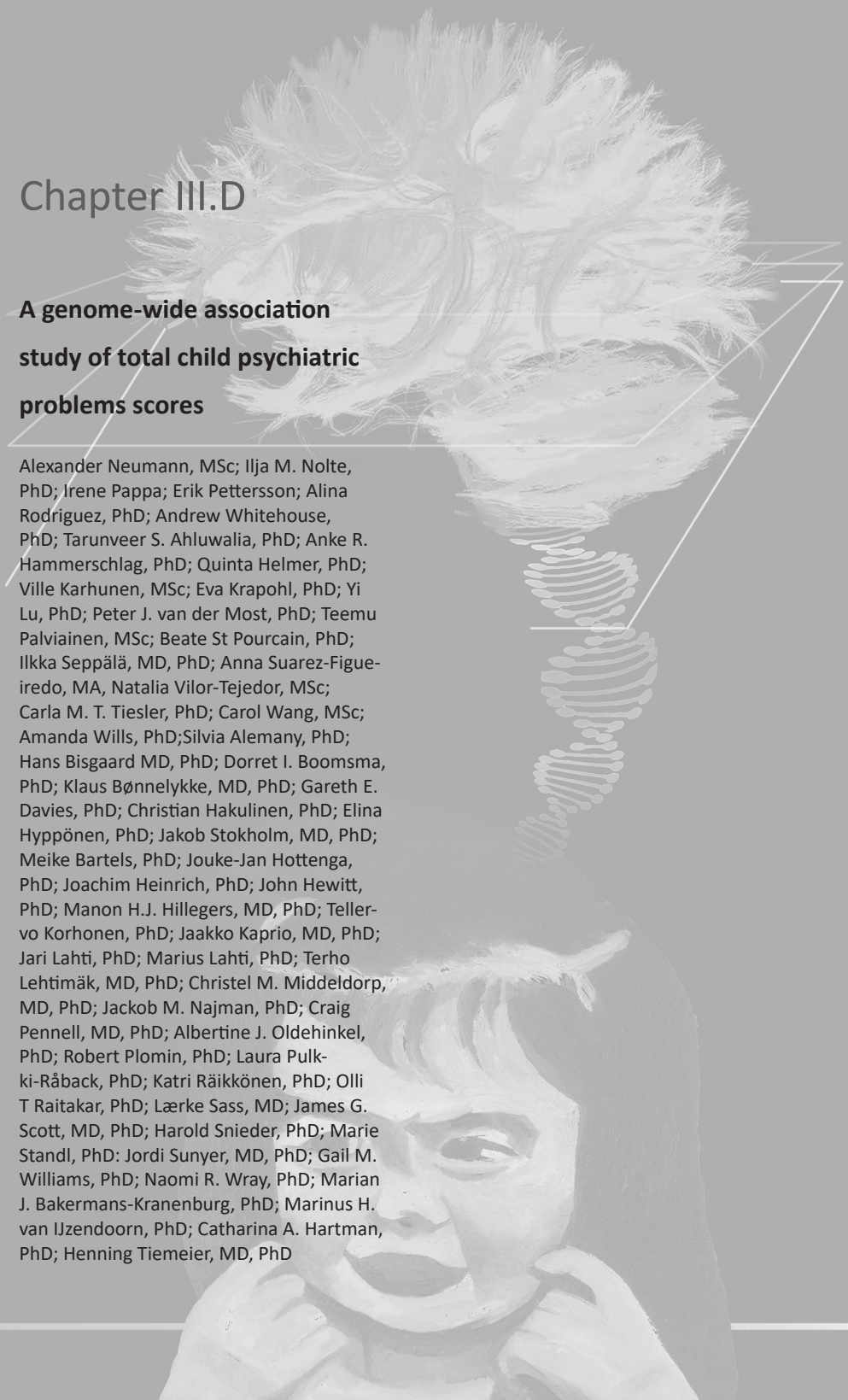
Chapter III.D

A genome-wide association study of total child psychiatric problems scores

Alexander Neumann, MSc; Ilja M. Nolte, PhD; Irene Pappa; Erik Pettersson; Alina Rodriguez, PhD; Andrew Whitehouse, PhD; Tarunveer S. Ahluwalia, PhD; Anke R. Hammerschlag, PhD; Quinta Helmer, PhD; Ville Karhunen, MSc; Eva Krapohl, PhD; Yi Lu, PhD; Peter J. van der Most, PhD; Teemu Palviainen, MSc; Beate St Pourcain, PhD; Ilkka Seppälä, MD, PhD; Anna Suarez-Figueiredo, MA, Natalia Vilor-Tejedor, MSc; Carla M. T. Tiesler, PhD; Carol Wang, MSc; Amanda Wills, PhD;Silvia Alemany, PhD; Hans Bisgaard MD, PhD; Dorret I. Boomsma, PhD; Klaus Bønnelykke, MD, PhD; Gareth E. Davies, PhD; Christian Hakulinen, PhD; Elina Hyppönen, PhD; Jakob Stokholm, MD, PhD; Meike Bartels, PhD; Jouke-Jan Hottenga, PhD; Joachim Heinrich, PhD; John Hewitt, PhD; Manon H.J. Hillegers, MD, PhD; Tellervo Korhonen, PhD; Jaakko Kaprio, MD, PhD; Jari Lahti, PhD; Marius Lahti, PhD; Terho Lehtimäk, MD, PhD; Christel M. Middeldorp, MD, PhD; Jackob M. Najman, PhD; Craig Pennell, MD, PhD; Albertine J. Oldehinkel, PhD; Robert Plomin, PhD; Laura Pulkki-Råback, PhD; Katri Räikkönen, PhD; Olli T Raitakar, PhD; Lærke Sass, MD; James G. Scott, MD, PhD; Harold Snieder, PhD; Marie Standl, PhD: Jordi Sunyer, MD, PhD; Gail M. Williams, PhD; Naomi R. Wray, PhD; Marian J. Bakermans-Kranenburg, PhD; Marinus H. van IJzendoorn, PhD; Catharina A. Hartman, PhD; Henning Tiemeier, MD, PhD



ABSTRACT

Previous research demonstrated substantial genetic correlations among psychiatric disorders and identified numerous cross-disorder genetic variants. To identify the genetic variants underlying global childhood psychopathology, we performed a genome-wide association study of a total psychiatric problem score. We analyzed 8,804,648 common SNPs in 29,446 school-aged children from 16 population-based cohorts participating in the EArly Genetics and Lifecourse Epidemiology (EAGLE) consortium. Gene-based analyses revealed, that the myotonic dystrophy (DM1) gene cluster, previously implicated in neurodevelopment, was associated with the total psychiatric problem score. No individual SNP reached genome-wide significance. Stronger association with the total psychiatric problems score at the gene level, corresponded to higher expression in brain tissue, in particular in limbic regions. The genetic effects underlying the total psychiatric problem score were shared with known genetic variants for common psychiatric disorders (ADHD, anxiety, depression, insomnia), intelligence, brain structure, educational attainment, wellbeing, insomnia, smoking and body fat ($rG > 0.23$), but not with less common disorders (schizophrenia, bipolar disorder, autism, or eating disorders) ($rG < 0.05$). In summary, the results suggest that many common genetic variants are non-specifically associated with various child psychiatric symptoms and related phenotypes, with the myotonic dystrophy gene cluster showing the most evidence. Further research is needed to establish causality and pleiotropic mechanisms.

INTRODUCTION

Psychiatric traits are moderately heritable, on average about 30-50% of the variability in symptoms can be explained by genetic differences between individuals.[1] Molecular genetic studies have established that to a large degree common genetic variants underlie these genetic effects. The joint effect of single nucleotide polymorphisms (SNP heritability) explains 5% to 30% of the variance in psychiatric disorders[2] and similarly of behavioral and emotional symptoms in children, although this varies depending on child age and informant.[3,4]

Recent family and molecular genetic studies have also demonstrated that much of the genetic effects underlying psychiatric disorders are not unique to particular diagnoses, but rather shared among several psychiatric diagnoses and symptoms.[2,5–9] This phenomenon is known as cross-phenotype association and suggests pleiotropy, i.e. the influence of a genetic variant on multiple traits.10 Several lines of evidence provide support for this notion, first that the SNP based genetic correlations between disorders from different domains, such as major depression, ADHD, bipolar disorder and schizophrenia are moderate to high,2 averaging 0.41. Second, measures of global psychopathology in children showed a common SNP heritability between 16% and 38%.[7,11] Third, in a GWAS meta-analysis of five disorders, loci were identified with associations across these disorders.[12]

The large extent of cross-phenotype associations presents a challenge in interpreting genome-wide association studies (GWAS), which typically analyze a single disorder to identify genetic loci. For instance, recently Brikell et al.[13] found that a polygenic score derived from a GWAS on ADHD symptoms was more strongly associated with a general psychopathology factor, than with specific hyperactivity or attention problems once general psychopathology was accounted for. This and the other aforementioned findings illustrate that multi-disorder approaches may benefit genetic studies.

Previous GWAS of child disorder, such as autism spectrum disorders, ADHD, aggression and internalizing disorders,[4,14–16] have been successful in providing insights into the genetic architecture of child psychiatric problems and into the genetic correlations between child psychiatric problems. However, with the notable exception of a very large recent ADHD study[17] and a GWAS on autism spectrum disorder14, these studies have mostly failed to identify individual genome-wide significant loci. Along with increasing the sample size, some researchers propose the inclusion of related phenotypes in analyses to increase power.[18,19] Genetic loci with pleiotropic effects may be missed in a GWAS of single psychiatric disorders. While a variant may modestly increase the risk of symptoms from different domains and be detectable in very large studies, any association with a specific disorder or trait may not be consistent enough or too weak to be detected in smaller studies. A focus on global psychopathology should therefore increase the power to detect unspecific genetic loci, which are associated with global psychiatric

vulnerabilities. While one previous GWAS[12] examined multiple disorders simultaneously, analyses on multiple dimensional measures of psychiatric problems in childhood are lacking. A focus on childhood problems is particularly important, as the incidence of many disorders changes across the life course. Moreover, childhood shows high heterotypic sequential continuity[20], thus a GWAS of total psychiatric problem scores in childhood can contribute to the understanding of the development of psychopathology.

Our aim was to identify genetic loci associated with a total psychiatric problem score including internalizing, externalizing, attention, neurodevelopmental and other psychiatric problems. We hypothesized that a large number of genetic loci may affect psychopathology in school-aged children. To identify these genetic variants, we performed a GWAS meta-analysis within the EArly Genetics and Lifecourse Epidemiology (EAGLE) consortium. In addition, we aggregated the SNP results to perform gene-based tests, which may have better power to identify loci. We also performed several follow-up analyses such as gene expression analyses, which may hint at which causal pathways are involved. We hypothesized that the stronger a gene is associated with the total psychiatric problem score, the higher the expression in brain tissue. Finally, we tested genetic correlations, i.e. the correlation due to shared genetic variants of the total psychiatric problem score with various psychiatric, psychological, neurological and lifestyle or educational characteristics. The GWAS was aimed at detecting loci associated with global psychopathology, however, the same loci may also impact many other phenotypes related to problem behavior, either through independent pathways, because the psychiatric symptoms caused by the genetic variant have an influence on other phenotypes, or because the geneitc loci first influence a related phenotype, which then causes psychopathology. To test the genetic correlations, we used information from previous GWAS available on LD hub[35]. We selected all traits, which were either psychiatric, psychological, neurological, or are affected by behavior. As the information came from mostly independent study populations, the genetic correlations with psychiatric phenotypes also serve as a validation of the GWAS results.

METHODS

Participants

Cohorts from the EAGLE consortium were invited to participate in the project, if parent-rated measures of psychiatric symptoms in the age range 5-16 years were available. Sixteen cohorts from Europe, the US and Australia contributed data to this meta-analysis. To avoid biases arising from population stratification we restricted the analysis to children of European ancestry. Sixteen cohorts with in sum 29,446 participants were

Table 1: Phenotype Characteristics

Cohort	n	Instrument	Informant	Age in years	Age SD	Score Mean	Score SD	% Female
ALSPAC	5461	SDQ	Maternal	9.64	0.12	6.71	4.82	49
BREATHE	1618	SDQ	Both	8.31	3.87	8.05	5.05	48
CADD	358	CBCL 4-18	Both	12.96	2.64	16.22	21.9	28
CATSS	6498	A-TAC	Both	12	-	5.39	7.48	49
FiNNTWIN	959	MPNI	Both	11.40	0.27	11.34	6.80	53
GenR	1847	CBCL 6-18	Maternal	9.70	0.28	17.29	15.18	51
Gini-Lisa	1389	SDQ	Maternal	10.04	0.20	7.3	5.15	48
Glaku	312	CBCL	Maternal	12.1	1.0	21.7	16.8	52
INMA	745	SDQ	Both	5.1	0.78	8.87	5.00	38
NFBC1986	3346	Rutter	Maternal	7.81	0.23	2.57	2.07	51
NTR	2563	CBCL 6-18	Maternal	9.9	0.98	19.32	15.89	52
RAINE	1366	CBCL 4-18	Both	10.58	0.20	21.12	18.62	48
TCHAD	2111	CBCL 6-18	Both	13	-	11.66	12.49	51
TEDS	2707	SDQ	Both	11.26	0.69	7.00	5.02	54
TRAILS	1283	CBCL 6-18	Maternal	11.08	0.56	0.24	0.16	52
YFS	1352	HES	Maternal	10.55	3.33	14.65	6.75	54

ALSPAC Avon Longitudinal Study of Parents and Children
CADD Center on Antisocial Drug Dependence
CATSS The Child and Adolescent Twin Study in Sweden
GenR Generation R
INMA Infancia y Medio Ambiente
NFBC1986 Northern Finland Birth Cohorts
NTR Netherlands Twin Register
TCHAD Twin Study of Child and Adolescent Development
TEDS Twins Early Development Study
TRAILS Tracking Adolescents' Individual Lives Survey
YFS The Cardiovascular Risk in Young Finns Study

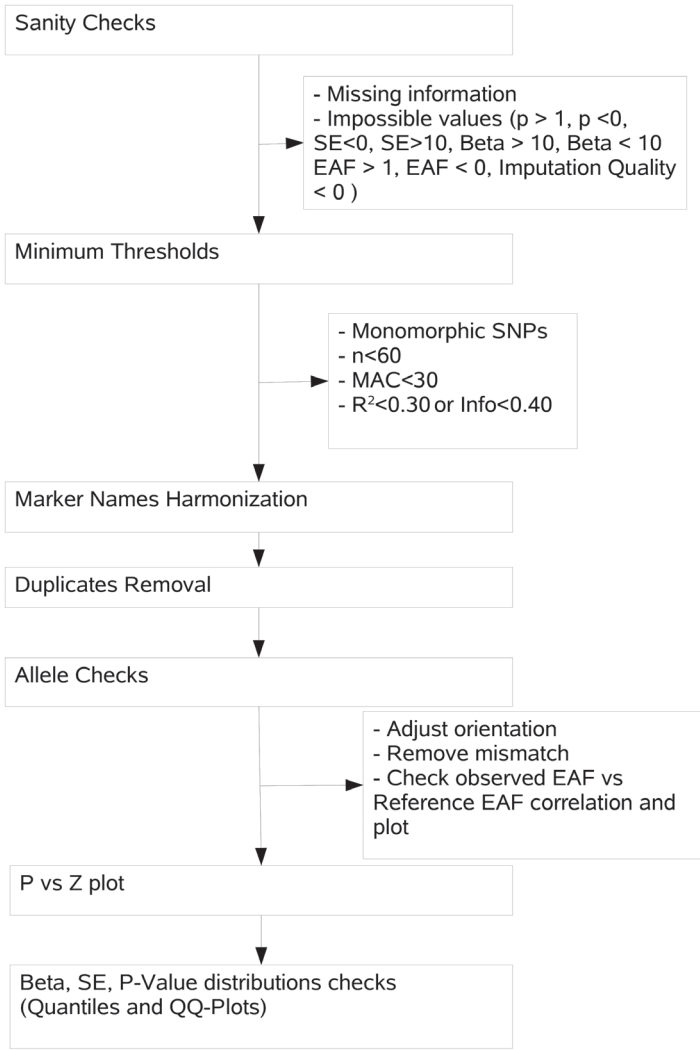


Figure 1: Quality control steps of individual cohort results.

meta-analyzed. Participants had mean ages of 9.9 years. See Table 1 for participant demographics per cohort.

Outcome

Psychiatric problems were assessed with parent-rated questionnaires at ages 5-16 years (mean=9.9, standard deviation (SD)=2.02) (Table 1). When information on psychopathology was available at multiple ages, we asked to use the assessment wave closest to the age of 10 as we aimed to investigate school-age psychopathology. All items of a broad psychiatric questionnaire were summed into a single total psychiatric sum score. Some questionnaires included items on sleep, thought, eating problems, and pervasive developmental disorders but in all cohorts at least internalizing, externalizing and attention problems were assessed. As expected in population-based cohorts, these total psychiatric problem scores had a skewed distribution. We applied a log transformation plus 1 to avoid bias due to non-normal residuals and influential observations, especially by low frequency variants. Since different scales were used, the log-transformed scores were converted to a z-score within cohorts to make units comparable across cohorts. Instruments included the Child Behavior Checklist[21], Strengths and Difficulties Questionnaire[22], Multidimensional Peer Nomination Inventory[23], Rutter Children' Behaviour Questionnaire[24], A-TAC[25], and items derived from the Health Examination Survey[26].

Genotyping and QC

Genotyping was performed in all cohorts using genome-wide genotyping arrays. Cohort-specific pre-imputation quality control was performed using established protocols. In all discovery cohorts, SNPs were imputed to the 1000 Genomes Phase 1 or Phase 3 reference panel.[27] Each cohort performed a GWAS and summary results were collected for meta-analysis. We omitted the X-chromosome from further analysis as most cohorts had no information on X-linked SNPs available. Pre-meta-analysis QC was performed with EasyQC.[28,29] Map and reference allele files were based on 1000G phase III version 5 data (Build 37). The QC steps are summarized in Figure 1. After meta-analysis, we excluded SNPs with low MAF (<5%), sample size (<5000) or a low numbers of cohorts with data on the tested SNP (<8). Finally, we checked the pooled results for spurious inflation by examining QQ-plots of the p distribution and by examining the LD score regression intercept (see statistical analysis).

Statistical analysis

Single SNP associations and meta-analysis

The z-scores of the total psychiatric problems scores were related to the SNP dosages in a linear model. Covariates included gender, age at assessment and principal components of ancestry. The number (0-10) of components were specified by cohort, depending on the extend of population stratification. CATSS and TCHAD additionally used a random effect to account for familial relatedness. FINNTWIN and NTR did not use

principal components, but applied a mixed model with two random effects to control for population stratification and relatedness. We pooled the results from the individual cohorts using a fixed-effects inverse variance meta-analysis. R 3.4.3 was used for QC, data preparation and analysis of results.[30] Meta-soft 2.0.1 was used for the meta-analysis of single SNP associations.[31] We based conclusions on a fixed effects model but also make available the results of a random effects model. For the latter, we used the Han and Eskin random effects approach[31], which does not assume that the true effect between studies is equal, but does assume homogeneity of null effects. Results, however, tended to be biased towards SNPs with information from fewer studies. Therefore we based our conclusions on the fixed effects model. Yet, the τ2 statistic obtained from the random effects model was used as measure of heterogeneity. The individual cohort results after quality control were examined and meta-analyzed independently by the first and second author with consistent results.

Gene-based and expression analysis

The results were explored with the FUMA web tool[32] First, we performed gene-based tests using MAGMA[33] in FUMA. MAGMA estimates the joint effect of all SNPs within a gene, while taking into account the LD structure, as well as gene size. We tested 18,168 protein coding genes and thus the genome-wide significant p-value was set at 2.752e-6 using Bonferroni correction. Second, we tested, whether the results from the gene-based tests are related to gene expression in several tissues. Specifically, we used MAGMA to test whether the association strength between genes and the total psychiatric problem score was related to the mean gene expression level in a specific tissue. Given that we expected genes to act via brain pathways, we tested expression in 13 brain regions (Table 4). However, as other pathways may also be possible, we also investigated gene expression levels on an organ level (Supplementary Table 3). Gene expression levels were obtained from the GTEx database, which is based on RNA sequence data.[34] FUMA uses expression values in reads per kilobase million (rpkm) after winsorization at 50 rpkm and log transformation after adding 1 to avoid zero values.

SNP heritability and genetic correlations

We estimated the SNP heritability of total psychiatric problem scores with LD score regression[35]. LD score regression is based on the assumption that SNPs with strong correlations with neighboring SNPs (high LD) are more likely to tag causal variants and thus have stronger associations with the phenotype. We used the online tool LD hub[36] to estimate common SNP heritability and genetic correlations with various psychiatric, psychological, neurological and lifestyle or educational characteristics. Genetic correlations are computed by regressing the product of two GWAS's association strength on the LD score. To compute the genetic correlations we used previously published GWAS summary statistics available on LD hub. Pubmed IDs for all traits are listed in Table 5, see these references for detailed genetic and phenotype methods of the correlates. LD

Table 2: SNPs with genome-wide suggestive (p<5E-06) results

SNP	Chr	BP	EA	OA	EAF	n_{stu}	n	β	SE	p
rs5837866	2	204457753	D	I	0.81	13	24864	0.06	0.01	6,00E-7
rs56324955	2	204461375	A	G	0.83	15	26544	0.06	0.01	8,00E-7
rs11189716	10	83164779	T	G	0.66	16	29257	0.04	0.01	1,00E-6
rs1763252	14	42577719	A	T	0.66	13	23849	-0.05	0.01	1,00E-6
rs4740712	9	3095668	G	C	0.08	15	26548	-0.08	0.02	1,00E-6
rs2070737	19	46282890	A	T	0.67	16	29238	-0.04	0.01	1,00E-6
rs8112096	19	5618508	T	G	0.96	11	22178	0.15	0.03	2,00E-6
rs74537605	2	36626313	A	C	0.98	10	22986	-0.18	0.04	2,00E-6
rs34820217	4	39357890	G	A	0.93	12	23544	-0.13	0.03	2,00E-6
rs56713366	6	148034509	C	T	0.93	13	25251	-0.1	0.02	2,00E-6
rs9564661	13	70718856	T	C	0.87	14	27057	-0.06	0.01	2,00E-6
rs1150459	5	5865408	A	C	0.63	16	29252	-0.04	0.01	2,00E-6
rs138233043	8	35693961	D	I	0.99	8	16867	-0.25	0.05	3,00E-6
rs725660	19	46262286	C	A	0.67	15	27887	-0.04	0.01	3,00E-6
rs11189717	10	83164890	G	T	0.66	14	27058	0.04	0.01	3,00E-6
rs2403298	11	18470082	G	A	0.40	14	24176	0.04	0.01	3,00E-6

Table 2: (Continued)

SNP	Chr	BP	EA	OA	EAF	n_{stu}	n	β	SE	p
rs9259913	6	29899653	T	C	0.92	11	20606	0.09	0.02	3,00E-6
rs10891091	11	110229530	C	T	0.93	15	26532	0.08	0.02	3,00E-6
rs35041294	11	78762524	G	A	0.20	12	22412	0.06	0.01	3,00E-6
rs35692111	11	78762526	C	A	0.20	12	22412	0.06	0.01	3,00E-6
rs4753887	11	110227485	T	C	0.93	15	26532	0.08	0.02	3,00E-6
rs3943756	11	78761301	T	C	0.19	14	24209	0.05	0.01	4,00E-6
rs7481948	11	78748152	G	T	0.19	14	24203	0.05	0.01	4,00E-6
rs6435192	2	204452173	T	C	0.81	14	26232	0.05	0.01	4,00E-6
rs2246727	5	5869657	A	G	0.63	16	29257	-0.04	0.01	4,00E-6
rs7938267	11	18432754	A	G	0.41	15	26516	0.04	0.01	4,00E-6
rs3858041	9	3084094	C	G	0.07	15	26545	-0.07	0.02	4,00E-6
rs4965595	15	100758977	A	G	0.68	14	27058	-0.05	0.01	4,00E-6
rs11213364	11	110227195	G	C	0.93	15	26531	0.08	0.02	4,00E-6
rs9676288	19	46292060	G	T	0.67	15	26530	-0.04	0.01	4,00E-6
rs7938942	11	18466247	C	T	0.40	14	24181	0.04	0.01	4,00E-6
rs17571725	19	46290822	C	T	0.67	15	26531	-0.04	0.01	4,00E-6
rs3904476	9	3081130	G	T	0.07	15	26544	-0.07	0.02	4,00E-6
rs149538014	6	148025311	G	A	0.93	10	22722	-0.11	0.02	5,00E-6
rs7256524	19	46224654	C	T	0.67	15	26509	-0.04	0.01	5,00E-6

Table 2: (Continued)

SNP	Chr	BP	EA	OA	EAF	n_{stu}	n	β	SE	p
rs3899008	9	3080960	T	C	0.07	15	26544	-0.07	0.02	5,00E-6

Chr Chromosome
BP Basepair
EA Effect Allele
OA Other Allele
EAF Effect Allele Frequency
n_{stu} Number of Studies
n Sample Size
β Beta
SE p p-value

Table 3: Genes with genome-wide significant (p<3e-6) or suggestive (p<3e-4) results

Gene	Chr	BP Start	BP Stop	n_{snps}	n	p
DMWD	19	46276205	46306060	48	25531	9,00E-7
DMPK	19	46262975	46295810	50	25255	1,00E-6
SIX5	19	46258043	46282484	35	25539	2,00E-6
TRAPPC6A	19	45656186	45691495	83	23784	7,00E-6
BLOC1S3	19	45672003	45695059	72	23884	2,00E-5
SLC40A1	2	190415305	190458484	97	25820	2,00E-5
TMEM56	1	95572894	95673163	294	26623	3,00E-5
NKPD1	19	45643008	45673408	60	24310	3,00E-5
FBXO46	19	46203887	46244162	68	24264	4,00E-5
TBCA	5	76976991	77174604	749	26665	7,00E-5
TRPS1	8	116410724	116831899	615	25482	8,00E-5
LRFN5	14	42066773	42383752	1086	25862	1,00E-4
RSPH6A	19	46288968	46328577	132	25712	1,00E-4
SLC25A26	3	66109285	66448540	775	24326	1,00E-4
LDHA	11	18405935	18439972	125	25821	1,00E-4
PAMR1	11	35443370	35561848	429	25908	2,00E-4
SLC9A3R2	16	2065357	2099027	86	21370	2,00E-4
SPATA7	14	88841268	88946694	136	25225	2,00E-4
RAB10	2	26246976	26370323	232	25544	3,00E-4
NPW	16	2049927	2080756	112	22389	3,00E-4
PTPN21	14	88922122	89031077	221	26175	3,00E-4
TMEM56-RWDD3	1	95573479	95722781	453	26725	3,00E-4
MARK4	19	45572546	45818541	777	24034	3,00E-4
DNASE2B	1	84854215	84890701	144	27431	3,00E-4

Chr Chromosome
BP Basepair
Start Basepair Position of gene start
Stop Basepair Position of gene end
$\mathbf{n}_{snps}$ Number of SNPs within the gene
n Number of participants
p p-value

Table 4: Tissue expression analysis (neural tissues)

Brain Region	β	β_{SD}	SE	p	q
Caudate basal ganglia	0.02	0.06	0.01	<0.01	0.02
Putamen basal ganglia	0.02	0.05	0.01	<0.01	0.02
Amygdala	0.02	0.05	0.01	<0.01	0.02
Nucleus accumbens basal ganglia	0.02	0.04	0.01	0.01	0.03
Hippocampus	0.02	0.04	0.01	0.01	0.03
Anterior cingulate cortex BA24	0.02	0.04	0.01	0.01	0.03
Substantia nigra	0.02	0.04	0.01	0.03	0.06
Hypothalamus	0.01	0.04	0.01	0.04	0.07
Frontal Cortex BA9	0.01	0.03	0.01	0.05	0.07
Cortex	0.01	0.03	0.01	0.05	0.07
Spinal cord cervical c1	0.01	0.03	0.01	0.06	0.07
Cerebellum	0.01	0.02	0.01	0.18	0.2
Cerebellar Hemisphere	0.00	0.01	0.01	0.22	0.22

β Beta
βSD Beta Standard Deviation
SE Standard Error
p P-value
q False Discovery Adjusted P-values

hub did not have summary statistics available for anxiety symptoms.[37] We therefore obtained this information from the psychiatric genetics consortium (https://www.med.unc.edu/pgc/results-and-downloads/downloads) and computed genetic correlations locally using ldsc 1.0.0 with the same settings as LD hub.

Heterogeneity and generalizability

The participating cohorts differed in instruments and genotyping platforms, in assessment age and population background. While only children with European ancestry were analyzed, they came from a variety of countries with different cultural backgrounds. The meta-analysis results may therefore vary depending on the exact composition of participating cohorts. To investigate how well the results generalize, we split the sample in two halves randomly, meta-analyzed both halves (eight cohorts each), and computed the genetic correlation between both halves. We repeated this step ten times and averaged the results. We also report the SNP heritability of these halves to assess the variability due to study selection.

RESULTS

Spurious inflation and SNP Heritability

We associated 8,804,648 SNPs with the z-score of the total psychiatric problem scores in 16 discovery cohorts in up to 29,446 children. The test statistics showed no spurious inflation according to a visual examination of a QQ-plot (see Figure 2) and the LD score regression intercept, which was close to 1 ($\beta 0 = 1.01$, standard error (SE)=0.02). The common SNP heritability was 8.4% (SE=0.02).

SNP and gene based tests

No SNP reached genome-wide significance, see Figure 3 for a Manhattan plot. 36 SNPs showed suggestive effects, defined as p<5E-06, see Table 2. Next we tested the association of 18,290 protein coding genes (as defined in FUMA) with the child total psychiatric problem score by aggregating the effects of the single SNPs. In contrast to the single SNP test, the MAGMA results showed higher number of positive test-results than expected under the null (Figure 4). Three genes within one locus on 19q13.32 reached genome-wide significance, DMWD, DMPK, and SIX5 (see Figure 5 for Manhattan plot and Table 3 for estimates).

Next we performed the MAGMA tissue expression analysis in 13 specific brain tissues (Table 4). Several subcortical structures survived multiple testing correction, specifically basal ganglia, amygdala, hippocampus and anterior cingulate. In addition we performed tissue expression analysis for 30 tissues on an organ level, see Supplementa-

Table 5: Genetic correlations based on LD score regression

Correlated trait	PMID	r_G	SE	p	q	h^2
Psychiatry						
ADHD	20732625	0.71	0.24	3,00E-3	9,00E-3	0.22
Insomnia	28604731	0.67	0.11	2,00E-9	2,00E-8	0.05
Depressive symptoms	27089181	0.56	0.11	3,00E-7	2,00E-6	0.05
Anxiety symptoms	26754954	0.49	0.23	3,00E-2	6,00E-2	0.03
Major depressive disorder	22472876	0.33	0.14	2,00E-2	5,00E-2	0.16
PGC cross-disorder analysis	23453885	0.13	0.09	2,00E-1	3,00E-1	0.17
Schizophrenia	25056061	-0.02	0.06	7,00E-1	8,00E-1	0.46
Autism spectrum disorder	28540026	-0.03	0.11	8,00E-1	8,00E-1	0.40
Bipolar disorder	21926972	-0.05	0.09	6,00E-1	7,00E-1	0.42
Anorexia Nervosa	24514567	-0.13	0.08	9,00E-2	1,00E-1	0.52
Neurology						

Table 5: (Continued)

Correlated trait	PMID	r_G	SE	p	q	h^2
Amyotrophic lateral sclerosis	27455348	0.16	0.17	3,00E-1	4,00E-1	0.05
Parkinsons disease	19915575	-0.03	0.10	8,00E-1	8,00E-1	0.35
Alzheimers disease	24162737	-0.14	0.16	4,00E-1	5,00E-1	0.04
Personality and Wellbeing						
Neuroticism	27089181	0.43	0.09	6,00E-7	3,00E-6	0.09
Neuroticism	24828478	0.30	0.15	4,00E-2	8,00E-2	0.01
Neo-openness to experience	21173776	0.05	0.15	7,00E-1	8,00E-1	0.11
Neo-conscientiousness	21173776	0.05	0.21	8,00E-1	8,00E-1	0.07
Subjective well being	27089181	-0.45	0.10	4,00E-6	2,00E-5	0.03

Table 5: (Continued)

Correlated trait	PMID	r_G	SE	p	q	h^2
Intelligence and educational attainment						
Childhood IQ	23358156	-0.37	0.13	5,00E-3	1,00E-2	0.28
College completion	23722424	-0.50	0.09	9,00E-9	7,00E-8	0.08
Years of schooling	27225129	-0.53	0.07	3,00E-14	10,00E-13	0.12
Intelligence	28530673	-0.54	0.09	7,00E-10	9,00E-9	0.19
Brain volume						
Mean Thalamus	25607358	0.01	0.17	1,00E+0	1,00E+0	0.13
Mean Pallidum	25607358	-0.04	0.15	8,00E-1	8,00E-1	0.17
Mean Caudate	25607358	-0.15	0.12	2,00E-1	3,00E-1	0.25
Mean Hippocampus	25607358	-0.16	0.16	3,00E-1	4,00E-1	0.13
Intracranial Volume	25607358	-0.20	0.16	2,00E-1	3,00E-1	0.17

Table 5: (Continued)

Correlated trait	PMID	r_G	SE	p	q	h^2
Mean Putamen	25607358	-0.29	0.12	1,00E-2	3,00E-2	0.29
Infant head circumference	22504419	-0.32	0.15	3,00E-2	6,00E-2	0.23
Mean Accumbens	25607358	-0.33	0.23	1,00E-1	2,00E-1	0.08
General health behaviors/outcomes						
Cigarettes smoked per day	20418890	0.55	0.18	3,00E-3	9,00E-3	0.06
Body fat	26833246	0.48	0.10	1,00E-6	5,00E-6	0.11
Body mass index	20935630	0.23	0.07	5,00E-4	2,00E-3	0.19
Sleep duration	27494321	-0.19	0.10	5,00E-2	8,00E-2	0.06
Age of smoking initiation	20418890	-0.34	0.15	3,00E-2	6,00E-2	0.07
Parent's age at death						
Mother's age at death	27015805	-0.26	0.13	5,00E-2	8,00E-2	0.04

Table 5: (Continued)

Correlated trait	PMID	r_G	SE	p	q	h^2
Father's age at death	27015805	-0.28	0.14	5,00E-2	8,00E-2	0.04
Reproduction						
Number of children ever born	27798627	0.27	0.09	3,00E-3	9,00E-3	0.03
Age of first birth	27798627	-0.50	0.08	1,00E-10	2,00E-9	0.06

PMID PubMed ID, **rG** Genetic Correlation, **SE** Standard Error, **p** P-value, **q** False Discovery Rate Adjusted P-values, **h^2** SNP heritability

ry Table 3. While expression in brain tissue reached nominal significance, expression in none of the other organs had statistically significant associations.

Genetic correlation

To quantify the extent to which the genetic associations of child psychiatric problems scores were shared with child and adult psychiatric, psychological, neurological and lifestyle or educational characteristics we performed genetic correlation analyses. After adjustment for false discovery rate, ADHD, insomnia, depressive symptoms, cigarettes smoked per day, body fat, neuroticism, major depressive disorder, number of children ever born and body mass index all showed positive genetic correlations between 0.23 and 0.71 with the total psychiatric problem score. Of these, the highest correlation of global psychopathology was with ADHD (Table 5). Mean putamen and accumbens volume, childhood IQ, subjective wellbeing, college completion, age of first birth, years of schooling and intelligence showed significant negative correlations with the total psychiatric problem score, ranging from -0.54 to -0.29. Of the psychiatric phenotypes tested, only the less common psychiatric disorders like schizophrenia, bipolar disorder, autism spectrum disorder, and anorexia showed no genetic correlation with the total psychiatric problem score.

Heterogeneity Analysis

We computed the genetic correlation between ten randomly split sets to assess heterogeneity between the cohorts in our GWAS (Supplementary Table S4). On average, the genetic correlation was 0.77 (SD=0.12), ranging from 0.66-0.98, indicating that heterogeneity of effects between cohorts was limited. The SNP heritability ranged from 8-16% and was on average 11% (SD=0.02).

DISCUSSION

The current study reports the first GWAS examining global psychopathology in children and the biggest GWAS to date examining a continuous measure of psychiatric symptoms in children. One locus including the genes *DMWD, DMPK* and *SIX5* reached genome-wide significance in a gene-based analysis, but no individual SNP was genome-wide significant. Stronger gene-based association with the total psychiatric problem score corresponded to higher expression in brain tissue, in particular limbic regions. The genetic effects underlying global psychopathology were shared with common psy-

chiatric disorders (ADHD, anxiety, depression, insomnia), but not with less common ones (schizophrenia, bipolar disorder, autism, eating disorders).

DMWD, DMPK and *SIX5* comprise a locus known as the myotonic dystrophy (DM1) gene cluster[38]. Myotonic dystrophy is a genetic disorder caused by a CTG repeat in DMPK leading to muscle wasting and function loss. In addition myotonic dystrophy is associated with neural and cognitive behavioral deficits.[38,39] *DMWD* is highly expressed in synapses and related to neurodevelopment in rats, and is therefore suspected to cause the neuropsychological problems in myotonic dystrophy.[38] Interestingly, *DMWD* was previously associated with Alzheimer disease, providing further evidence that the locus is important for neural functioning.[40] Our results suggest as well that this locus is not only involved in rare neurological disorders, but that common variants in the myotonic dystrophy gene cluster impact psychiatric problems together with many other genetic variants and environmental influences.

Not only are these three genes, which we discovered in the gene-based analyses, expressed in the brain, but the genome-wide gene expression analysis also suggests, that loci more strongly associated with the total psychiatric problem score are more likely to be expressed in the limbic system of the brain. The limbic system includes evolutionary preserved regions responsible for emotion regulation and motivation[41], which were previously implicated in affective disorders, ADHD and OCD.[42,43] The present study suggests that genetic variants affecting psychopathology in general are most likely to act via pathways involving limbic structures and therefore supports the notion, that variation in this brain region may act as intermediate phenotype of psychopathology and potential target for intervention[44].

The total psychiatric problem scores were based on various instruments, which all included items for common psychiatric internalizing, attention, and externalizing symptoms. Therefore, it is not surprising that common psychiatric symptoms and disorders such as ADHD, anxiety and depression shared 25% or more of the genetic variation with the total psychiatric problem score. ADHD showed the highest genetic correlation, perhaps because ADHD symptoms are well covered in the instruments tested and because the source GWAS was well powered to estimated the SNP effects precisely. In contrast, the extent to which the questionnaires used in this study covered other less common problems, such as psychotic, bipolar or autistic symptoms varied greatly by instrument, with e.g. the SDQ only covering emotional, conduct, hyperactivity and peer relationship problems, but not thought disorders, such as the CBCL. Furthermore, age of onset for schizophrenia and bipolar disorder is typically in adolescence and adulthood.[45,46] In the case of autism spectrum disorder the age of onset is early, but the prevalence in the cohorts low. Thus the total psychiatric problem score was very broad but not representative of psychiatric disorders with lower prevalence rates or emergence at later ages. This most likely explains why the genetic correlations with the less common psychiatric disorders autism, schizophrenia and bipolar disorder with the total psychiatric problem score measured at age 10 were very low to absent. The differential genetic correlations

with common and not the relatively rare disorders suggests that there is a continuum of genetic effects from very specific variants, variants which underlie either common or less common disorders, to variants which underlie most psychiatric problems.[2,8] The latter set of variants may be better detected with measures of global psychopathology in older children, when thought disorders such as schizophrenia and bipolar disorder occur.

In this study we observed 8% SNP heritability, which is equivalent to the heritability of ADHD symptoms in children4, but somewhat higher than the 5% heritability for depression symptoms47 and 7% for anxiety symptoms[37] in adulthood. Previous studies observed heritability estimates to increase from childhood to adulthood for most psychiatric symptoms, such as anxiety (3% per year), depression and externalizing behaviors (1% per year for both).[48] An exception appears to be ADHD heritability, which tends to decrease somewhat across age.[49,50] Explanations for this increase with age include gene-environment correlation effects and timing differences of puberty in dizygotic boy/girl twin pairs, though the latter would not impact GWAS studies with unrelated participants.[48] Importantly, the lack of precision resulting from the contextual effects of ratings from an external observer, i.e. the parent, which are inherent to child psychiatric assessments may also play a role in the increase of heritabilities from childhood to adulthood. Further research is needed to investigate whether the heritability of a total psychiatric problem score is age dependent similarly to affective and externalizing problems and perhaps should be explicitly modeled in a longitudinal GWAS design.

Next to high genetic correlations with common psychiatric disorders and symptoms, we also observed high genetic correlations with phenotypes such as intelligence, educational attainment, wellbeing, insomnia, smoking and body fat. For illustration, the results suggest that more than 20% of the genetic effects underlying total psychiatric problem scores in children are shared with the amount of body fat individuals have or the number of cigarettes they smoke at later ages. Negative genetic correlations with mean putamen and nucleus accumbens volume suggest these structures as potential intermediate phenotypes. Alternative explanations would be, that these regions are affected by psychiatric symptoms or that they are independent outcomes of the genetic effects.

A limitation of this study is the large heterogeneity in methods. The cohorts included in this study used different genotyping arrays and different instruments to measure psychopathology. On the one hand, the variety of methods is an advantage, since detected associations are expected to be more generalizable. Genetic correlations calculated for random splits of the sample in half generally were high, indicating that similar results on a genome-wide level for both subsamples were found and therefore suggesting that the genetic effects in the full sample are generalizable. On the other hand, the high heterogeneity might reduce power and limit the detectability of less robustly associated variants. The gene-based tests showed better test results than expected under the null, which may indicate bias or power. In this GWAS the higher lambda of the gene-based

tests is more likely the result of less tests to correct for and hence higher power, given that the SNP results did not indicate biases due to population stratification, so a true polygenic signal is likely. Finally, as in any other GWAS study, the extent to which the found associations can be interpreted causally is difficult. Due to linkage disequilibrium it is unclear whether all three of the identified genes (DMWD, DMPK and SIX5) have a causal influence on psychopathology, only a subset or perhaps none, as they may be marker for nearby causal variants. However, as discussed, given the prior literature on the functioning of the genes, a direct causal influence seems plausible.

In conclusion, the GWAS of total psychiatric problem scores supports the notion that part of psychiatric genetics is shared among internalizing, externalizing, attention and other psychiatric problems in childhood. Moreover, the results suggest that genes involved in the total psychiatric problem scores act mostly via neural pathways, in particular limbic regions. The pleiotropy was not restricted to psychiatric phenotypes, but also included intelligence, neural features, educational attainment, wellbeing, insomnia, smoking, body fat and reproduction. Interestingly, we did not find shared genetic effects with autism, schizophrenia and bipolar disorder. We identified the myotonic dystrophy gene cluster as locus associated with child psychiatric problems in general, however, further investigation is needed to confirm these findings and explore potential causal mechanisms.

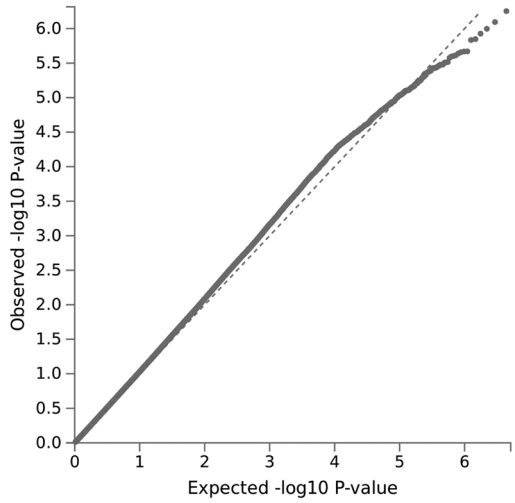


Figure 2: Quantile-quantile plot of observed -log 10 p values vs expected -log 10 p values assuming chance findings in single SNP analysis. Diagonal line indicates a p value distribution compatible with chance finding. Upward deviations indicate p values more significant than expected.

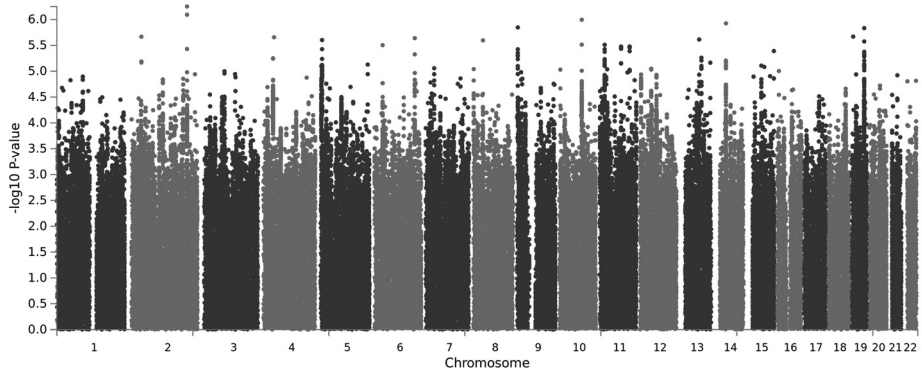


Figure 3: Manhattan plot of -log 10 p values vs SNP position for single SNP analysis.

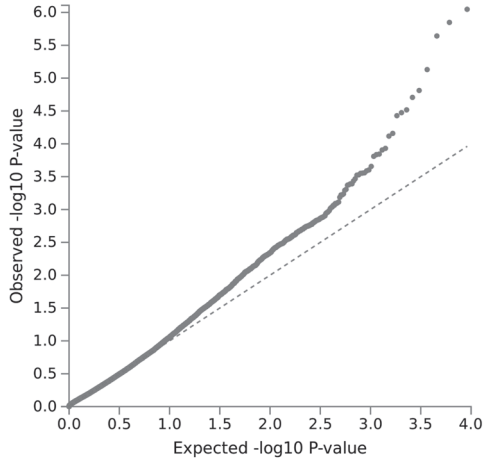


Figure 4: Quantile-quantile plot of observed -log 10 p values vs expected -log 10 p values assuming chance findings in gene based analysis. Diagonal line indicates a p value distribution compatible with chance finding. Upward deviations indicate p values more significant than expected.

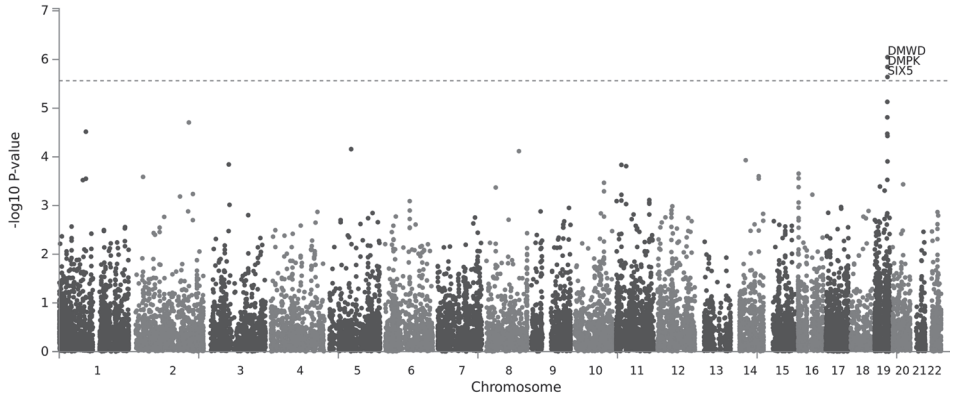


Figure 5: Manhattan plot of -log 10 p values vs SNP position for gene based analysis. Genes above the red horizontal line indicate genome-wide significant findings.

REFERENCES

1. Polderman TJC, Benyamin B, de Leeuw CA, et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nat Genet. 2015;47(7):702-709. doi:10.1038/ng.3285
2. Lee SH, Ripke S, Neale BM, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet. 2013;45(9):984-994. doi:10.1038/ng.2711
3. Pappa I, Fedko IO, Mileva-Seitz VR, et al. Single Nucleotide Polymorphism Heritability of Behavior Problems in Childhood: Genome-Wide Complex Trait Analysis. J Am Acad Child Adolesc Psychiatry. 2015;54(9):737-744. doi:10.1016/j.jaac.2015.06.004
4. Middeldorp CM, Hammerschlag AR, Ouwens KG, et al. A Genome-Wide Association Meta-Analysis of Attention-Deficit/Hyperactivity Disorder Symptoms in Population-Based Pediatric Cohorts. J Am Acad CHILD Adolesc PSYCHIATRY. 2016;55(10):896-905. doi:10.1016/j.jaac.2016.05.025
5. Lahey BB, Van Hulle CA, Singh AL, Waldman ID, Rathouz PJ. Higher-order genetic and environmental structure of prevalent forms of child and adolescent psychopathology. Arch Gen Psychiatry. 2011;68(2):181-189. doi:10.1001/archgenpsychiatry.2010.192
6. Spatola CAM, Fagnani C, Pesenti-Gritti P, Ogliari A, Stazi M-A, Battaglia M. A general population twin study of the CBCL/6-18 DSM-oriented scales. J Am Acad Child Adolesc Psychiatry. 2007;46(5):619-627. doi:10.1097/CHI.0b013e3180335b12
7. Neumann A, Pappa I, Lahey BB, et al. Single Nucleotide Polymorphism Heritability of a General Psychopathology Factor in Children. J Am Acad Child Adolesc Psychiatry. 2016;55(12):1038-1045. doi:10.1016/j.jaac.2016.09.498
8. Anttila V, Bulik-Sullivan B, Finucane HK, et al. Analysis of Shared Heritability in Common Disorders of the Brain. Cold Spring Harbor Labs Journals; 2016. doi:10.1101/048991
9. Group C, Consortium PG. Supplementary appendix. 2013;6736(12).
10. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet. 2013;14(7):483-495. doi:10.1038/nrg3461
11. Alnæs D, Kaufmann T, Doan NT, et al. Association of Heritable Cognitive Ability and Psychopathology With White Matter Properties in Children and Adolescents. JAMA Psychiatry. 2018;75(3):287-295. doi:10.1001/jamapsychiatry.2017.4277
12. Consortium C-DG of the PG. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet. 2013;381(9875):1371-1379. doi:10.1016/S0140-6736(12)62129-1
13. Brikell I, Larsson H, Lu Y, et al. The contribution of common genetic risk variants for ADHD to a general factor of childhood psychopathology. bioRxiv. 2017. doi:http://dx.doi.org/10.1101/193573
14. The Autism Spectrum Disorders Working Group of The Psychiatric Genomics. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Mol Autism. 2017;8(1):21. doi:10.1186/s13229-017-0137-9
15. Benke KS, Nivard MG, Velders FP, et al. A genome-wide association meta-analysis of preschool internalizing problems. J Am Acad Child Adolesc Psychiatry. 2014;53(6):667-676.e7. doi:10.1016/j.jaac.2013.12.028

16. Pappa I, St Pourcain B, Benke K, et al. A genome-wide approach to children's aggressive behavior: The EAGLE consortium. Am J Med Genet Part B Neuropsychiatr Genet. 2016;171(5):562-572. doi:10.1002/ajmg.b.32333
17. Demontis D, Walters RK, Martin J, et al. Discovery Of The First Genome-Wide Significant Risk Loci For ADHD. bioRxiv. 2017:145581. doi:10.1101/145581
18. Galesloot TE, van Steen K, Kiemeney L a LM, Janss LL, Vermeulen SH. A comparison of multivariate genome-wide association methods. PLoS One. 2014;9(4):e95923. doi:10.1371/journal.pone.0095923
19. Wang Y, Thompson WK, Schork AJ, et al. Leveraging Genomic Annotations and Pleiotropic Enrichment for Improved Replication Rates in Schizophrenia GWAS. PLOS Genet. 2016;12(1):1-22. doi:10.1371/journal.pgen.1005803
20. Shevlin M, McElroy E, Murphy J. Homotypic and heterotypic psychopathological continuity: a child cohort study. Soc Psychiatry Psychiatr Epidemiol. 2017;52(9):1135-1145. doi:10.1007/s00127-017-1396-7
21. Achenbach TM, Rescorla LA. Manual for the ASEBA School-Age Forms and Profiles. . (University of Vermont Youth, and Families RC for C, ed.). Burlington, VT; 2001.
22. Goodman R. The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry. 1997;38(5):581-586. doi:10.1111/j.1469-7610.1997.tb01545.x
23. Pulkkinen L, Kaprio J, Rose RJ. Peers, teachers and parents as assessors of the behavioural and emotional problems of twins and their adjustment: the Multidimensional Peer Nomination Inventory. Twin Res. 1999;2(4):274-285. doi:10.1375/twin.2.4.274
24. Rutter M. A CHILDREN'S BEHAVIOUR QUESTIONNAIRE FOR COMPLETION BY TEACHERS: PRELIMINARY FINDINGS. J Child Psychol Psychiatry. 1967;8(1):1-11. doi:10.1111/j.1469-7610.1967.tb02175.x
25. Hansson SL, Röjvall AS, Rastam M, Gillberg C, Gillberg C, Anckarsäter H. Psychiatric telephone interview with parents for screening of childhood autism - Tics, attention-deficit hyperactivity disorder and other comorbidities (A-TAC): Preliminary reliability and validity. Br J Psychiatry. 2005;187(SEPT.):262-267. doi:10.1192/bjp.187.3.262
26. Wells E. Behavioral patterns of children in school. Vitality Heal Stat. 1980;77:113.
27. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526(7571):68-74. doi:10.1038/nature15393
28. Thomas Winkler. EasyQC. www.genepi-regensburg.de/easyqc. Accessed December 21, 2017.
29. Winkler TW, Day FR, Croteau-Chonka DC, et al. Quality control and conduct of genome-wide association meta-analyses. Nat Protoc. 2014;9(5):1192-1212. doi:10.1038/nprot.2014.071
30. R Core Team. R: A Language and Environment for Statistical Computing. 2016. https://www.r-project.org/.
31. Han B, Eskin E. Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. Am J Hum Genet. 2011;88(5):586-598. doi:10.1016/j.ajhg.2011.04.014
32. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8(1):1826. doi:10.1038/s41467-017-01261-5
33. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLoS Comput Biol. 2015;11(4):1-19. doi:10.1371/journal.pcbi.1004219

34. Lonsdale J, Thomas J, Salvatore M, et al. The genotype-tissue expression (GTEx) project. Nat Genet. 2013;45(6):580-585.
35. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47(3):291-295. doi:10.1038/ng.3211
36. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics. 2017;33(2):272-279. doi:10.1093/bioinformatics/btw613
37. Otowa T, Hek K, Lee M, et al. Meta-analysis of genome-wide association studies of anxiety disorders. Mol Psychiatry. 2016;21:1391. http://dx.doi.org/10.1038/mp.2015.197.
38. Westerlaken JHAM, Van Der Zee CEEM, Peters W, Wieringa B. The DMWD protein from the myotonic dystrophy (DM1) gene region is developmentally regulated and is present most prominently in synapse-dense brain areas. Brain Res. 2003;971(1):116-127. doi:10.1016/S0006-8993(03)02430-2
39. Machuca-Tzili L, Brook D, Hilton-Jones D. Clinical and molecular aspects of the myotonic dystrophies: A review. Muscle and Nerve. 2005;32(1):1-18. doi:10.1002/mus.20301
40. Ghanbari M, Ikram MA, Looper HWJ De, et al. Genome-wide identification of microRNA-related variants associated with risk of Alzheimer's disease. Sci Rep. 2016;(January):1-9. doi:10.1038/srep28387
41. Morgane PJ, Galler JR, Mokler DJ. A review of systems and networks of the limbic forebrain/limbic midbrain. Prog Neurobiol. 2005;75(2):143-160. doi:10.1016/j.pneurobio.2005.01.001
42. Bora E, Fornito A, Pantelis C, Yücel M. Gray matter abnormalities in Major Depressive Disorder: A meta-analysis of voxel based morphometry studies. J Affect Disord. 2011. doi:10.1016/j.jad.2011.03.049
43. Norman LJ, Carlisi C, Lukito S, et al. Structural and Functional Brain Abnormalities in Attention-Deficit/Hyperactivity Disorder and Obsessive-Compulsive Disorder: A Comparative Meta-analysis. JAMA Psychiatry. 2017;73(8):815-825. doi:10.1001/jamapsychiatry.2016.0700
44. Buchheim A, Viviani R, Kessler H, et al. Changes in Prefrontal-Limbic Function in Major Depression after 15 Months of Long-Term Psychotherapy. Bruce A, ed. PLoS One. 2012;7(3):e33745. doi:10.1371/journal.pone.0033745
45. McGrath JJ, Saha S, Al-Hamzawi AO, et al. Age of onset and lifetime projected risk of psychotic experiences: Cross-national data from the world mental health survey. Schizophr Bull. 2016;42(4):933-941. doi:10.1093/schbul/sbw011
46. Post RM, Altshuler LL, Kupka R, et al. Age of onset of bipolar disorder: Combined effect of childhood adversity and familial loading of psychiatric disorders. J Psychiatr Res. 2016;81:63-70. doi:10.1016/j.jpsychires.2016.06.008
47. Okbay A, Baselmans BML, De Neve JE, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nat Genet. 2016;48(6):624-633. doi:10.1038/ng.3552
48. Bergen SE, Gardner CO, Kendler KS. Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood. Twin Res Hum Genet. 2007;10(Journal Article):423-433.
49. Larsson H, Chang Z, D'Onofrio BM, Lichtenstein P. The heritability of clinically diagnosed attention deficit hyperactivity disorder across the lifespan. Psychol Med. 2014;44(10):2223-

2229. doi:10.1017/S0033291713002493

50. Keyes CLM, Myers JM, Kendler KS. The structure of the genetic and environmental influences on mental well-being. Am J Public Health. 2010;100(12):2379-2384. doi:10.2105/AJPH.2010.193615

SUPPLEMENTARY MATERIAL

COHORT-SPECIFIC METHODS

ALSPAC Ethical approval for the study was obtained from the ALSPAC[1] Ethics and Law Committee and the Local Research Ethics Committees

http://www.bristol.ac.uk/alspac/, 1. Boyd A, Golding J, Macleod J, et al. Cohort profile: The 'Children of the 90s'-The index offspring of the avon longitudinal study of parents and children. Int J Epidemiol. 2013;42(1):111-127.

BREATHE The BREATHE[2] project (European Commission: FP7-ERC-2010-AdG, ID 268479) is a population-based cohort of primary schoolchildren designed to analyze the association between air pollution and behavior, cognitive function and brain morphology (Sunyer et al. 2015). Thirty-six of the 416 schools in Barcelona were selected based on modeled levels of traffic-related nitrogen dioxide. Thirty-eight schools were located in Barcelona and one school was in an adjacent municipality, Sant Cugat del Vallés. All families of children without special needs who were enrolled in 2nd, 3rd, and 4th grades at the selected schools were invited to participate in the study (2012). A total of 2897 children aged 7 to 11 years accepted the invitation and participated in the project. Genotype data were available for 1667 children of European ethnic origin. All parents or legal guardians gave written informed consent, and the study was approved by the IMIM-Parc de Salut Mar Research Ethics Committee (No. 2010/41221/I), Barcelona, Spain; and the FP7-ERC-2010-AdG Ethics Review Committee (268479-22022011).

http://www.creal.cat/projectebreathe/descripcion.html, 2. Sunyer J, Esnaola M, Alvarez-Pedrerol M, et al. Association between Traffic-Related Air Pollution in Schools and Cognitive Development in Primary School Children: A Prospective Cohort Study. PLoS Med. 2015;12(3):1-24.

CADD CADD[3] is a longitudinal study of adolescent substance use and associated comorbid conditions in clinical and community cases. Genotyping was done on 1901 subjects, of which 358 had parental CBCL data, and met inclusion criteria.

https://www.colorado.edu/ibg/human-research-studies/center-antisocial-drug-dependence, 3. Derringer J, Corley RP, Haberstick BC, Young SE, Demmitt BA, Howrigan DP, Kirkpatrick RM, Iacono WG, McGue M, Keller MC, Brown S, Tapert S, Hopfer CJ, Stallings MC, Crowley TJ, Rhee SH, Krauter K, Hewitt JK, McQueen MB (2015). Genome-Wide Association Study of Behavioral Disinhibition in a Selected Adolescent Sample. Behavior Genetic, 45(4):375-381

CATSS The Child and Adolescent Twin Study in Sweden (CATSS)[4] is an ongoing longitudinal twin study targeting all twins born in Sweden since July 1, 1992. Subjects are protected by the informed consent process, in which they are informed of what is being collected and repeatedly given the option to withdraw their consent and discontinue their participation. The CATSS-9/12 study has ethical approval from the Karolinska Institute Ethical Review Board: Dnr 03-672 and 2010/507-31/1, CATSS-9 – clinical

2010/1099-31/3 CATSS-15 Dnr: 2009/1599-32/5, CATSS-15/DOGSS Dnr: 03-672 and 2010/1356/31/1, and CATSS-18 Dnr: 2010/1410/31/1.

https://ki.se/meb/catss, 4. Anckarsäter H, Lundström S, Kollberg L, et al. The child and adolescent twin study in Sweden (CATSS). Twin Res Hum Genet. 2011;14(6):495-508.

FiNNTWIN FinnTwin12[5] is a longitudinal twin study launched in 1994 to investigate the developmental epidemiology of health-related behaviors. From 1994 to 1998, all Finnish families with twins born in 1983–1987 were identified from Finland's Population Register Centre and enrolled into a two-stage sampling design. The first stage included questionnaire assessments of all twins and parents at baseline (87% participation rate, 2724 families) conducted during the year in which the consecutive twin cohorts reached age 11, with follow-up of all twins at ages of 14 and 17½ years, and as young adults (age 22). Data collection procedures were approved by the Ethics Committee of Helsinki University, Finland and the Institutional Review Board of Indiana University, Bloomington, USA.

5. Rose RJ, Dick DM, Viken RJ, Pulkkinen L, Kaprio J. Drinking or abstaining at age 14? A genetic epidemiological study. Alcohol Clin Exp Res. 2001;25(11):1594-1604. doi:10.1111/j.1530-0277.2001.tb02166.x

GenR Generation R[6] is a population-based birth cohort aiming to identify early environmental and genetic determinants of development and health. All parents gave informed consent for their children's participation. The Generation R Study is conducted in accordance with the World Medical Association Declaration of Helsinki and study protocols have been approved by the Medical Ethics Committee of the Erasmus Medical Center, Rotterdam.

www.generationr.nl, 6. Kooijman MN, Kruithof CJ, van Duijn CM, et al. The Generation R Study: design and cohort update 2017. Eur J Epidemiol. 2016;31(12):1243-1264. doi:10.1007/s10654-016-0224-9

Gini-Lisa The influence of Life-style factors on the development of the Immune System and Allergies in East and West Germany PLUS the influence of traffic emissions and genetics (LISAplus) Study is a population based birth cohort study.[7,8] A total of 3094 healthy, full-term neonates were recruited between 1997 and 1999 in Munich, Leipzig, Wesel and Bad Honnef. The participants were not preselected based on family history of allergic diseases. A total of 5991 mothers and their newborns were recruited into the German Infant study on the influence of Nutrition Intervention PLUS environmental and genetic influences on allergy development (GINIplus) between September 1995 and June 1998 in Munich and Wesel. Infants with at least one allergic parent and/or sibling were allocated to the interventional study arm investigating the effect of different hydrolysed formulas for allergy prevention in the first year of life. All children without a family history of allergic diseases and children whose parents did not give consent for the intervention were allocated to the non-interventional arm. DNA was collected at the age 6 and 10 years and 1511 children from the Munich study center from both studies were genotyped. For both studies, approval by the local Ethics

Committees and written consent from participant's families were obtained.

www.helmholtz-muenchen.de/epi/research/research-groups/allergy-epidemiology/projects/giniplus/index.html, 7. Heinrich J, Bolte G, Hölscher B, et al. Allergens and endotoxin on mothers' mattresses and total immunoglobulin E in cord blood of neonates. Eur Respir J. 2002;20(3):617-623. , 8. Strandberg TE, Järvenpää AL, Vanhanen H, McKeigue PM. Birth outcome in relation to licorice consumption during pregnancy. Am J Epidemiol. 2001;153(11):1085-1088. doi:10.1093/aje/153.11.1085

Glaku The adolescents came from an urban community-based cohort comprising 1049 infants born between March and November 1998 in Helsinki, Finland.[9] Ethics Committees of the City of Helsinki Health Department and Children's Hospital in Helsinki University Central Hospital approved the study protocol. Each child and her/ his parent(s) provided their written informed consent at both follow-ups.

9. Guxens M, Ballester F, Espada M, et al. Cohort profile: The INMA-INfancia y Medio Ambiente-(environment and childhood) project. Int J Epidemiol. 2012;41(4):930-940. doi:10.1093/ije/dyr054

INMA The INMA—INfancia y Medio Ambiente—(Environment and Childhood) Project[10] is a network of birth cohorts in Spain that aim to study the role of environmental pollutants in air, water and diet during pregnancy and early childhood in relation to child growth and development (http://www.proyectoinma.org/) (Guxens et al. 2012). The study has been approved by Ethical Committee of each participating centre and written consent was obtained from participating parents. Data for this study comes from INMA Sabadell and Valencia subcohorts.

http://www.proyectoinma.org/, 10. Jarvelin M-R, Elliott P, Kleinschmidt I, et al. Ecological and individual predictors of birthweight in a northern Finland birth cohort 1986. Paediatr Perinat Epidemiol. 1997;11:298-312.

NFBC1986 NFBC1986[11] is a population-based birth cohort aiming to identify early environmental and genetic determinants of development and health. All parents gave informed consent for their children's participation. NFBC1986 is conducted in accordance with the World Medical Association Declaration of Helsinki and study protocols have been approved by the Ethics Committee of Northern Ostrobotnia Hospital District, Finland

www.oulu.fi/nfbc, 11. \Van Beijsterveldt CEM, Groen-Blokhuis M, Hottenga JJ, et al. The young Netherlands twin register (YNTR): Longitudinal twin and family studies in over 70,000 children. Twin Res Hum Genet. 2013;16(1):252-267. doi:10.1017/thg.2012.118

NTR NTR[12] is a population-based twin cohort of twins registered shortly after birth by their parents.

http://www.tweelingenregister.org/, 12. Van Beijsterveldt CEM, Groen-Blokhuis M, Hottenga JJ, et al. The young Netherlands twin register (YNTR): Longitudinal twin and family studies in over 70,000 children. Twin Res Hum Genet. 2013;16(1):252-267.

RAINE The Western Australian Pregnancy Cohort (RAINE) Study[13] is a longitudinal study of 2900 pregnant women and their offspring consecutively recruited from maternity units between 1989 and 1991(2). The inclusion criteria were (i) English lan-

guage skills sufficient to understand the study demands, (ii) an expectation to deliver at King Edward Memorial Hospital (KEMH), and (iii) an intention to remain in Western Australia to enable future follow-up of their child. Ninety percent of eligible women agreed to participate in the study. From the original cohort, 2868 children have been followed over two decades. The study was conducted with appropriate institutional ethics approval, and written informed consent was obtained from mothers at all follow-ups and participants at the year 17 follow-up. The RAINE sample is representative of the larger Australian population (88% Caucasian). DNA samples have been collected using standardized procedures at 14 or 16 years of age. Only those children with both biological parents of White European origin were included in the current analyses.

www.rainestudy.org.au, 13. McKnight CM, Newnham JP, Stanley FJ, et al. Birth of a cohort — The first 20 years of the raine study. Med J Aust. 2012;197(11):608-610.

TCHAD TCHAD[14] is a longitudinal study of all 1480 twin pairs born in Sweden between May 1985 and December 1986 followed with four waves of measurements from childhood (age 8–9), throughout early (age 13–14) and late adolescence (age 16–17), into emerging adulthood (age 19–20). The TCHAD (Dnr 94-277, 98-486, 02-271, 05-628, 12-2107) study has been approved by the Ethics Committee at Karolinska Institutet.

https://ki.se/en/research/swedish-twin-registry-for-researchers, 14. Lichtenstein P, Tuvblad C, Larsson H, Carlström E. The Swedish twin study of child and adolescent development: the TCHAD-study. Twin Res Hum Genet. 2007;10(1):67-73.

TEDS TEDS[15] is a multivariate longitudinal study that recruited more than 11,000 twin pairs born in England and Wales in 1994, 1995, and 1996. The TEDS sample is representative of the UK population compared with the data obtained by the Office of National Statistics. The project received approval from the King's College London Institute of Psychiatry ethics committee, and parental consent was obtained before data collection.

https://www.teds.ac.uk/, 15. Haworth CMA, Davis OSP, Plomin R. Twins early development study (TEDS): A genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. Twin Res Hum Genet. 2013;16(1):117-125.

TRAILS TRAILS[16] (TRacking Adolescents' Individual Lives Survey) is a prospective cohort study of Dutch adolescents with bi- or triennial measurements from age 11 to at least age 25 and consists of a general population and a clinical cohort (for a cohort profile see Huisman et al., 2008). In the population cohort, four assessment waves have been completed to date, which ran from March 2001 to July 2002 (T1), September 2003 to December 2004 (T2), September 2005 to August 2007 (T3), and October 2008 to September 2010 (T4). Data for the present study were collected during the third assessment wave. TRAILS is conducted in accordance with the World Medical Association Declaration of Helsinki and the protocol was approved by the Central Committee on Research Involving Human Subjects (CCMO), The Hague, the Nether-

lands. All participating adolescents and their parents gave written informed consent.

www.trails.nl, 16. Huisman M, Oldehinkel AJ, de Winter A, et al. Cohort profile: The Dutch TRacking Adolescents Individual Lives Survey; TRAILS. Int J Epidemiol. 2008;37(6):1227-1235.

YFS The Young Finns Study[17,18] has been financially supported by the Academy of Finland: grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research ; Finnish Cultural Foundation; The Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; and EU Horizon 2020 (grant 755320 for TAXINOMISIS); European Research Council (grant 742927 for MULTIEPIGEN project); and The Wihuri Foundation.

http://youngfinnsstudy.utu.fi/, 17. Juonala M, Viikari JSA, Raitakari OT. Main findings from the prospective Cardiovascular Risk in Young Finns Study. Curr Opin Lipidol. 2013;24(1):57-64., 18. Raitakari OT, Juonala M, Rönnemaa T, et al. Cohort profile: The cardiovascular risk In young Finns study. Int J Epidemiol. 2008;37(6):1220-1226.

Table S 3: Tissue expression analysis (organs)

Tissue	β	β_{SD}	SE	p	q
Brain	0.01	0.04	0.01	0.02	0.57
Liver	0.01	0.02	0.01	0.11	0.86
Breast	0.02	0.05	0.02	0.14	0.86
Pituitary	0.01	0.02	0.01	0.15	0.86
Adipose tissue	0.01	0.04	0.01	0.15	0.86
Nerve	0.01	0.03	0.01	0.17	0.86
Kidney	0.01	0.02	0.01	0.23	0.94
Skin	0.01	0.01	0.01	0.03	0.94
Muscle	0.00	0.01	0.01	0.37	0.94
Vagina	0.00	0.01	0.01	0.38	0.94
Bladder	0.00	0.01	0.02	0.04	0.94
Colon	0.00	0.01	0.02	0.04	0.94
Stomach	0.00	0.00	0.01	0.45	0.94
Testis	0.00	0.00	0.01	0.51	0.94
Blood	0.00	0.00	0.01	0.53	0.94
Pancreas	0.00	0.00	0.01	0.54	0.94
Salivary Gland	0.00	0.00	0.01	0.55	0.94
Cervix Uteri	0.00	-0.01	0.02	0.58	0.94
Small intestine	0.00	-0.01	0.01	0.61	0.94
Blood vessel	0.00	-0.01	0.01	0.63	0.94
Prostate	-0.01	-0.03	0.01	0.75	0.94
Heart	-0.01	-0.02	0.01	0.76	0.94
Uterus	-0.01	-0.03	0.01	0.77	0.94
Esophagus	-0.01	-0.04	0.02	0.08	0.94
Spleen	-0.01	-0.02	0.01	0.85	0.94

Table S 3: (Continued)

Adrenal gland	-0.01	-0.03	0.01	0.87	0.94
Fallopian tube	-0.02	-0.05	0.02	0.88	0.94
Thyroid	-0.02	-0.04	0.01	0.09	0.94
Lung	-0.02	-0.04	0.01	0.91	0.94
Ovary	-0.02	-0.05	0.01	0.96	0.96

β Beta
βSD Beta Standard Deviation
SE Standard Error
p P-value
q False Discovery Rate Adjusted P-values

Table S 4: Generalizability of the results

Random Split	r_G	SE	p	Sample 1 h^2	Sample 2 h^2
1	0.66	0.21	2,00E-3	0.16	0.07
2	0.65	0.21	2,00E-3	0.08	0.13
3	0.84	0.22	1,00E-4	0.10	0.10
4	0.73	0.23	1,00E-3	0.12	0.09
5	0.80	0.21	1,00E-3	0.10	0.11
6	0.71	0.20	5,00E-4	0.11	0.10
7	0.63	0.16	7,00E-5	0.13	0.11
8	0.93	0.20	5,00E-6	0.11	0.09
9	0.79	0.18	1,00E-5	0.13	0.08
10	0.98	0.23	3,00E-5	0.12	0.08
	r_G	SD		h^2	SD
Average	0.77	0.12		0.11	0.02

Cohorts were split into two halves, each half was meta-analyzed and genetic correlations between both halves were computed. This was repeated ten times.
PMID PubMed ID
rG Genetic Correlation
SE Standard Error
p P-value
h^2 SNP heritability
SD Standard Deviation of genetic correlations