**Supporting Information**

**ADHD and Intelligence polygenic scores and their associations with dimensional developmental characteristics within a paediatric transdiagnostic cohort**

Andrea M. Santangelo1, Olena Ohlei2, Silvana Mareva1,3, Diandra Brkic1,4, CALM team1, Lars Bertram2, Joni Holmes1,5, Duncan Astle1,6 and Kate Baker1,7

1 MRC Cognitive and Brain Sciences Unit, University of Cambridge, UK

2 Lubeck Interdisciplinary Platform for Genome Analysis (LIGA), University of Luebeck, Germany

3 Psychology Department, University of Exeter, UK

4Brain Health Department, Nestlé Institute of Health Sciences, Switzerland

5 School of Psychology, University of East Anglia, UK

6 Department of Psychiatry, University of Cambridge, UK

7 Department of Medical Genetics, University of Cambridge, UK

**List of content**

Supporting materials and methods

Supporting Figure S1

Supporting table S1-S6

**Supporting materials and methods**

***Genotyping, QC filtering and imputation procedures***

DNA samples were collected from saliva in vials using the Oragene® DNA self-collection kits (DNA Genotek Inc). Children were asked to produce a saliva sample by first rubbing their cheeks gently for 30 s to create saliva, and then they were asked to spit in a pot. For children who find it hard to create saliva, a small amount (max ¼ tsp) of white table sugar was available to place on the child’s tongue. The saliva samples were stored in the Oragene® kits at room temperature (15–30 °C) until DNA extraction, which was performed using prepIT∙L2P Kit (DNA Genotek Inc) following manufacturer instructions. Once extracted, DNA was stored at -80 °C at the Wellcome Trust-MRC Institute of Metabolic Science at Addenbrooke’s Hospital, and subsequently transferred to the BioRepository for long-term storage.

Genome-wide SNP genotyping was performed in two batches (CALM1 N=632, and CALM2 N=143) using the Global Screening Array with shared custom-content (GSA, Illumina, Inc., USA) on an iScan instrument according to manufacturer’s recommendations. The raw intensity data was subjected to genotype calling, extensive pre-imputation quality control and SNP imputation using an automated computational workflow described previously (Hong et al., 2020, 2021). A total of 7 samples were excluded because of low quality DNA or evidence of cross-contamination. We estimated the genetic sex of participants using the command --check-sex 0.25 0.75 in PLINK (Purcell et al., 2007). This led to the exclusion of 10 participant in whom genetic sex could not be determined unambiguously or in whom there was a discrepancy with the clinically reported sex. One more sample was removed because of excessive missing data (--mind 0.05). The LD pruned dataset was also used for principal component analysis (PCA; using PLINK command ‘--pca') along with the reference dataset of the 1000 Genome Project Consortium Phase 3 (The 1000 Genomes Project Consortium, 2015) to assign ethnic descent groups using the five 1000G super-populations by k-nearest neighbour classification (k-NN; k=9; using R package ‘class’ in R 2.3.2 (Venables & Ripley, 2002). This resulted in assigning “European descent” to 678 out of all 757 samples across both genotyping batches; only these n=678 European-descent samples were used in the subsequent analyses.

Genotype imputation was performed on the QC’ed and filtered data with MiniMac3 (Das et al., 2016) software using the "Haplotype Reference Consortium" (HRC; v1.1 [EGAD00001002729 including 39,131,578 SNPs from ~11K individuals]) reference panel (McCarthy et al., 2016). Overall, this procedure resulted in 39,131,578 imputed SNPs for each 561 and 117 subjects in CALM1 and CALM2 datasets, respectively.

***PGS calculation***

To aggregate data on multiple variants per individual we computed PGS for each individual as described previously (Choi et al., 2020). Allele status and effect-size estimates were taken from the summary statistics of the largest ADHD (Demontis et al., 2023) and Intelligence (Savage et al., 2018) meta-GWAS published to date. First, we removed duplicate and ambiguous SNPs (A/T and C/G) from the list of considered variants. We only used SNPs with MAF >0.05 (because our target dataset is N<1000 samples), and imputation quality r2 >0.8. The remaining variants were 5,685,483 (out of 6,774,224 SNPs) for ADHD-GWAS, and 4,673,194 (out of 9,295,118 SNPs) for Intelligence-GWAS. The target CALM sample did not overlapped with any of these GWAS datasets. Both GWAS and the CALM genotyping data were annotated on the GRCh37/hg19 genome build. VCF files containing the genotyping data from CALM1 and CALM2 were merged using BCFTOOLS. The merged file was converted into PLINK binary files using --vcf. The non-referred individuals were removed, remaining only 557 samples. Further QC included filtering for samples with extreme heterozygosity (more than 3 SD from the mean), by performing first pruning (--indep-pairwise 200 50 0.25) and then calculating the F coefficient for each individual (--het), leading to the exclusion of 9 samples. Finally, 24 closely related individuals were also removed (--rel-cutoff 0.125), giving a final sample size of **N=524**. Duplicate SNPs were removed using --exclude in PLINK, with 38,913,048 out of 39,131,578 variants remaining. After further removing all SNPs with MAF<0.01, P value for HWE < 1e-6, and genotyping efficiency <99% (--geno 0.01), 7,568,187 variants remained. Variant matching and recoding between the ADHD/Intelligence GWAS base and CALM target datasets was performed using R (version 3.5.3), resulting in 5,583,323 and 3,994,342 overlapping SNPs for ADHD and Intelligence, respectively.

To account for LD, we performed clumping in PLINK (--clump-p1 1, --clump-r2 0.1, --clump-kb 250) to remove the SNPs highly linked to the index SNP and with low association to the GWAS-specific phenotype. To account for population stratification, we performed a PCA on the pruned SNP pool (--indep-pairwise 200 50 0.25) and extracted the first six PC (--pca 6). Actual PGS were computed using the --score command for a variety of P-value thresholds (Pt) in the primary GWAS data (0.001, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50).

**References**

Choi, S. W., Mak, T. S. H., & O’Reilly, P. F. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*, *15*(9), 2759–2772. https://doi.org/10.1038/S41596-020-0353-1

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A. E., Kwong, A., Vrieze, S. I., Chew, E. Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P. R., Iacono, W. G., Swaroop, A., Scott, L. J., Cucca, F., Kronenberg, F., Boehnke, M., … Fuchsberger, C. (2016). Next-generation genotype imputation service and methods. *Nature Genetics*, *48*(10), 1284–1287. https://doi.org/10.1038/NG.3656

Demontis, D., Walters, G. B., Athanasiadis, G., Walters, R., Therrien, K., Nielsen, T. T., Farajzadeh, L., Voloudakis, G., Bendl, J., Zeng, B., Zhang, W., Grove, J., Als, T. D., Duan, J., Satterstrom, F. K., Bybjerg-Grauholm, J., Bækved-Hansen, M., Gudmundsson, O. O., Magnusson, S. H., … Børglum, A. D. (2023). Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nature Genetics 2023 55:2*, *55*(2), 198–208. https://doi.org/10.1038/s41588-022-01285-8

Hong, S., Dobricic, V., Ohlei, O., Bos, I., Vos, S. J. B., Prokopenko, D., Tijms, B. M., Andreasson, U., Blennow, K., Vandenberghe, R., Gabel, S., Scheltens, P., Teunissen, C. E., Engelborghs, S., Frisoni, G., Blin, O., Richardson, J. C., Bordet, R., Lleó, A., … Bertram, L. (2021). TMEM106B and CPOX are genetic determinants of cerebrospinal fluid Alzheimer’s disease biomarker levels. *Alzheimer’s & Dementia : The Journal of the Alzheimer’s Association*, *17*(10), 1628–1640. https://doi.org/10.1002/ALZ.12330

Hong, S., Prokopenko, D., Dobricic, V., Kilpert, F., Bos, I., Vos, S. J. B., Tijms, B. M., Andreasson, U., Blennow, K., Vandenberghe, R., Cleynen, I., Gabel, S., Schaeverbeke, J., Scheltens, P., Teunissen, C. E., Niemantsverdriet, E., Engelborghs, S., Frisoni, G., Blin, O., … Bertram, L. (2020). Genome-wide association study of Alzheimer’s disease CSF biomarkers in the EMIF-AD Multimodal Biomarker Discovery dataset. *Translational Psychiatry*, *10*(1). https://doi.org/10.1038/S41398-020-01074-Z

McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A. R., Teumer, A., Kang, H. M., Fuchsberger, C., Danecek, P., Sharp, K., Luo, Y., Sidore, C., Kwong, A., Timpson, N., Koskinen, S., Vrieze, S., Scott, L. J., Zhang, H., Mahajan, A., … Marchini, J. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, *48*(10), 1279–1283. https://doi.org/10.1038/NG.3643

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., De Leeuw, C. A., Nagel, M., Awasthi, S., Barr, P. B., Coleman, J. R. I., Grasby, K. L., Hammerschlag, A. R., Kaminski, J. A., Karlsson, R., Krapohl, E., Lam, M., Nygaard, M., Reynolds, C. A., Trampush, J. W., … Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics*, *50*(7), 912–919. https://doi.org/10.1038/S41588-018-0152-6

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S*. https://doi.org/10.1007/978-0-387-21706-2

**Supporting Figure S1**



**Figure S1**: Frequency distribution of the Index of Multiple Deprivation (IMD) in deciles.

**Supporting tables**

|  |
| --- |
| **Table S1: Bonferroni corrected p values** |
|  | **Dimensions** | **HiTOP moldels** | **ADHD diagnosis** |
|  | **Measures N=6** | **1 factor** | **2 factor** | **3 factor** | **ADHD scores N=2** |
|  | alpha/6 | alpha/1 | alpha/2 | alpha/3 | alpha/2 |
| **ADHD-PGS** | p<.008 | p<.05 | p<.025 | p<.017 | p<.025 |
| **Intelligence-PGS** | p<.008 | p<.05 | p<.025 | p<.017 | NA |

|  |
| --- |
| **Table S2: Linear regression between PGS and behavioural and cognitive dimensions using sex, age and the first 6 ancestry PC as covariates (\* p< .008 after Bonferroni correction, Table S1).** |
| **ADHD-PGS** |
|  | **Hyp/Imp** | **Inatt** | **Intern** | **Extern** | **Non-verb Intelligence** | **Verbal Intelligence** |
| **Pt** | 0.4 | 0.4 | 0.001 | 0.2 | 0.4 | 0.3 |
| **r2** | 0.030 | 0.019 | 0.007 | 0.043 | 0.011 | 0.020 |
| **r** | 0.173 | 0.139 | 0.083 | 0.208 | -0.105 | -0.140 |
| **Beta** | 38672 | 19114 | 1415 | 24739 | -40704 | -143298 |
| **P** | **3.06E-05\*** | **1.44E-03\*** | 5.19E-02 | **1.06E-06\*** | 1.57E-02 | **1.02E-03\*** |
| **Intelligence-PGS** |
| **Pt** | 0.1 | 0.5 | 0.05 | 0.001 | 0.5 | 0.001 |
| **r2** | 0.000 | 0.006 | 0.002 | 0.004 | 0.046 | 0.054 |
| **r** | -0.008 | -0.077 | 0.041 | -0.064 | 0.213 | 0.232 |
| **Beta** | -2209 | -28405 | 7687 | -3348 | 222200 | 85549 |
| **P** | 8.56E-01 | 7.95E-02 | 3.35E-01 | 1.36E-01 | **7.39E-07\*** | **4.01E-08\*** |

|  |
| --- |
| **Table S3: Linear regression between PGS and behavioural and cognitive dimensions using sex, age, the first 6 ancestry PC, and IMD as covariates (\* p< .008 after Bonferroni correction, Table S1).** |
| **ADHD-PGS** |
|  | **Hyp/Imp** | **Inatt** | **Intern** | **Extern** | **Non-verb Intelligence** | **Verbal Intelligence** |
| **Pt** | 0.4 | 0.4 | 0.001 | 0.2 | 0.4 | 0.3 |
| **r2** | 0.024 | 0.014 | 0.005 | 0.041 | 0.009 | 0.011 |
| **r** | 0.154 | 0.117 | 0.071 | 0.202 | -0.095 | -0.105 |
| **Beta** | 35164 | 16060 | 1209 | 24304 | -37906 | -109345 |
| **P** | **2.62E-04\*** | 9.91E-03 | 1.02E-01 | **3.06E-06\*** | 3.48E-02 | 1.69E-02 |
| **Intelligence-PGS** |
| **Pt** | 0.2 | 0.5 | 0.05 | 0.001 | 0.5 | 0.001 |
| **r2** | 0.000 | 0.003 | 0.003 | 0.003 | 0.049 | 0.043 |
| **r** | 0.008 | -0.058 | 0.058 | -0.056 | 0.220 | 0.208 |
| **Beta** | 3058 | -21274 | 10898 | -2892 | 234104 | 77214 |
| **P** | 8.58E-01 | 2.02E-01 | 1.86E-01 | 2.05E-01 | **8.14E-07\*** | **1.78E-06\*** |

|  |
| --- |
| **Table S4: Linear regression between PGS and factor models of psychopathology using sex, age and the first 6 ancestry PC as covariates (\* p< .05, \*\*p< .025, \*\*\*p< .017 after Bonferroni corrections Table S1).** |
| **ADHD-PGS** |
|  | **P factor** | **Broad Ext** | **Broad Int** | **Social adj** | **Specific Int** | **Neurodev** |
| **Pt** | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.4 |
| **r2** | 0.028 | 0.029 | 0.012 | 0.033 | 0.014 | 0.012 |
| **r** | 0.167 | 0.171 | 0.111 | 0.183 | 0.118 | 0.107 |
| **Beta** | 2769 | 2809 | 1819 | 3051 | 1942 | 4355 |
| **P** | **0.014\*** | **0.011\*\*** | 0.103 | **0.006\*\*\*** | 0.082 | 0.114 |
| **Intelligence-PGS** |
| **Pt** | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.5 |
| **r2** | 0.015 | 0.012 | 0.010 | 0.016 | 0.010 | 0.008 |
| **r** | -0.124 | -0.109 | -0.098 | -0.127 | -0.102 | -0.091 |
| **Beta** | -1596 | -1389 | -1250 | -1648 | -1294 | -10009 |
| **P** | 0.067 | 0.104 | 0.148 | 0.058 | 0.135 | 0.181 |

|  |
| --- |
| **Table S5: Linear regression between PGS and factor models of psychopathology using sex, age, the first 6 ancestry PC, and IMD as covariates (\* p< .05, \*\*p< .025, \*\*\*p< .017 after Bonferroni corrections Table S1).** |
| **ADHD-PGS** |
|  | **P factor** | **Broad Ext** | **Broad Int** | **Social adj** | **Specific Int** | **Neurodev** |
| **Pt** | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.4 |
| **r2** | 0.013 | 0.017 | 0.003 | 0.018 | 0.004 | 0.009 |
| **r** | 0.112 | 0.131 | 0.059 | 0.134 | 0.066 | 0.093 |
| **Beta** | 1840 | 2163 | 950 | 2257 | 1062 | 3829 |
| **P** | 0.091 | 0.045 | 0.390 | 0.039 | 0.336 | 0.177 |
| **Intelligence-PGS** |
| **Pt** | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.5 |
| **r2** | 0.004 | 0.004 | 0.002 | 0.006 | 0.003 | 0.006 |
| **r** | -0.066 | -0.062 | -0.049 | -0.075 | -0.051 | -0.081 |
| **Beta** | -839 | -788 | -607 | -967 | -633 | -9005 |
| **P** | 0.318 | 0.343 | 0.475 | 0.252 | 0.456 | 0.241 |

|  |
| --- |
| **Table S6: Linear regression between ADHD-PGS and ADHD symptom scores using sex, age, the first 6 ancestry PC and ADHD diagnosis as covariates (\*\*p< .025 after Bonferroni corrections Table S1).** |
| **Without IMD as covariate** |
|  | **Hyperactivity/Impulsivity** | **Inattention** |
| **Pt** | 0.5 | 0.3 |
| **r2** | 0.012 | 0.011 |
| **r** | 0.11 | 0.10 |
| **Beta** | 28089 | 12461 |
| **P** | **3.05E-03\*\*** | **1.56E-02\*\*** |
| **With IMD as covariate** |
|  | **Hyperactivity/Impulsivity** | **Inattention** |
| **Pt** | 0.5 | 0.3 |
| **r2** | 0.011 | 0.008 |
| **r** | 0.11 | 0.09 |
| **Beta** | 27027 | 10620 |
| **P** | **6.11E-03\*\*** | 4.61E-02 |