Data supplement for Harich et al., From Rare Copy Number Variants to Biological Processes in ADHD. Am J Psychiatry (doi: 10.1176/appi.ajp.2020.19090923)

SUPPLEMENTAL METHODS, FIGURES, AND TABLES

SQL query script for the UCSC database, GWAS meta-analyses data set for ADHD and Gene-based association analyses for ADHD GWAS meta-analyses data:

```
#select below for SQL query:

select
CNV.name as Study, CNV.type as
Type, Gene.name as Refseq,
Gene.name2 as Name,
Gene.strand as Strand, CNV.chrom
as Chromosome,
CNV.chromStart as CNV_Start,
CNV.chromEnd as CNV_End,
Gene.txStart as Gene_Start,
Gene.txEnd as Gene_End,
(Gene.txEnd-Gene.txStart) as Gene_Size,
(CNV.chromEnd-CNV.chromStart) as CNV_Size,

#CREATE COLUMN LOCATION
if (CNV.chromStart<Gene.txEnd and Gene.txStart<CNV.chromEnd, #The CNV
overlaps with the gene, where does it overlap?
if (Gene.txStart<CNV.chromStart,
#It overlaps with right, what part of gene? if
(Gene.strand='+',"Cterm","Nterm"),
if (CNV.chromEnd<Gene.txEnd,
#It overlaps with left, what part of gene? if
(Gene.strand='+',"Nterm","Cterm"), #Neither
left or right, thus whole gene "CDS")
),
#The CNV does not overlap with the gene, does it overlap with 2kb of regulatory region:
if(Gene.txEnd<CNV.chromStart,if(Gene.strand='+',"3'UTR","PROM"),if(CNV.chromEnd<Gene.txSt
art,if(Gene.strand='+',"PROM","3'UTR"),"NULL"))

) as CNV_Location,

#CREATE COLUMN OVERLAP
if (CNV.chromStart<Gene.txEnd and Gene.txStart<CNV.chromEnd, #The
CNV overlaps with the gene, how large is the overlap?
if (Gene.txStart<CNV.chromStart,Gene.txEnd-CNV.chromStart,
if (CNV.chromEnd<Gene.txEnd,CNV.chromEnd-Gene.txStart,Gene.txEnd-Gene.txStart)
```

```
),
#The CNV does not overlap with the gene: "0")
as Coding_Overlap,

#CREATE COLUMN GENEPERCENTAGE

if (CNV.chromStart<Gene.txEnd and Gene.txStart<CNV.chromEnd, #The
CNV overlaps with the gene, how large is the overlap?
if (Gene.txStart<CNV.chromStart,(Gene.txEnd-CNV.chromStart)/(Gene.txEnd-Gene.txStart),
if (CNV.chromEnd<Gene.txEnd,(CNV.chromEnd-Gene.txStart)/(Gene.txEnd-Gene.txStart),"1")
),
#The CNV does not overlap with the gene:
"0") as Coding_Fraction,

#CREATE COLUMN PROMOTER
if(Gene.strand='+',if(CNV.chromStart<Gene.txStart,Gene.txStart-
CNV.chromStart,"NULL"),if(Gene.txEnd<CNV.chromEnd,CNV.chromEnd-Gene.txEnd,"NULL")) as
"5'UTR_Overlap",

#CREATE COLUMN UTR
if(Gene.strand='-',if(CNV.chromStart<Gene.txStart,Gene.txStart-
CNV.chromStart,"NULL"),if(Gene.txEnd<CNV.chromEnd,CNV.chromEnd-Gene.txEnd,"NULL")) as
"3'UTR_Overlap",

#CREATE COLUMN GAP
if(cnv.chromend<gene.txstart,gene.txstart-
cnv.chromend,if(gene.txend<cnv.chromstart,cnv.chromstart-gene.txend,"NULL")) as
"Gap_CNV-Gene"

# DECLARE THE TABLE:
##########################################################################

from 140822_transcoded_corrected

##########################################################################################
###########
as CNV

left join refGene as Gene on

(CNV.chrom=Gene.chrom and not(Gene.txEnd<(CNV.chromStart-2000) or
(CNV.chromEnd+2000)<Gene.txStart))
```

GWAS meta-analyses data set for ADHD

The cohorts include eleven clinical collections of the PGC and 37,076 samples of the Danish Bloodspot efforts (iPSYCH). Samples were of Caucasian or Han Chinese origin and met diagnostic criteria according to the DSM-IV.[1] Written informed consent was obtained from all participants. Each study was approved by the respective institutional review board or local ethics committee. The meta-analytic data used in this study were available as summary statistics, including genome-wide SNP data with corresponding P- values and odds ratios. Detailed procedures of DNA isolation, whole-genome genotyping and imputation were described previously.[1] Shortly, genome-wide data was obtained from different genotyping arrays and was imputed using 1000 Genomes data as a reference panel (phase 3, version 5 (1KGP3v5)) in NCBI build 37 (hg19) coordinates) for autosomal SNPs.[2] Meta-analytic data were processed through a stringent quality control pipeline applied at the PGC.[1] Only SNPs with an imputation quality score of INFO ≥ 0.8 and a minor allele frequency ≥ 0.01 were included in our analyses.

Gene-based association analyses for ADHD GWAS meta-analyses data

We used data from the recent meta-analysis of genome-wide association studies (GWAS) of 20,183 patients with ADHD and 35,191 controls as performed by the Psychiatric Genomics Consortium (PGC) ADHD Working Group and the Danish iPSYCH Initiative. Details on the samples and quality control (as described above and in Demontis et.al 2019).[1]

Gene-based association analyses were performed using the Multi-marker Analysis of GenoMic Annotation (MAGMA) software package (version 1.05)[3]. First, genome-wide SNP data from a reference panel (1000 Genomes, v3 phase1)[4] was annotated to NCBI Build 37.3 gene locations using a symmetric 100 kb flanking window and both files were downloaded from http://ctglab.nl/software/magma. Next, the gene annotation file was used to map the genome-wide SNP data, to assign SNPs to genes, and to calculate gene-based p-values. For the gene-based analyses, single SNP p-values within a gene were transformed into a gene-statistic by taking the mean of the $\chi 2$-statistic among the SNPs in each gene. To account for linkage disequilibrium (LD), the 1000 Genomes Project European sample was used as a reference to estimate the LD between SNPs within (the vicinity of) the genes (http://ctglab.nl/software/MAGMA/ref_data/g1000_ceu.zip).

Gene-wide p-values were converted to z-values reflecting the strength of the association of each gene with the phenotype, with higher z-values corresponding to stronger associations. All individual genes were investigated, by reviewing their gene test-statistics. Genes were considered gene-wide significant if they reached the Bonferroni correction threshold-adjusted for the number of genes tested ($p<0.05/26$).

References

1. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
2. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
3. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11**, 1–19 (2015).
4. Consortium, T. 1000 G. P. A map of human genome variation from population scale sequencing. *Nature* **467**, 1061–1073 (2010).

FIGURE S1. Regional association of genes most strongly associated with ADHD. Regional association plots showing association signal for ADHD in the PGC-iPSYCH GWAS meta-analysis data for the two most strongly associated high priority ADHD candidate genes, including flanking regions of 100kb. (A) *POLR3C* locus with the top-SNP (rs376814422) indicated by the black arrow. (B) *RBFOX1* locus with the top-SNP (rs6500945) indicated by the purple dot. Results are shown as -log (p value) for genotyped and imputed SNPs. The color of each marker reflects its LD (r2) with the SNP indicated by the purple dot. The recombination rate is plotted in blue. cm/Mb, centimorgan/megabase. Chr, chromosome.

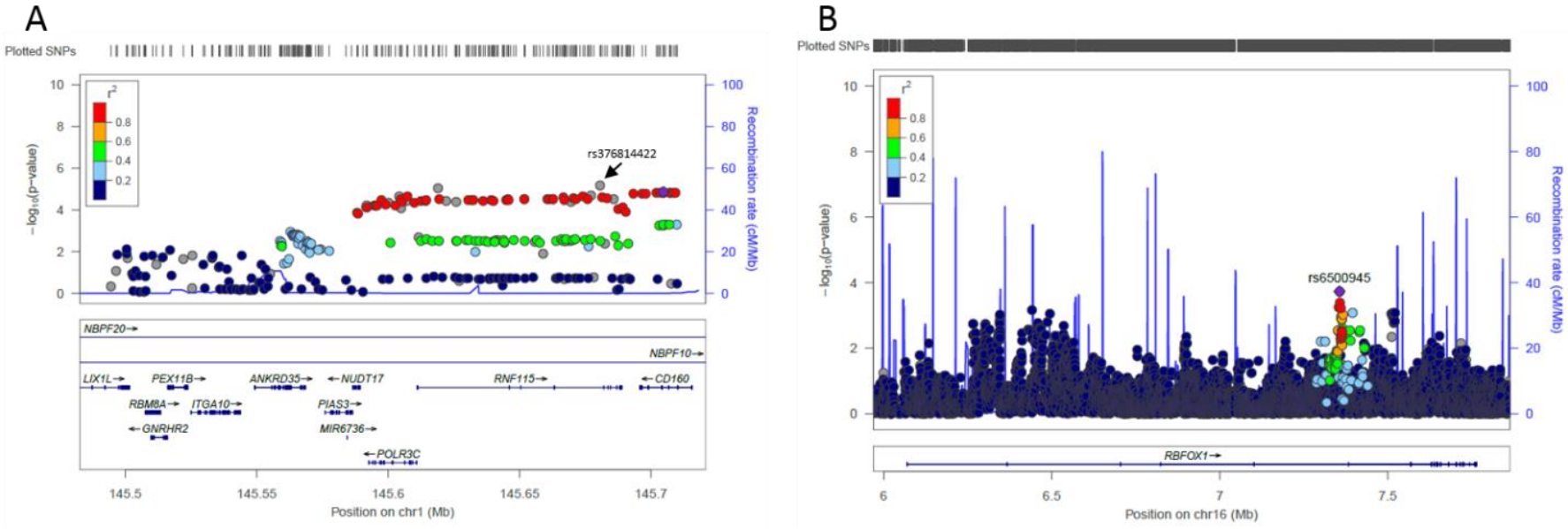

FIGURE S2. For prioritized genes, gene expression heatmaps were constructed with GTEx, version 8 (54 tissues; upper figure) and BrainSpan for 29 different ages of brain samples data (lower figure). Genes and tissues are ordered by clusters for the GTEx heatmap. For the BrainSpan heatmap, genes are ordered by expression clusters and developmental stages are ordered chronologically. The upper panel shows the expression of the 26 recurrently identified candidate genes in different tissues, including the brain; in the lower panel the brain expression at different timepoints during development is shown.

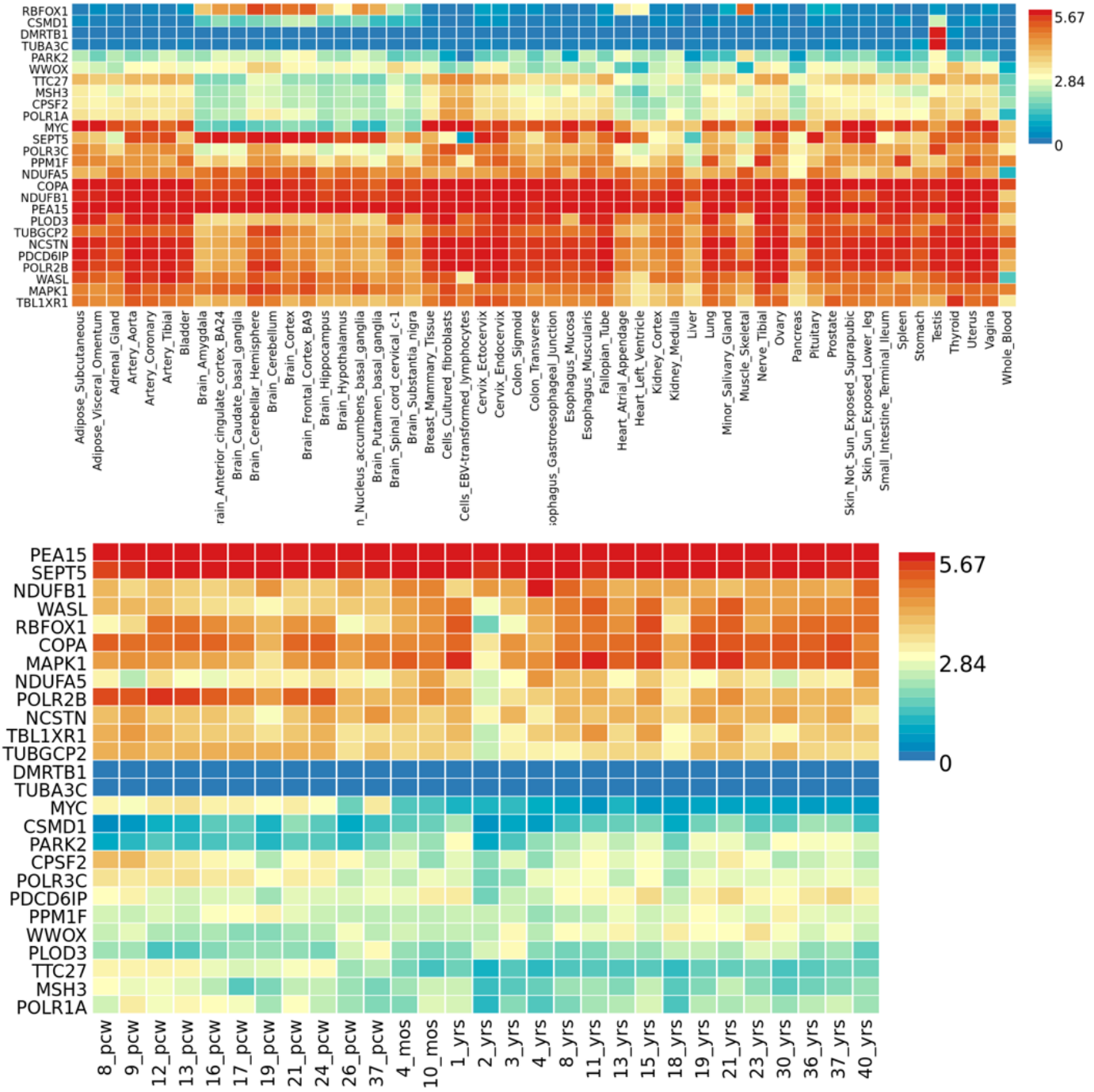

TABLE S1. Overview of all included studies reporting rare CNVs in an ADHD cohort

Study	Cohort	Cases (n)	Controls (n)	CNV size	Age distribution
2010 Elia	Caucasian (Philadelphia)	335 trios	2026	incl. Small	Children 6–18 years
2010 Williams	Caucasian (UK)	366	1047	> 500kb	Children 5–17 years
2011 Lesch	Caucasian (Germany)	99 + parents	100	one 75kb, the rest larger	Children 6–18 years, mean age 11 years
2011 Lionel	Caucasian (Ontario, Germany)	248	2357	>20kb	Children 5–17 years
2012 Elia	Caucasian (Philadelphia)	1013	4105	incl. Small	Children 6–17 years
2012 Stergiakouli	Caucasian (Wales, Scotland, Ireland)	727	5081	> 500kb	Children
2012 Wiliams	Caucasian (European)	732	2455	> 100kb	Children
2014 Jarick	Caucasian (Germany)	489	1285	> 500kb	Children 6 – 18 years, mean age 11 years
2014 Martin	Caucasian (Wales, Scotland, Ireland)	727	5081	> 500kb	Children
2014 Ramos-Quiroga	Caucasian (Spain)	400	526	> 100kb (except Table S4)	Adult samples, average age 31.2 years
2013 Yang	Asian (China)	1040	963	> 100kb	Children 6 – 16 years, mean age 9.7 years

TABLE S2. Full lists of raw input data points extracted from 11 ADHD CNV studies (Tab 1), High priority gene list; altered copy number observed in more than one patient and none of the controls (Tab 2) and Low-priority gene list; altered copy number in only one patient and no controls (Tab 3).

******See the Excel spreadsheet******

TABLE S3. Full list of all selected ADHD core phenotype labels in the Monarch Initiative database query

Search term in Monarch Initiative database
attention
behavioral
behavioural
hyperactive
hyperactivity
impulse
impulsivity
increased frequency
increased occurrence
increased rate
increased speed locomotion
increased vertical activity
locomot
locomotory behavior

Results			
	object	object_label	Monarch genes 12-jan-2017
	FBcv:0000392	hyperactive	47
	HP:0000734	Disinhibition	36
	HP:0000736	Short attention span	125
	HP:0000752	Hyperactivity	684
	HP:0007018	Attention deficit hyperactivity disorder	114
	HP:0100710	Impulsivity/abnormal impulsive behavior control	62
	MP:0002574	increased vertical activity	77
	MP:0002629	hyperactivity elicited by ethanol administration	8
	MP:0008911	induced hyperactivity	43
	MP:0009751	enhanced behavioral response to alcohol	17

TABLE S4. Significant connected genes of the high-priority gene list after DAPPLE analysis and the assigned modules in Figure 3.

Gene	P corrected	Assigned Module in Figure3
WWOX	0,0297	Module 1
POLR1A	0,0356	Module 1
CTNNA3	0,0415	Module 2
DENR	0,0317	Module 3
BIRC6	0,0120	Module 4
RAB15	0,0396	Module 4
SEPT5	0,0396	Module 4
PHKB	0,0139	Module 4
SDK1	0,0020	no assigned module
UXS1	0,0020	no assigned module
C12orf65	0,0020	no assigned module
PRRT2	0,0080	no assigned module
FAM84B	0,0100	no assigned module
TAGLN2	0,0159	no assigned module
FIS1	0,0179	no assigned module
SLC6A12	0,0219	no assigned module
ADSS	0,0278	no assigned module
ATP1A4	0,0317	no assigned module
FKBP3	0,0337	no assigned module
SETD8	0,0454	no assigned module

TABLE S5. Results MAGMA analysis of the 26 genes consistently observed in all the different approaches.

******See the Excel spreadsheet******

TABLE S6. Results MAGMA analyses of the 26 genes with the combined measure in the PGC cross-disorder GWAS meta-analysis ($n_{no23\&Me}$=438,997) (32).

******See the Excel spreadsheet******