1) | 1.04E+00 | 4.82E-01 |
| ZNF37A | ZNF37A | 1.04E+00 | 4.82E-01 |

**Supplementary Table S14. Top 15 Transcription Factors from $PGS_{nonCog}$ (controls)**

| Gene IDs | Transcription Factors | *z*-score | *p*-value |
|---|---|---|---|
| DLX5 | DLX5 | 1.86E+00 | 4.16E-02 |
| CASZ1 | CASZ1 | 1.21E+00 | 2.62E-01 |
| MBD1 | MBD1 | 1.21E+00 | 2.62E-01 |
| ZBTB38 | ZBTB38 | 1.13E+00 | 3.10E-01 |
| TOX2 | TOX2 | 1.13E+00 | 3.10E-01 |
| ZNF304 | ZNF304 | 1.13E+00 | 3.10E-01 |
| MPP4 | NFAT-90 | 1.13E+00 | 3.10E-01 |
| MKX | IRXL1 | 1.05E+00 | 3.67E-01 |
| ATF7 | ATF-7 | 1.05E+00 | 3.67E-01 |
| LMX1B | LMX1B | 1.05E+00 | 3.67E-01 |
| NACC1 | Nac1 | 1.05E+00 | 3.67E-01 |
| KLF13 | KLF13 | 1.05E+00 | 3.67E-01 |
| BNC1 | Basonuclin | 1.05E+00 | 3.67E-01 |
| TLX1 | TLX3 | 9.54E-01 | 4.33E-01 |
| HEYL | HEYL | 9.54E-01 | 4.33E-01 |

## References


1. Vilhjálmsson, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
2. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: Better, faster, stronger. *Bioinformatics* **36**, 5424–5431 (2020).
3. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).
4. Demange, P. A. *et al.* Investigating the genetic architecture of noncognitive skills using GWAS-by-subtraction. *Nat. Genet.* **53**, 35–44 (2021).
5. Privé, F., Albiñana, C., Arbel, J., Pasaniuc, B. & Vilhjálmsson, B. J. Inferring disease architecture and predictive ability with LDpred2-auto. *Am. J. Hum. Genet.* **110**, 2042–2055 (2023).
6. Prive, F., Aschard, H., Ziyatdinov, A. & Blum, M. G. B. Efficient analysis of large-scale genome-wide data with two R packages: Bigstatsr and bigsnpr. *Bioinformatics* **34**, 2781–2787 (2018).
7. Privé, F., Arbel, J., Aschard, H. & Vilhjálmsson, B. J. Identifying and correcting for misspecifications in GWAS summary statistics and polygenic scores. *Human Genetics and Genomics Advances* **3**, (2022).
8. Murphy, A. E., Schilder, B. M. & Skene, N. G. MungeSumstats: A Bioconductor package for the standardization and quality control of many GWAS summary statistics. *Bioinformatics* **37**, 4593–4596 (2021).