

University of Groningen

## Genetics of early-life head circumference and genetic correlations with neurological, psychiatric and cognitive outcomes

Early Growth Genetics Consortium; Vogelesang, Suzanne; Bradfield, Jonathan P.; Grant, Struan F. A.; Felix, Janine F.; Jaddoe, Vincent W. V.

*Published in:*  
BMC Medical Genomics

*DOI:*  
[10.1186/s12920-022-01281-1](https://doi.org/10.1186/s12920-022-01281-1)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2022

[Link to publication in University of Groningen/UMCG research database](#)

### *Citation for published version (APA):*

Early Growth Genetics Consortium, Vogelesang, S., Bradfield, J. P., Grant, S. F. A., Felix, J. F., & Jaddoe, V. W. V. (2022). Genetics of early-life head circumference and genetic correlations with neurological, psychiatric and cognitive outcomes. *BMC Medical Genomics*, 15, [124]. <https://doi.org/10.1186/s12920-022-01281-1>

### **Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### **Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

RESEARCH

Open Access



# Genetics of early-life head circumference and genetic correlations with neurological, psychiatric and cognitive outcomes

Suzanne Vogelezang<sup>1,2,3†</sup>, Jonathan P. Bradfield<sup>4,5†</sup>, the Early Growth Genetics Consortium, Struan F. A. Grant<sup>4,6,7,8†</sup>, Janine F. Felix<sup>1,2†</sup> and Vincent W. V. Jaddoe<sup>1,2\*†</sup>

## Abstract

**Background:** Head circumference is associated with intelligence and tracks from childhood into adulthood.

**Methods:** We performed a genome-wide association study meta-analysis and follow-up of head circumference in a total of 29,192 participants between 6 and 30 months of age.

**Results:** Seven loci reached genome-wide significance in the combined discovery and replication analysis of which three loci near *ARFGF2*, *MYCL1*, and *TOP1*, were novel. We observed positive genetic correlations for early-life head circumference with adult intracranial volume, years of schooling, childhood and adult intelligence, but not with adult psychiatric, neurological, or personality-related phenotypes.

**Conclusions:** The results of this study indicate that the biological processes underlying early-life head circumference overlap largely with those of adult head circumference. The associations of early-life head circumference with cognitive outcomes across the life course are partly explained by genetics.

**Keywords:** Head circumference, Genome-wide association study, Genetic correlations, Infancy

## Background

Head circumference is a complex trait, commonly used as an indicator of brain volume during development and associated with child and adult intelligence [1–3]. It is also used as a measure of skeletal growth in fetal life, at birth and in early childhood [4, 5]. Twin studies show heritability estimates ranging from 75 to 90%, which are consistent across the life course [6]. Large genome-wide association studies (GWAS) have identified multiple

loci associated with child and adult head circumference, intracranial volume and brain volume [7–10]. Heritability estimates from GWAS range from 10 to 31% [8]. However, only two genetic loci associated with head circumference between 6 and 30 months have been identified so far [11]. Identifying additional genetic loci related to early-life head circumference may contribute towards our understanding of early brain development. This is important since observational studies have associated early brain development with several neurological and psychiatric diseases, such as Alzheimer's disease, schizophrenia and autism [12–17]. The underlying mechanisms are poorly understood. Both genetics and environmental factors play a role [18]. Additionally, the shared genetic contribution between early-life head circumference and later-life outcomes is yet unknown. Unravelling this shared genetic contribution may help us to better

<sup>†</sup>Suzanne Vogelezang, Jonathan P. Bradfield, Struan F. A. Grant, Janine F. Felix and Vincent W. V. Jaddoe have contributed equally to this work

A list of authors and their affiliations appears at the end of the paper

\*Correspondence: v.jaddoe@erasmusmc.nl

<sup>1</sup>The Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

Full list of author information is available at the end of the article



understand the etiology of later-life outcomes related to early-life head circumference.

We examined the genetic background of early-life head circumference by performing a two-stage GWAS meta-analysis comprising 25 studies with a combined sample size of 29,192 European-ancestry participants between 6 and 30 months of age. We also examined genetic correlations of early-life head circumference with anthropometrics, brain volume-related, neurological, psychiatric, cognitive, and personality related traits.

## Methods

### Study design

We conducted a two-stage meta-analysis in children of European ancestry to identify genetic loci associated with birth and early-life head circumference. Sex- and age-adjusted standard deviation scores (SDS) were created for head circumference between 6 to 30 months (closest to 18 months, if multiple measurements were available) using Growth Analyzer 3.0 across all studies [19]. In the case of twin pairs and siblings, only one of each twin or sibling pair was included, either randomly or based on genotyping or imputation quality.

In the discovery stage, we performed a meta-analysis of early-life head circumference in 21 studies ( $N = 22,279$ ), including the Amsterdam Born Children and their Development-Genetic Enrichment Study (ABCD,  $N = 1018$ ), the Avon Longitudinal Study of Parents and Children (ALSPAC,  $N = 3960$ ), Children's Hospital of Philadelphia (CHOP,  $N = 856$ ), the Copenhagen Prospective Studies on Asthma in Childhood 2000 (COPSAC2000,  $N = 325$ ) and 2010 (COPSAC2010,  $N = 603$ ), the Danish National Birth Cohort- preterm birth study (DNBC-PTB,  $N = 508$ ), the Generation R Study (GenerationR,  $N = 2299$ ), the Danish National Birth Cohort—the Genetics of Overweight Young Adults offspring study (DNBC GOYA-offspring,  $N = 230$ ), the INfancia y Medio Ambiente [Environment and Childhood] Project, with two subcohorts that were entered into the meta-analysis separately (INMA-Sabadell and Valencia subcohort and INMA Menorca subcohort,  $N = 550$ ), the German Infant Study on the influence of Nutrition Intervention PLUS environmental and genetic influences on allergy development & Influence of life-style factors on the development of the immune system and allergies in East and West Germany (GINIplus&LISA,  $N = 1455$ ), the Leipzig Research Center for Civilization Diseases—Child study (LIFE-Child,  $N = 1365$ ), the Norwegian Mother Child Cohort (MoBa,  $N = 836$ ), the Northern Finland Birth Cohort 1966 (NFBC 1966,  $N = 4603$ ), the Northern Finland Birth Cohort 1986 (NFBC 1986,  $N = 826$ ), the Physical Activity and Nutrition in Children Study (PANIC,  $N = 372$ ), the Raine Study (Raine Study,  $N = 1373$ ), the

Småbørns Kost Og Trivsel study, including two subcohorts (SKOT 1,  $N = 170$  and SKOT 2,  $N = 98$ ), the Special Turku Coronary Risk factor Intervention Project (STRIP,  $N = 505$ ), and the TEENs of Attica: Genes and Environment (TEENAGE,  $N = 327$ ).

In the replication stage of early-life head circumference analysis, we included 4 studies ( $N = 6913$ ): 319 additional children from the INfancia y Medio Ambiente [Environment and Childhood] (INMA) Project (INMA-Gipuzkoa subcohort), 5644 additional children from the Norwegian Mother and Child Cohort (MoBa), the European Childhood Obesity Project (CHOP Study,  $N = 366$ ), and the Exeter Family Study of Childhood Health (EFSOCH,  $N = 584$ ). Characteristics of discovery and replication studies can be found in Additional file 2: Table S1. The study design of birth head circumference can be found in the Additional file 1.

### Study-level analyses

Genome-wide association analyses were first run in all discovery cohorts for birth and early-life head circumference separately. Studies used high-density Illumina or Affymetrix Single Nucleotide Polymorphism (SNP) arrays, followed by imputation to the 1000 Genomes Project or Haplotype Reference Consortium (HRC). Before imputation, studies applied study specific quality filters on sample and SNP call rate, minor allele frequency and Hardy–Weinberg disequilibrium (see Additional file 2: Table S1 for details). Linear regression models assuming an additive genetic model were run in each study, to assess the association of each SNP with SDS head circumference, adjusting for principal components if this was deemed needed in the individual studies. As SDS head circumference is age and sex specific, no further adjustments were made. Before the meta-analysis, we applied quality filters to each study, filtering out SNPs with a minor allele frequency (MAF) below 1% and SNPs with poor imputation quality ( $MACH\ r2\_hat \leq 0.3$ ,  $IMPUTE\ proper\_info \leq 0.4$  or  $info \leq 0.4$ ).

### Meta-analysis

We performed fixed-effects inverse-variance weighted meta-analysis of all discovery samples using Metal [20]. Genomic control was applied to every study before the meta-analysis. Individual study lambdas before genomic control ranged from 0.99 to 1.03 (Additional file 2: Table S1). The lambda of the discovery meta-analysis was 1.02. Linkage Disequilibrium (LD) score regression analysis showed an intercept of 1.0, indicating that the slight inflation was mainly caused by polygenicity of early-life head circumference and not by population stratification, cryptic relatedness or other confounders [21, 22]. After the meta-analysis, we excluded SNPs for which

information was available in less than 50% of the studies or in less than 50% of the total sample size.

Genome-wide Complex Trait Analysis (GCTA) was used to select the independent SNPs for each locus [23]. We performed conditional analyses based on summary-level statistics and LD estimation between SNPs from the Generation R Study as a reference sample to select independently associated SNPs on the basis of conditional  $P$  values [23]. For early-life head circumference, 27 genome-wide significant or suggestive loci ( $P$  values  $< 5 \times 10^{-8}$  and  $P$  values  $< 5 \times 10^{-6}$ , respectively) were taken forward for replication in the 4 replication cohorts. Fixed-effects inverse variance weighted meta-analysis was performed for these 27 SNPs combining the discovery samples and all replication samples, giving a combined analysis beta, standard error and  $P$  value (Table 1). SNPs that reached genome-wide significance ( $P$  value  $< 5 \times 10^{-8}$ ) in the combined analysis were considered to be significantly associated with SDS-head circumference. For birth head circumference, SNPs were taken forward for replication, using the same methodology.

#### Functional mapping and annotation of genetic associations (FUMA)

To obtain predicted functional consequences for the SNPs that reached genome-wide significance in the combined meta-analysis, we used SNP2FUNC in FUMA, a web-based platform to facilitate and visualize functional annotation of GWAS results [24]. To annotate the nearest genes of the seven SNPs in biological context, we used the GENE2FUNC option in FUMA, which provides hypergeometric tests of enrichment of the list of nearest genes in 53 GTEx tissue-specific gene expression sets [24, 25]. We used GENE2FUNC for two sets of genes: 1. Nearest genes of seven SNPs; 2. Genes located in a region of 500 kb to either side of the 7 SNPs [24].

#### Colocalization analysis

We used Bayesian colocalization analysis to examine evidence for colocalization between early-life head circumference and eQTL signals (GTEx v7). Colocalization analyses were conducted using the R package *coloc*, <https://cran.r-project.org/web/packages/coloc>, as described previously [26]. Briefly, in each of the GTEx v7 tissues, all cis-eQTLs at FDR  $< 5\%$  were identified. For each eQTL, GWAS summary statistics were extracted for all SNPs that were present in  $> 50\%$  of the studies and  $> 50\%$  of the total sample size and that were in common to both GWAS and eQTL studies, within 1 MB of the transcription start site of the gene. For each such locus, colocalization analyses were done with default parameters, testing the following hypotheses [26]:

- $H_0$ : No association with either trait;
- $H_1$ : Association with early-life head circumference only;
- $H_2$ : Association with gene expression only;
- $H_3$ : Association with early-life head circumference and gene expression, two distinct causal variants;
- $H_4$ : Association with early-life head circumference and gene expression, one shared causal variant.

Support for each hypothesis was quantified in terms of posterior probabilities, defined at SNP level and indicated by  $PP_0$ ,  $PP_1$ ,  $PP_2$ ,  $PP_3$  or  $PP_4$ , corresponding to the five hypotheses and measuring how likely these hypotheses are. In most pairs, no evidence for association was found with either trait. In case association was observed, it was mostly with a single trait. To define colocalization we used restriction to pairs of early-life head circumference and eQTL signals with a high posterior probability for colocalization, indicated by a  $PP_4/(PP_3 + PP_4) > 0.9$ .

#### The Database for Annotation, Visualization and Integrated Discovery (DAVID)

To explore biological processes, we used DAVID, with the seven nearest genes as input, using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [27, 28].

#### Linkage-disequilibrium score regression

The use of LD score regression to estimate genetic correlations between two phenotypes has been described in detail previously [29]. Briefly, LD score is a measure of how much a genetic variation is tagged by each variant. A high LD score indicates that a variant is in high LD with many nearby polymorphisms. Variants with high LD scores are more likely to contain true signals and have a higher chance of overlap with genuine signals between GWAS. To estimate LD scores, summary statistics from GWAS meta-analysis are used to calculate the cross-product of test statistics of per SNP, which is regressed on the LD score. The slope of the regression is a function of the genetic covariance between traits [29]:

$$E(z_{1j}z_{2j}) = \frac{\sqrt{N_1N_2}\rho_g}{M}l_j + \frac{\rho N_s}{\sqrt{N_1N_2}}$$

where  $N_i$  is the sample size of study  $i$ ,  $\rho_g$  is the genetic covariance,  $M$  is the number of SNPs in the reference panel with a MAF between 5 and 50%,  $l_j$  is the LD score for SNP  $j$ ,  $N_s$  quantifies the number of individuals that overlap both studies, and  $\rho$  is the phenotypic correlation amongst the  $N_s$  of overlapping samples. A sample overlap or cryptic relatedness between samples only affects the intercept from the regression but not the slope. Thus, estimates are robust even in presence of sample overlap

**Table 1** Results of the 27 SNPs with  $P$  values  $< 5 \times 10^{-6}$  in the discovery phase

SNP	CHR	Position	Nearest gene	EA/non_EA	P value discovery	P value replication	EAF <sub>a</sub>	Beta (SE) combined	P value combined	Heterogeneity I <sup>2</sup> combined analysis
rs8756 <sub>b,c,d</sub>	12	66,359,752	HMG42	C/A	$5.18 \times 10^{-12}$	$4.62 \times 10^{-3}$	0.50	0.065 (0.009)	$1.84 \times 10^{-13}$	0
rs9795522 <sub>b</sub>	12	123,730,935	C12orf65	C/A	$3.20 \times 10^{-9}$	$3.86 \times 10^{-5}$	0.23	0.072 (0.010)	$7.37 \times 10^{-12}$	0
rs10883848 <sub>d</sub>	10	104,973,061	NT5C2	A/G	$3.06 \times 10^{-6}$	0.03	0.35	0.058 (0.010)	$9.99 \times 10^{-10}$	54.8
rs6095360	20	47,532,536	ARFGF2	G/A	$3.53 \times 10^{-8}$	$6.19 \times 10^{-4}$	0.32	0.055 (0.009)	$4.39 \times 10^{-9}$	36.6
rs3134614	1	40,363,054	MYCL1	C/G	$4.31 \times 10^{-6}$	$7.75 \times 10^{-4}$	0.88	0.075 (0.013)	$1.43 \times 10^{-8}$	0
rs6016511	20	39,709,255	TOP1	T/C	$4.55 \times 10^{-7}$	0.02	0.68	0.052 (0.010)	$4.02 \times 10^{-8}$	0
rs7792211 <sub>d</sub>	7	50,707,141	GRB10	G/A	$7.62 \times 10^{-7}$	0.01	0.42	0.051 (0.009)	$4.43 \times 10^{-8}$	11.3
rs1490384 <sub>c,d</sub>	6	126,851,160	CENPW	T/C	$5.46 \times 10^{-8}$	0.14	0.51	0.049 (0.009)	$5.70 \times 10^{-8*}$	43.9
rs116536930	17	43,855,156	CRHR1	G/T	$4.19 \times 10^{-8}$	0.53	0.51	0.055 (0.011)	$2.27 \times 10^{-7*}$	40.1
rs72952297	2	180,379,814	ZNF385B	T/G	$2.07 \times 10^{-6}$	0.11	0.93	0.087 (0.018)	$9.19 \times 10^{-7*}$	0
rs12350281	9	20,637,773	MLL3	G/A	$4.22 \times 10^{-6}$	0.03	0.05	0.152 (0.031)	$9.98 \times 10^{-7*}$	43.1
rs7658192	4	140,023,559	ELF2	T/G	$2.29 \times 10^{-7}$	0.46	0.33	0.046 (0.010)	$1.70 \times 10^{-6*}$	62.9
rs143094271 <sub>e</sub>	17	7,463,102	TNFSF12-TNFSF13	A/G	$5.96 \times 10^{-7}$	0.44	0.02	0.133 (0.028)	$2.02 \times 10^{-6*}$	33.7
rs1014098	1	119,458,215	TBX15	C/T	$2.67 \times 10^{-6}$	0.17	0.79	0.051 (0.011)	$2.05 \times 10^{-6*}$	30.3
rs77690628	8	26,189,223	PPP2R2A	A/T	$1.02 \times 10^{-6}$	0.31	0.07	0.080 (0.017)	$3.26 \times 10^{-6*}$	24.6
rs9853018	3	141,101,961	ZBTB38	T/C	$5.41 \times 10^{-7}$	0.62	0.45	0.040 (0.009)	$5.30 \times 10^{-6}$	38.4
rs4674101	2	217,487,622	RPL37A	A/G	$8.66 \times 10^{-7}$	0.57	0.31	0.043 (0.010)	$6.56 \times 10^{-6}$	65.3
rs150266910	1	23,442,265	LUZP1	C/T	$1.42 \times 10^{-6}$	0.65	0.83	0.052 (0.012)	$1.04 \times 10^{-5}$	58.2
rs13067734	3	141,963,015	GK5	A/C	$2.83 \times 10^{-6}$	0.55	0.52	0.037 (0.009)	$1.22 \times 10^{-5}$	26.5
rs17223831	14	90,601,920	KCNK13	A/G	$1.24 \times 10^{-6}$	0.18	0.14	0.067 (0.015)	$1.30 \times 10^{-5}$	67.6
rs62580922	9	125,631,594	RC3H2	C/T	$1.13 \times 10^{-8}$	0.35	0.07	0.072 (0.017)	$1.52 \times 10^{-5}$	77.8
rs2448415	1	93,469,310	FAM69A	A/G	$1.97 \times 10^{-6}$	0.81	0.47	0.037 (0.009)	$2.56 \times 10^{-5}$	60.7
rs261752	5	39,328,506	C9	A/G	$1.39 \times 10^{-6}$	0.70	0.55	0.035 (0.009)	$7.63 \times 10^{-5}$	63.7
rs72979897	6	140,881,181	MIR3668	T/A	$4.13 \times 10^{-6}$	0.88	0.04	0.100 (0.025)	$8.26 \times 10^{-5}$	50.6
rs503783	6	161,959,649	PARK2	T/C	$2.98 \times 10^{-6}$	0.66	0.27	0.037 (0.010)	$1.59 \times 10^{-4}$	74.3
rs1113397574	5	31,968,320	PDZD2	T/C	$4.56 \times 10^{-6}$	0.61	0.02	0.118 (0.032)	$2.40 \times 10^{-4}$	58.5
rs77535478	9	77,533,287	TRPM6	A/G	$1.28 \times 10^{-6}$	0.39	0.04	0.081 (0.026)	$2.05 \times 10^{-3}$	75.4

**Bolded P values indicate genome-wide significance in the combined analysis**

**CHR** chromosome, **EA** effect allele, **EAF** effect allele frequency, **SE** standard error

\*Genome-wide suggestive loci

<sup>a</sup> From combined analysis

<sup>b</sup> Locus previously reported for infant head circumference [11]

<sup>c</sup> Locus previously reported for adult intracranial volume [7, 9]

<sup>d</sup> Locus previously reported for adult head brain volume [8]

<sup>e</sup> Locus previously reported for childhood and adult head circumference and intracranial volume [10]



when comparing traits across distinct GWAS populations. Estimates of genetic covariance are therefore not biased by overlapping samples. Similarly, in case of population stratification, the intercept is affected but it has only minimal impact on the slope since population stratification does not correlate with LD between variants.

Because of the correlation between the imputation quality and LD score, imputation quality is a confounder for LD score regression. Therefore, SNPs were excluded according to the following criteria:  $MAF < 0.01$  or  $INFO \leq 0.9$ . The filtered GWAS results were uploaded on the online webtool, a web service with many GWAS meta-analyses available on which LD score regression has been implemented by the developers of the LD score regression method. In case multiple GWAS meta-analyses were available for the same phenotype, the genetic correlation with early-life head circumference was estimated using the most recent meta-analysis. Genetic correlations are shown in Fig. 3 and Additional file 1: Table S7.

#### Genetic risk score and percentage of variance explained

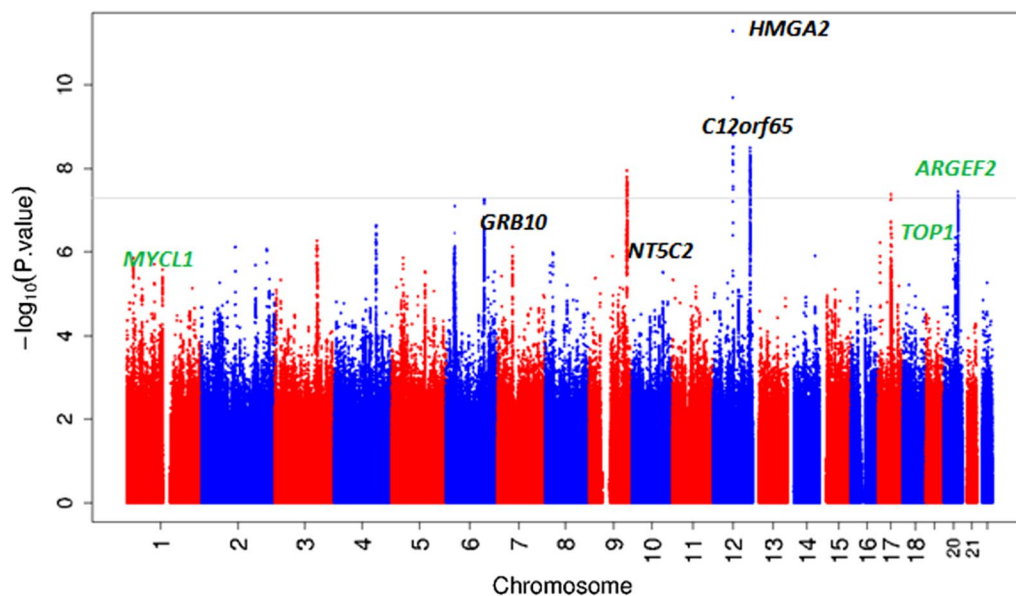
We combined the seven genome-wide significant SNPs from the combined meta-analysis into a Genetic Risk Score (GRS) by summing up the number of alleles that increase the SDS head circumference, weighted by the effect sizes from the combined meta-analysis. The GRS was rescaled to a range from 0 to 14, which is the maximum number of head circumference SDS increasing

alleles and rounded to the nearest integer. Linear regression analysis was used to examine the associations of the risk score with head circumference and intracranial volume at different ages. For these analyses data from the Generation R Study and UK Biobank were used. When calculating the risk score for the Generation R study, effect estimates from the combined meta-analysis were used after excluding Generation R from the meta-analysis. The variance explained was estimated by the adjusted  $R^2$  of the models.

## Results

### Identification of genetic loci associated with early-life head circumference

Individual study characteristics are shown in Additional file 2: Table S1. In the discovery stage, we performed a fixed-effects inverse variance-weighted meta-analysis including data imputed to the 1000 Genomes or the Haplotype Reference Consortium (HRC) reference panels from 21 studies ( $N = 22,279$ ). Using data from the discovery cohorts, single nucleotide polymorphisms (SNPs) at five independent loci reached genome-wide significance ( $P$  values  $< 5 \times 10^{-8}$ ) and SNPs at another 22 loci showed suggestive associations with early-life head circumference ( $5 \times 10^{-8} < P$  values  $< 5 \times 10^{-6}$ ). A Manhattan plot of the discovery meta-analysis is shown in Fig. 1. No evidence of inflation by population stratification or cryptic relatedness was found (genomic inflation factor ( $\lambda$ ) = 1.02 and LD-score regression intercept = 1.0) (Additional



**Fig. 1** Manhattan plot of results of the discovery meta-analysis of 21 GWAS. On the x-axis the chromosomes are shown. On the y-axis the  $-\log_{10}$  of the  $P$  value is shown. Novel SNPs are shown in green. Known SNPs are shown in black. The genome wide significance cutoff of  $5 \times 10^{-8}$  is represented by the grey dotted line

file 1: Fig. S1) [21]. The index SNPs from each of the 27 genome-wide and suggestive loci were followed up in four replication cohorts ( $N=6913$ ). The results of the discovery, replication and combined analyses are shown in Table 1 and Additional file 1: Tables S2 and S3. Results of the discovery analysis for SNPs with  $P$  values  $< 5 \times 10^{-6}$  are shown in Additional file 3: Table S4.

Of the 27 SNPs identified in the discovery meta-analysis, seven reached genome-wide significance in the combined meta-analysis, in which we used data from the discovery and replication stage. An identified locus was defined to be a known locus if the index SNP was within a range of 500 kb upstream to 500 kb downstream of and in LD ( $r^2 \geq 0.2$ ) with a previously reported SNP for head circumference, intracranial volume, or brain volume at any age [7–11]. Of the seven genome-wide significant SNPs, three were novel: rs6095360 near *ARFGEF2*, rs3134614 near *MYCL1*, and rs6016511 near *TOPI* (Table 1 and Additional file 1: Tables S2 and S3). Regional plots of these three loci are shown in Fig. 2. The remaining four SNPs mapped to loci previously identified from GWAS on infant head circumference, adult intracranial volume, and/or adult brain volume (nearest genes: *HMG2A*, *C12orf65*, *NT5C2*, and *GRB10*) [7, 10, 11].

Six SNPs located within 500 kb (upstream or downstream) from rs6095360 (*ARFGEF2*), rs3134614 (*MYCL1*), and rs6016511 (*TOPI*) have been previously reported in relation to adult height [30]. The linkage disequilibrium (LD) of the three novel SNPs near *ARFGEF2*, *MYCL1*, and *TOPI* with these six adult height SNPs was weak to moderate. We found suggestive evidence of association for rs6095360 (*ARFGEF2*) with early-life length in 28,949 participants between 6 and 30 months of age in an unpublished GWAS meta-analysis of 24 cohorts ( $P$  value  $4.58 \times 10^{-7}$ ), but the other two novel SNPs did not show evidence of association (Additional file 1: Table S5).

We also performed a meta-analysis of birth head circumference in a total of 32,084 participants. None of the SNPs reached genome-wide significance in this analysis. A total of 11 SNPs with  $P$  values between  $5 \times 10^{-8}$  and  $5 \times 10^{-6}$  were taken forward for replication ( $N=3750$ ) and combined analyses, but none were genome-wide significant in the combined analysis. Therefore, no follow-up analyses were performed for birth head circumference. A Manhattan plot and a Quantile–Quantile plot of the discovery meta-analysis of birth head circumference are

shown in Additional file 1: Figs. S2 and S3. Results of the discovery analysis for SNPs with  $P$  values  $< 5 \times 10^{-6}$  are shown in Additional file 4: Table S6.

### Functional characterization

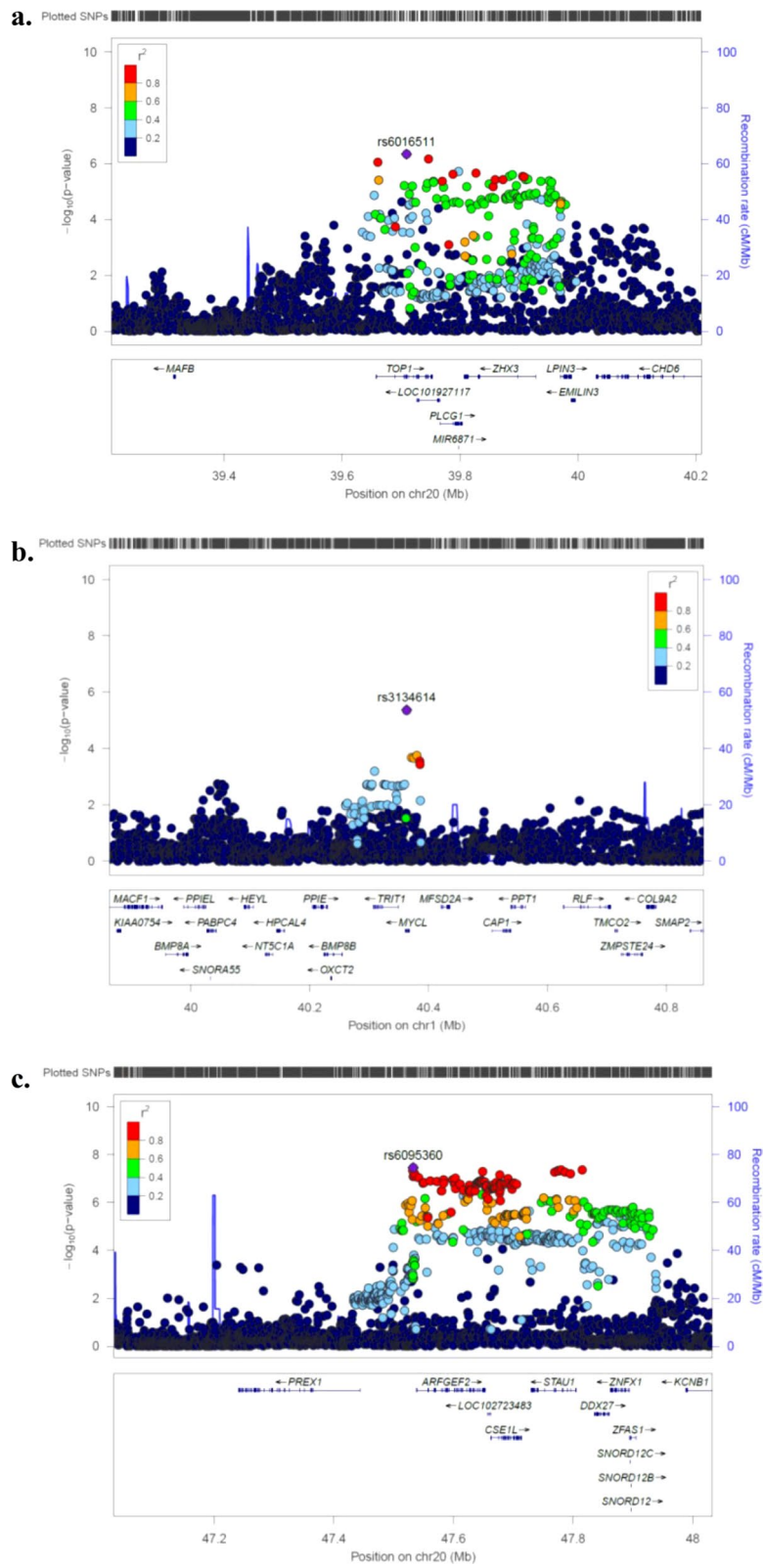
To gain insight into the function of the seven SNPs associated with early-life head circumference, we used several strategies. First, using Bayesian colocalization analysis, we examined evidence of colocalization between GWAS and eQTL signals for the seven index SNPs (GTEx v7), but did not find a signal at any of the seven loci. Second, to explore biological processes, we used the Kyoto Encyclopedia of Genes and Genomes (KEGG) database in the Database for Annotation, Visualization and Integrated Discovery (DAVID) with the seven SNPs and their nearest genes as input [27, 28], but no enriched biological processes were identified. Third, we did a look-up of the seven nearest genes in mouse-knockout data but there was no phenotypic information available for any of these gene knockouts [31]. Fourth, we examined gene expression profiles for the nearest genes to the seven SNPs with GTEx v7 in 53 tissues, using the tool for Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) [24, 25]. We did not find significant differential expression for these seven nearest genes. Going a step further, we included all genes within a range of 500 kb upstream to 500 kb downstream of the seven index SNPs and found significant differential expression in several brain structures, including the putamen, amygdala, hippocampus, caudate, nucleus accumbens, substantia nigra, and anterior cingulate cortex, and in other tissues such as the heart, pancreas and liver [25].

### Shared genetic background of early-life head circumference with childhood and adult outcomes

First, to examine whether associations between early-life head circumference and phenotypes in later life from observational studies were at least partly genetically explained, we estimated genetic correlations between early-life head circumference and other traits, using LD score regression, which is based on the genome-wide meta-analysis results [29]. We found positive genetic correlation coefficients for early-life head circumference with birth length ( $R_g=0.57$ ,  $P$  value  $= 3.61 \times 10^{-12}$ ), birth weight ( $R_g=0.48$ ,  $P$  value  $= 2.51 \times 10^{-8}$ ), adult height ( $R_g=0.34$ ,  $P$  value  $= 9.77 \times 10^{-11}$ ), adult body mass index

(See figure on next page.)

**Fig. 2 a–c** Locus zoom plots of the 3 novel loci. Shown are the results of the meta-analysis. Regional association plot of the 3 novel loci. SNPs are plotted with their  $P$  values from the discovery stage (as  $-\log_{10}$ ; left y-axis) as a function of genomic position (x-axis). Estimated recombination rates (right y-axis) taken from 1000 Genomes, March 2012 release are plotted to reflect the local LD-structure around the top associated SNP (indicated with purple color) and the correlated proxies (indicated in colors)



**Fig. 2** (See legend on previous page.)

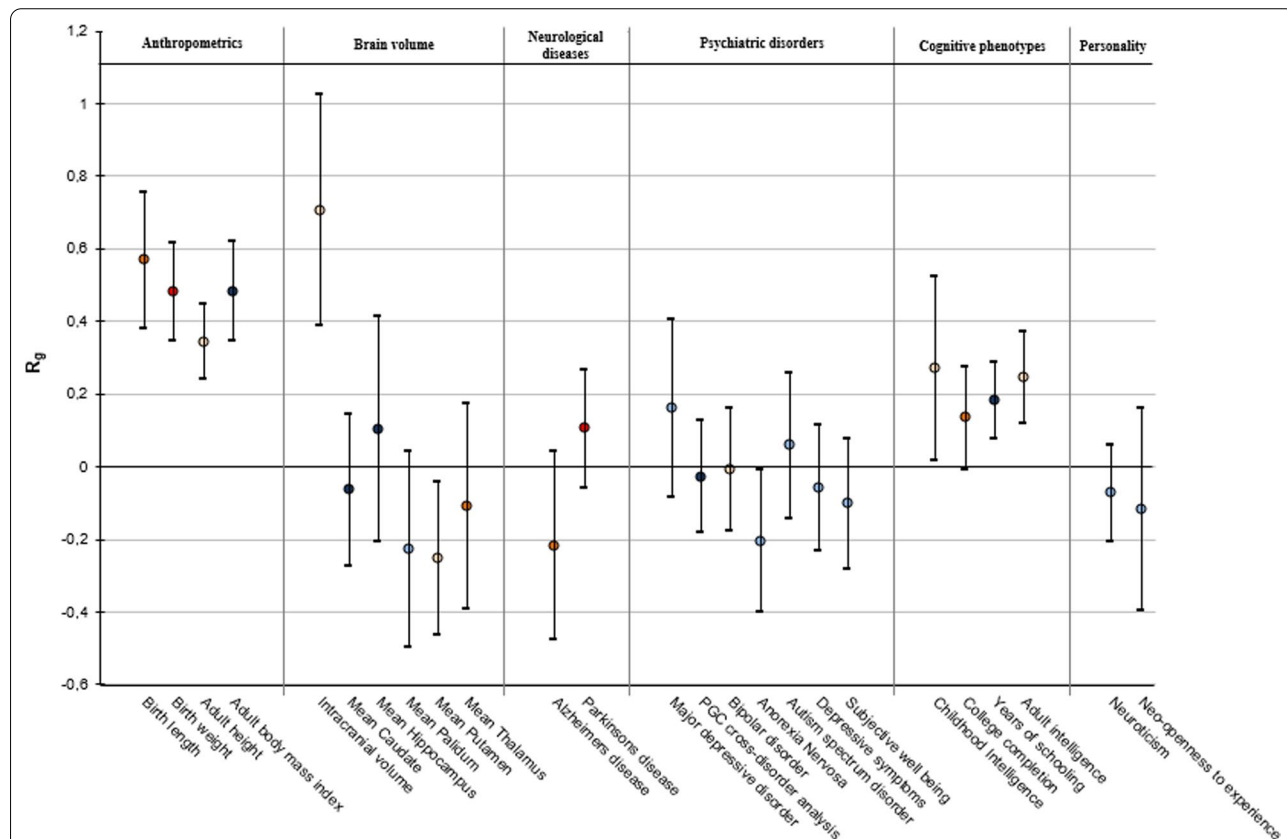


(BMI) ( $R_g=0.13$ ,  $P$  value=0.03), adult intracranial volume ( $R_g=0.71$ ,  $P$  value= $1.44 \times 10^{-5}$ ) and several cognitive outcomes, including childhood intelligence (age range 6–18 years) ( $R_g=0.27$ ,  $P$  value=0.04), years of schooling ( $R_g=0.18$ ,  $P$  value= $6.0 \times 10^{-4}$ ), and adult intelligence ( $R_g=0.25$ ,  $P$  value= $1.0 \times 10^{-4}$ ), (Fig. 3 and Additional file 1: Table S7). We did not find genetic correlations with any psychiatric (for instance bipolar disorder and autism spectrum disorder), neurological (Parkinson's disease, Alzheimer's disease) or personality-related outcomes (for instance neuroticism) ( $P$  values > 0.05) (Additional file 1: Table S7).

Second, we performed a look-up of the seven identified SNPs in GWAS meta-analyses data on potentially related phenotypes, including adult intracranial volume, intelligence, Alzheimer's disease, neuroticism, depression and educational attainment [7, 32–35]. Rs10883848 (*NT5C2*) was genome-wide significantly associated with adult intracranial volume ( $P$  value= $1.48 \times 10^{-9}$ ). All effect estimates for intracranial volume were in the same direction as those for head circumference in early-life

(Additional file 1: Table S8). Additionally, rs6095360 (*ARFGF2*) was associated with adult intelligence at a genome-wide level ( $P$  value= $2.31 \times 10^{-16}$ ). We did not observe evidence of genome-wide significant associations with Alzheimer's disease, neuroticism, depression or educational attainment. Suggestive evidence for association was observed for rs8756 (*HMGA2*) with adult intracranial volume ( $P$  value= $6.32 \times 10^{-8}$ ), for rs9795522 (*C12orf65*) with adult intelligence ( $P$  value= $1.41 \times 10^{-6}$ ) and for rs8756 (*HMGA2*) with educational attainment ( $P$  value= $1.41 \times 10^{-6}$ ) (Additional file 1: Tables S9–13). None of the SNPs showed evidence of association with birth head circumference in the GWAS meta-analysis (Additional file 1: Table S14).

Third, we calculated a combined genetic risk score (GRS) using the seven index SNPs identified in the current study. We summed the number of head circumference-increasing alleles weighted by the effect sizes from the combined meta-analysis after excluding the Generation R Study, in which we tested the GRS longitudinally. The GRS was associated with fetal head circumference



**Fig. 3** Genome-wide genetic correlations between early-life head circumference and adult traits and diseases. On the x axis the traits and diseases are shown. The y-axis shows the genetic correlations ( $R_g$ ) and corresponding standard errors, indicated by error bars, between early-life head circumference and each trait, estimated by LD score regression. The genetic correlation estimates ( $R_g$ ) are colored according to their intensity and direction. Red indicates a positive correlation, blue indicates a negative correlation. References can be found in Additional file 1: Table S7

in the third trimester of pregnancy (N=1984), at post-natal ages 1 month (N=1501), 6 months (N=1662), 11 months (N=1528), and 6 years (N=4010) and at the mean (SD) age of 64 (7.5) years (N=22,152) in UK Biobank data (*P* values < 0.05) (Fig. 4 and Additional file 1: Table S15).

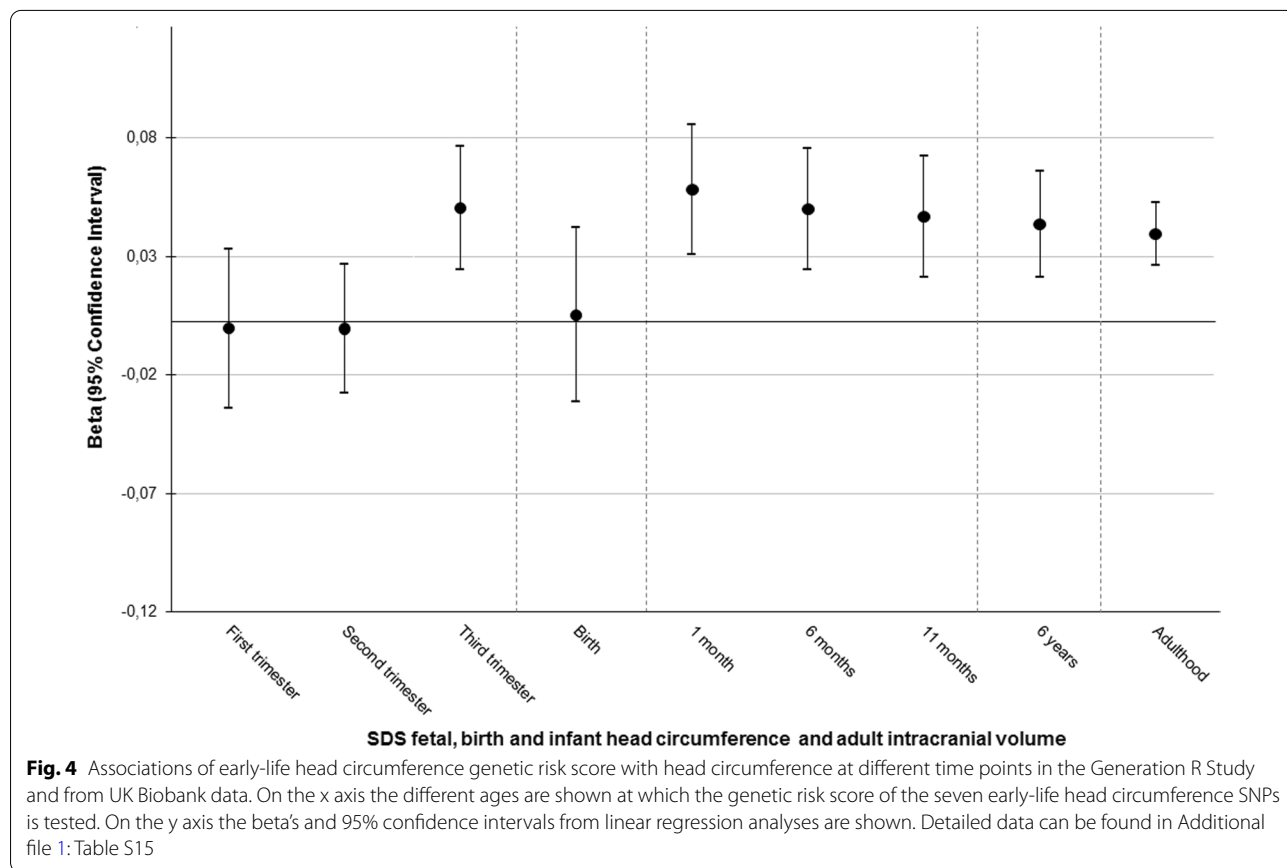
**Discussion**

In a GWAS meta-analysis including 29,192 participants of European ancestry aged 6 to 30 months of age, we identified seven genome-wide significant SNPs associated with early-life head circumference, of which three were novel and had not been related with head circumference, intracranial volume or brain volume before. We observed positive genetic correlations between early-life head circumference and adult intracranial volume as well as cognitive outcomes.

We used multiple approaches to identify potential underlying mechanisms. As there is no strong evidence linking the nearest genes to the seven SNPs as causal genes, we included all genes within 500 kb to either side of the genome-wide significant SNPs in a GTEx analysis. We found differential expression of these genes in different brain structures that are related to cognitive

functions and emotional control, indicating a potential functional role of these genes in the brain [36–39]. However, as donors aged 20–79 years were included in the GTEx data source, we were not able to look at expression of the genes in brain structures in early life. Using colocalization analysis, no potentially causal genes were identified [26]. Future studies should also determine whether the nearest genes, identified in this study, are indeed the causal genes and assess their expression in child brain structures.

The potential roles of the nearest genes to the novel loci are still poorly understood. *MYCL1*, (*MYCL* proto-oncogene, BHLH transcription factor), and *TOP1* (DNA topoisomerase 1) have been suggested to play a role in various types of cancer [40–44]. The role of these genes in the development of head circumference in early life is currently unknown. A mutation in *ARFGEF2* has been previously associated with several phenotypes related to brain development, including microcephaly [45, 46]. The three novel SNPs are located near regions that have been previously reported for adult height, indicating that they might represent loci involved in growth [30]. However, the strong association of rs6095360 (nearest gene: *ARFGEF2*) with adult intelligence (*P* value =  $2.31 \times 10^{-16}$ )



might indicate a role in brain development as well [32]. Future functional studies should investigate the role of these genes and should determine whether these genes are indeed causal.

Observational studies suggest that early life head circumference is not only related to intracranial volume in adults, but also to adult intelligence, Alzheimer's disease, schizophrenia and autism [1–3, 12–17]. In observational studies of such associations, effect estimates may be influenced by confounding factors and reverse causation, potentially evoking spurious associations [47, 48]. Genetic studies such as ours can provide more insight into the etiology of complex diseases. We found a strong genetic correlation of head circumference in early life with intracranial volume in adults, underlining the idea that early-life head circumference is a valid measure for brain growth during early development [3, 7]. Abnormal growth trajectories of head circumference are related to adverse neurological outcomes [49]. Additionally, variation within the normal range of head circumference has been reported to be associated with cognitive and behavioral traits [2, 50, 51]. In the current study, we observed positive genetic correlations for head circumference in early life with childhood intelligence, years of schooling and adult intelligence. These findings indicate that the association of early-life head circumference with cognitive function from observational studies is at least partly explained by a shared genetic background, which is in line with the observed positive genetic correlations between intracranial volume and cognitive function in the literature [7, 8, 52]. Altogether, the findings from the current study and from previous literature suggest that the associations between measures of early-life brain volume and cognition decades later are partly genetically explained.

It has been suggested that heritability estimates are consistent from childhood onwards [6]. Whether this genetic stability starts from early life onwards, is currently not well studied. We observed evidence for association of two of the seven SNPs with adult intracranial volume [7]. We combined the seven index SNPs into a weighted GRS. Although we have used effect sizes of the meta-analysis after excluding the Generation R Study, the discovery of the seven SNPs was based on the meta-analysis including the Generation R Study, potentially resulting in overfitting of the GRS. We found an association of the GRS with fetal head circumference in third trimester, head circumference in infancy and childhood and intracranial volume in adulthood. The effect estimates were largely similar for the different time points. We did not observe an association of the GRS with birth head circumference. This may be explained by the larger variance in birth head circumference that might be present

due to the deformation of the head during birth. Also, in the GWAS meta-analysis of head circumference at birth, no SNPs were genome-wide significantly associated with birth head circumference. Thus, the genetic background of early-life head circumference seems to partially overlap with the genetic background of related measures in later life. The SNPs identified in infancy seem to represent effects across multiple ages. However, as not all SNPs identified for early-life head circumference were associated with adult intracranial volume, it has been suggested that some of the underlying mechanisms are age-specific.

## Conclusions

We identified seven SNPs associated with early-life head circumference. Three of these are novel and four mapped to loci that are known for head circumference, intracranial volume, or brain volume. We observed a strong positive genetic correlation of early-life head circumference with adult intracranial volume and cognitive outcomes in childhood and adulthood. The well-known associations of head circumference with later cognitive phenotypes are partly explained by genetics. Our findings may contribute to the understanding of the early-life brain development, which may lay the foundation for diseases later in life.

## Abbreviations

GWAS: Genome-wide association studies; SDS: Standard deviation scores; MAF: Minor allele frequency; KEGG: Kyoto Encyclopedia of Genes and Genomes; HRC: Haplotype Reference Consortium; SNPs: Single nucleotide polymorphisms; LD: Linkage disequilibrium; DAVID: Database for Annotation, Visualization and Integrated Discovery; FUMA: Functional Mapping and Annotation of Genome-Wide Association Studies; BMI: Body mass index; GRS: Genetic risk score.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12920-022-01281-1>.

**Additional file 1.** Additional data includes 15 tables, 3 figures and additional cohort information and methods.

**Additional file 2.** Characteristics of discovery and replication studies.

**Additional file 3.** Results of the discovery analysis for SNPs with P-values  $< 5 \times 10^{-6}$ .

**Additional file 4.** Results of the discovery analysis for SNPs with P-values  $< 5 \times 10^{-6}$  for birth head circumference.

## Acknowledgements

SV, JFF and VVWJ designed the research and wrote the paper. SV, JFF and JPB analyzed the meta-analysis data. IN, JFF, VVWJ, GVD, KB, HB, KFM, LL, JCH, PR, VG, BK, JPL, EV, JE, DG, TAL, VL, MA, AK, WK, UG, CEP, SL, LW, OTR, KP, HN, TGMV, MM, BF, WK, and JH were responsible for the study design of an individual contributing study. JFF, VVWJ, IN, GVD, JTL, EW, TSA, KB, HB, MVL, KFM, LL, CFB, MAVL, JCH, NJT, GDS, PR, VG, BK, JPL, EV, JE, DE, TAL, MV, AK, WK, HH, SFAG, BAK, ATH, RMF, LSM, CI, MT, MV, CEP, SL, LW, HHHH, OTR, KP, HN, MM, BF, WK, JH,

and MFH were responsible for the management of an individual contributing study. SV, IN, GVD, JGE, RV, KB, HB, CSM, EAN, PRN, SJ, ØH, MV, CFB, MAVL, JCH, NJT, GDS, PR, VG, BK, JPL, EV, JE, DG, TAL, VL, MA, MV, AK, WK, HH, SFAG, BAK, ATH, LSM, MC, CEP, SL, LB, TEE, OTR, KP, HN, TGMV, FG, and BF were involved in sample collection and phenotyping of an individual contributing study. FR, EZ, JGE, TSA, HB, TMS, CTH, CSM, TH, NG, PRN, SJ, ØH, MV, CFB, MAVL, OP, JCH, ES, NJT, GDS, BP, MW, TAL, MA, MS, HH, SFAG, JMV, RNB, NFJ, JRB, CEP, SL, LW, OTR, MHZ, AGU, MMcC, and TGMV genotyped the data of an individual contributing study. SV, FR, JFF, IN, KP, NJ, JTL, TSA, TMS, CTH, NG, ØH, MV, BP, PR, VL, RJ, MS, JPB, RNB, IT, NFJ, JRB, CW, MJK, NP, MHZ, XL, FG, BF, VDJ, and SLRS analyzed data of an individual contributing study. All authors provided comments and consultation regarding the analyses, interpretation of the results and manuscript. SV, JFF and VVVJ had primary responsibility for final content. All authors gave final approval of the version to be published.

### The Early Growth Genetics Consortium

Suzanne Vogelezang<sup>1-3</sup>, Jonathan P Bradfield<sup>4,5</sup>, Stefan Johansson<sup>6,7</sup>, Evie Stergiakouli<sup>8-10</sup>, Elisabeth Thiering<sup>11,12</sup>, Craig E Pennell<sup>13</sup>, Tarunveer S Ahluwalia<sup>14-16</sup>, Ville Karhunen<sup>17,18</sup>, Markus Scholz<sup>19,20</sup>, Xueping Liu<sup>21</sup>, Carmen Iniguez<sup>22,23</sup>, Olli T Raitakari<sup>24-26</sup>, Jonas Bacelis<sup>27,28</sup>, Theresia M Schnurr<sup>29</sup>, Timo A Lakka<sup>30-32</sup>, Ioanna Ntalla<sup>33</sup>, Mads V Lind<sup>34</sup>, Lotte Lauritzen<sup>34</sup>, Sauli Herrala<sup>18</sup>, Frederick TJ Lin<sup>35</sup>, Christine Frithioff-Bjørnsøe<sup>29,36</sup>, Robin N Beaumont<sup>37</sup>, Mohammed H Zafarmand<sup>38,39</sup>, Peter Rzehak<sup>40</sup>, Jose R Bilbao<sup>41-43</sup>, Nella Junna<sup>44</sup>, Judith M Vonk<sup>45</sup>, Sheryl I Rifas-Shiman<sup>46</sup>, Vimala D Janjanam<sup>47</sup>, Maria J Knol<sup>3</sup>, Shweta Ramdas<sup>48</sup>, Lawrie Beilin<sup>49</sup>, Klaus Bønnelykke<sup>14</sup>, Maribel Casas<sup>23,50,51</sup>, Johan G Eriksson<sup>52-54</sup>, Joaquin Escribano<sup>55</sup>, Tavia E Evans<sup>35,56</sup>, Ulrike Gehring<sup>57</sup>, Frank Geller<sup>21</sup>, Veit Grote<sup>40</sup>, Darek Gruszfeld<sup>58</sup>, Hakon Hakonarson<sup>4,59</sup>, Andrew T Hattersley<sup>37,60</sup>, Christian T Have<sup>29</sup>, M Geoffrey Hayes<sup>35,61,62</sup>, Joachim Heinrich<sup>11,63</sup>, Øvind Helgeland<sup>64,64</sup>, Jhon Holloway<sup>65</sup>, Raimo Joro<sup>30</sup>, Julius Juodakis<sup>27</sup>, Bridget A Knight<sup>60</sup>, Bert Koletzko<sup>40</sup>, Antje Körner<sup>20,66</sup>, Jean-Paul Langhendries<sup>67</sup>, Jaakko T Leinonen<sup>44</sup>, Virpi Lindi<sup>68</sup>, Morten AV Lund<sup>36,69</sup>, Stephen Lye<sup>70</sup>, Mads Melbye<sup>21,71,72</sup>, Kim F Michaelsen<sup>34</sup>, Camilla S Morgen<sup>73,74</sup>, Harri Niinikoski<sup>26,75</sup>, Katja Pahlkala<sup>24,26,76</sup>, Kalliope Panoutsopoulou<sup>77,78</sup>, Oluf Pedersen<sup>29</sup>, Fernando Rivadeneira<sup>3,79</sup>, Loreto Santa-Marina<sup>80-82</sup>, Barbera DC Schaik<sup>39</sup>, Denise Scholtens<sup>83</sup>, Sylvain Sebert<sup>18,84</sup>, Ibon Tamayo<sup>85,86</sup>, Nicholas J Timpson<sup>8,9</sup>, Maties Torrent<sup>87,88</sup>, André G Uitterlinden<sup>3,79,89</sup>, Marc Vaudel<sup>6</sup>, Elvira Verduci<sup>90</sup>, Rebecca Vinding<sup>14</sup>, Mandy Vogel<sup>20,66</sup>, Eleftheria Zeggini<sup>77,91</sup>, Christopher Brown<sup>92</sup>, Hieab HH Adams<sup>3,56,93</sup>, Wilfried Karmaus<sup>47</sup>, Marie-France Hivert<sup>46,94,95</sup>, Gerard H Koppelman<sup>96</sup>, Elisabeth Widén<sup>44</sup>, Nora Fernandez-Jimenez<sup>41,42</sup>, Melanie Waldenberger<sup>97</sup>, Tanja GM Vrijkotte<sup>38</sup>, Rachel M Freathy<sup>27</sup>, Jens-Christian Holm<sup>29,36</sup>, William L Lowe, Jr.<sup>35</sup>, Niels Grarup<sup>29</sup>, Torben Hansen<sup>29</sup>, George V Dedoussis<sup>98</sup>, Mustafa Atalay<sup>30</sup>, Ellen A Nohr<sup>99,100</sup>, Bo Jacobsson<sup>101-103</sup>, Niina Pitkänen<sup>24,26</sup>, Martine Vrijheid<sup>23,50,51</sup>, Bjarke Feenstra<sup>21</sup>, Wieland Kiess<sup>20,66</sup>, Marjo-Riita Jarvelin<sup>17,18,84</sup>, Hans Bisgaard<sup>14</sup>, Carol Wang C<sup>13</sup>, Marie Standl<sup>11</sup>, Mark I McCarthy<sup>104-106</sup>, Beate St Pourcain<sup>8,107,108</sup>, Pål R Njølstad<sup>6,109</sup>, Struan FA Grant<sup>4,59,110-111</sup>, Janine F Felix<sup>1,2</sup>, Vincent WV Jaddoe<sup>1,2</sup>

These authors contributed equally: Suzanne Vogelezang and Jonathan P Bradfield.

These authors contributed equally: Struan FA Grant, Janine F Felix, Vincent WV Jaddoe.

1. The Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, The Netherlands. 2. Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam, The Netherlands. 3. Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, The Netherlands. 4. Center for Applied Genomics, Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, USA. 5. Quantinuum Research LLC, San Diego, CA, USA. 6. KG Jebsen Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, Norway. 7. Department of Medical Genetics, Haukeland University Hospital, Bergen, Norway. 8. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK. 9. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. 10. School of Oral and Dental Sciences, University of Bristol, Bristol, UK. 11. Institute of Epidemiology, Helmholtz Zentrum München- German Research Center for Environmental Health, Neuherberg, Germany. 12. Division of Metabolic and Nutritional Medicine, Dr. von Hauner Children's Hospital, University of Munich Medical Center, Munich, Germany. 13. University of Newcastle, School of Medicine and Public Health, Faculty of Medicine and Health. 14. COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark. 15. Steno Diabetes Center Copenhagen, Herlev, Denmark. 16. The Bioinformatics Center, Department of Biology, University of

Copenhagen, Copenhagen, Denmark. 17. Department of Epidemiology and Biostatistics, Imperial College London, London, UK. 18. Center for Life-Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland. 19. Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany. 20. LIFE Research Center for Civilization Diseases, University of Leipzig, Leipzig, Germany. 21. Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark. 22. Department of Statistics & Computational research, Universitat de València, València (Spain). 23. CIBER of Epidemiology and Public Health (CIBERESP). 24. Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Finland. 25. Department of Clinical Physiology and Nuclear Medicine, University of Turku, Turku University Hospital, Turku, Finland. 26. Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland. 27. Department of Obstetrics and Gynecology, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg Sweden. 28. Region Västra Götaland, Sahlgrenska University Hospital, Department of Obstetrics and Gynecology, Gothenburg Sweden. 29. Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 30. Institute of Biomedicine, Physiology, University of Eastern Finland, Kuopio, Finland. 31. Foundation for Research in Health Exercise and Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio, Finland. 32. Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland. 33. Clinical Pharmacology, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom. 34. Department of Nutrition, Exercise and Sports, Faculty of Science, University of Copenhagen. 35. Division of Endocrinology, Metabolism, and Molecular Medicine, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL. 36. The Children's Obesity Clinic, Department of Pediatrics, Copenhagen University Hospital Holbæk, Holbæk, Denmark. 37. Institute of Biomedical and Clinical Science, University of Exeter Medical School, University of Exeter, Royal Devon and Exeter Hospital, Exeter, EX2 5DW, UK. 38. Department of Public Health, Amsterdam Public Health Research Institute, Amsterdam UMC, location Academic Medical Center, University of Amsterdam, 1105 AZ Amsterdam, The Netherlands. 39. Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam Public Health Research Institute, Amsterdam UMC, location Academic Medical Center, University of Amsterdam, 1105 AZ Amsterdam, The Netherlands. 40. Division of Metabolic and Nutritional Medicine, Dr. von Hauner Children's Hospital, Ludwig-Maximilians Universität München (LMU), Munich, Germany. 41. Dpt. of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country (UPV/EHU) Leioa, Spain. 42. Biocruces-Bizkaia Health Research Institute, Barakaldo, Spain. 43. Spanish Biomedical Research Center in Diabetes and Associated Metabolic Disorders (CIBERDEM), Spain. 44. Institute for Molecular Medicine Finland, FIMM, University of Helsinki, Finland. 45. Department of Epidemiology, GRIAC (Groningen Research Institute for Asthma and COPD), University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. 46. Department of Population Medicine, Harvard Medical School, Harvard Pilgrim Health Care Institute, Boston, MA, USA. 47. Division of Epidemiology, Biostatistics, and Environmental Health, University of Memphis, USA. 48. Department of Genetics, University of Pennsylvania. 49. University of Western Australia, Faculty of Health and Medical Sciences. 50. ISGlobal, Barcelona (Spain). 51. Pompeu Fabra University (UPF). 52. Department of General Practice and Primary Health Care, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. 53. Folkhälsan Research Center, Helsinki Finland. 54. Chronic Disease Prevention Unit, National Institute for Health and Welfare, Helsinki, Finland. 55. Pediatrics, Nutrition and Development Research Unit, Universitat Rovira i Virgili, IISPV, Reus, Spain. 56. Department of Radiology and Nuclear Medicine, Erasmus MC University Medical Center, Rotterdam, The Netherlands. 57. Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands. 58. Neonatal Department, Children's Memorial Health Institute, Warsaw, Poland. 59. Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. 60. NIHR Exeter Clinical Research Facility, University of Exeter Medical School and Royal Devon and Exeter NHS Foundation Trust, Exeter, EX2 5DW, UK. 61. Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL. 62. Department of Anthropology, Northwestern University, Evanston, IL. 63. Institute and Outpatient Clinic for Occupational, Social and Environmental Medicine, Inner City Clinic, University Hospital Munich, Ludwig Maximilian University of Munich, Munich, Germany. 64. Department of Genetics and

Bioinformatics, Health Data and Digitalization, Norwegian Institute of Public Health, Oslo, Norway. 65. Faculty of Medicine, University of Southampton, Southampton, UK. 66. Center of Pediatric Research, University Hospital for Children and Adolescents, University of Leipzig, Leipzig, Germany. 67. CHC, St Vincent, Liège-Rocourt, Belgium. 68. Library, University of Eastern Finland, Kuopio, Finland. 69. Department of Biomedical Sciences, University of Copenhagen, Copenhagen N, Denmark. 70. Lunenfeld-Tanenbaum Research Institute Mount Sinai Hospital Joseph & Wolf Lebovic Health Complex. 71. Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 72. Department of Medicine, Stanford University School of Medicine, Stanford, California, USA. 73. National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark. 74. Department of Public Health, Section of Epidemiology, University of Copenhagen, Copenhagen, Denmark. 75. Department of Physiology and Department of Pediatrics, University of Turku, Turku, Finland. 76. Paavo Nurmi Centre, Sports and Exercise Medicine Unit, Department of Health and Physical Activity, University of Turku, Finland. 77. Institute of Translational Genomics, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany. 78. Wellcome Trust Sanger Institute, Hinxton, CB10 1SA, UK. 79. Department of internal medicine, Erasmus MC, University Medical Center Rotterdam, The Netherlands. 80. Basque Government Public Health Department-Gipuzkoa, San Sebastian, Spain. 81. Biodonostia Health Research Institute, San Sebastian, Spain. 82. Spanish Biomedical Research Center in Epidemiology and Public Health (CIBERESP), Spain. 83. Department of Preventive Medicine, Division of Biostatistics, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA. 84. Biocenter Oulu, Oulu, Finland. 85. Division of Immunology and Immunotherapy, Cima Universidad de Navarra. 86. Instituto de Investigación Sanitaria de Navarra (IdISNA). 87. Area de Salut de Menorca ib-salut. 88. Institut d'Investigació Sanitària Illes Balears (IdISBa). 89. Netherlands Genomics Initiative (NGI)-sponsored Netherlands Consortium for Healthy Aging NCHA), Leiden, The Netherlands. 90. Department of Pediatrics, San Paolo Hospital, University of Milan, Milan, Italy. 91. Technical University of Munich (TUM) and Klinikum Rechts der Isar, TUM School of Medicine, Munich, Germany. 92. Department of Genetics, Perelman School of Medicine, University of Pennsylvania. 93. Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, The Netherlands. 94. Diabetes Unit, Massachusetts General Hospital, Boston, MA, USA. 95. Department of Medicine, Université de Sherbrooke, QC, Canada. 96. Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, GRIAC (Groningen Research Institute for Asthma and COPD), University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. 97. Research Unit of Molecular Epidemiology, Institute of Epidemiology, Helmholtz Zentrum Muenchen, Munich, Germany. 98. Harokopi University of Athens, Athens, Greece. 99. Institute of Clinical Research, University of Southern Denmark, Odense, Denmark. 100. Centre of Women's, Family and Child Health, University of South-Eastern Norway, Kongsberg, Norway. 101. Department of Obstetrics and Gynecology, Institute of Clinical Science, Sahlgrenska Academy, University of Gothenburg, Gothenburg Sweden (SA). 102. Region Västra Götaland, Sahlgrenska University Hospital, Department of Obstetrics and Gynecology, Gothenburg Sweden (SU). 103. Department of Genetics and Bioinformatics, Domain of Health Data and Digitalization, Institute of Public Health, Oslo, Norway (FHI). 104. Oxford National Institute for Health Research (NIHR) Biomedical Research Centre, Churchill Hospital, Oxford, UK. 105. Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. 106. Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK. 107. Max Planck Institute for Psycholinguistics, The Netherlands. 108. Donders Institute for Brain, Cognition and Behaviour, Radboud University, The Netherlands. 109. Department of Pediatrics and Adolescents, Haukeland University Hospital, Bergen, Norway. 110. Division of Endocrinology and Diabetes, The Children's Hospital of Philadelphia, Philadelphia, PA, USA. 111. Center for Spatial and Functional Genomics, Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, USA.

#### Author contributions

SV, JFF and VVWJ designed the research and wrote the paper. SV, JFF and JPB analyzed the meta-analysis data. JFF and VVWJ were responsible for the study design of an individual contributing study. JFF, VVWJ and SFAG were responsible for the management of an individual contributing study. SV and SFAG were involved in sample collection and phenotyping of an individual contributing study. SFAG genotyped the data of an individual contributing study.

SV, JFF and JPB analyzed data of an individual contributing study. All authors provided comments and consultation regarding the analyses, interpretation of the results and manuscript. SV, JFF and VVWJ had primary responsibility for final content. All authors gave final approval of the version to be published. All authors read and approved the final manuscript.

#### Funding

##### Amsterdam Born Children and their Development

The ABCD study has been supported by grants from The Netherlands Organisation for Health Research and Development (ZonMw) and Sarphati Amsterdam. Genotyping was funded by the BBMRI-NL grant CP2013-50. Dr M.H. Zafarmand was supported by BBMRI-NL (CP2013-50). Dr. T.G.M. Vrijkotte was supported by ZonMw (TOP 40-00812-98-11010).

##### Avon Longitudinal Study of Parents and Children (ALSPAC)

The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. A comprehensive list of grants funding for ALSPAC is available online (<http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf>). ES, NJT and GDS work in a unit that receives funding from the University of Bristol and the UK Medical Research Council (MC\_UU\_00011/1).

##### CHOP Europe

The research of the CHOP study reported herein was partially supported by the Commission of the European Community, specific RTD Programme "Quality of Life and Management of Living Resources" within the 5th Framework Programme (research grant nos. QLRT-2001-00389 and QLKI-CT-2002-30582); the 6th Framework Programme contract no. 007036 (FP6-007036); the European Union's Seventh Framework Programme Project EarlyNutrition under grant agreement no. 289346 (FP7-289346), the Horizon 2020 research and innovation programme DYNAHEALTH (no. 633595) and the European Research Council Advanced Grant META-GROWTH (ERC-2012-AdG—no. 322605). Additional support from the German Ministry of Education and Research, Berlin (Grant Nr. 01 GI 0825) and the University of Munich Innovative Research Priority Project MC-Health is gratefully acknowledged. This manuscript does not necessarily reflect the views of the Commission and in no way anticipates the future policy in this area. The funders of this study had no role in study design, data collection, data analysis, data interpretation, decision to publish, or preparation of the manuscript.

##### Copenhagen Prospective Studies on Asthma in Childhood (COPSAC Cohorts (COPSAC 2000, COPSAC 2010 and COPSAC registry):

All funding received by COPSAC is listed on [www.copsac.com](http://www.copsac.com). The Lundbeck Foundation (Grant no R16-A1694); The Ministry of Health (Grant no 903516); Danish Council for Strategic Research (Grant no 0603-00280B) and The Capital Region Research Foundation have provided core support to the COPSAC research center.

##### DNBC GOYA-offspring

The Danish National Research Foundation has established the Danish Epidemiology Science Centre that initiated and created the Danish National Birth Cohort (DNBC). The cohort is furthermore a result of a major grant from this foundation. Additional support for the Danish National Birth Cohort is obtained from the Pharmacy Foundation, the Egmont Foundation, the March of Dimes Birth Defects Foundation, the Augustinus Foundation, and the Health Foundation. The DNBC biobank is a part of the Danish National Biobank resource, which is supported by the Novo Nordisk Foundation.

##### The Danish National Birth Cohort-Preterm Birth Study (DNBC-PTB)

DNBC: The Danish National Research Foundation has established the Danish Epidemiology Science Centre that initiated and created the Danish National Birth Cohort (DNBC). The cohort is furthermore a result of a major grant from this foundation. Additional support for the Danish National Birth Cohort is obtained from the Pharmacy Foundation, the Egmont Foundation, the March of Dimes Birth Defects Foundation, the Augustinus Foundation, and the Health Foundation. The DNBC biobank is a part of the Danish National Biobank resource, which is supported by the Novo Nordisk Foundation. The DNBC preterm birth (DNBC-PTB) study is a nested study within the DNBC. The generation of GWAS genotype data for the DNBC-PTB sample was carried out within the Gene Environment Association Studies (GENEVA) consortium with funding provided through the National Institutes of Health's Genes, Environment, and Health Initiative (U01HG004423; U01HG004446; U01HG004438).

##### EFSOCH

The Exeter Family Study of Childhood Health (EFSOCH) was supported by South West NHS Research and Development, Exeter NHS Research and Development, the Darlington Trust and the Peninsula National Institute of



Health Research (NIHR) Clinical Research Facility at the University of Exeter. The opinions given in this paper do not necessarily represent those of NIHR, the NHS or the Department of Health. Genotyping of the EFSOCH study samples was funded by the Wellcome Trust and Royal Society grant 104150/Z/14/Z.

#### **Generation R**

The general design of Generation R Study is made possible by financial support from the Erasmus Medical Center, Rotterdam, the Erasmus University Rotterdam, the Netherlands Organization for Health Research and Development (ZonMw), the Netherlands Organisation for Scientific Research (NWO), the Ministry of Health, Welfare and Sport and the Ministry of Youth and Families. The project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 633595 (DynaHEALTH) and under grant agreements No 733206 (LifeCycle).

#### **HAPO**

HAPO is supported by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and the National Institute of Diabetes and Digestive and Kidney Diseases (R01-HD34242 and R01-HD34243); the National Center for Research Resources (M01-RR00048 and M01-RR00080); and the American Diabetes Association; and grants to local field centers from Diabetes UK (RD04/0002756), Kaiser Permanente Medical Center, KK Women's and Children's Hospital, Mater Mother's Hospital, Novo Nordisk, the Myre Sim Fund of the Royal College of Physicians of Edinburgh, and the Howard and Carol Bernick Family Foundation.

#### **HBCS**

Helsinki Birth Cohort Study has been supported by grants from the Academy of Finland (JGE grant no. 129369, 129907, 135072, 129255 and 126775), the Finnish Diabetes Research Society, Samfundet Folkhälsan, Novo Nordisk Foundation, Finska Läkaresällskapet, Juho Vainio Foundation, Signe and Ane Gyllenberg Foundation, Liv och Hälsa, University of Helsinki and Ministry of Education. JTL has been supported by the Finnish Cultural Foundation.

#### **Infancia y Medio Ambiente [Environment and Childhood] (INMA) Project**

The INMA-GIP cohort was funded by the Spanish Biomedical Research Center in Epidemiology and Public Health (CIBERESP) and by grants from the Instituto de Salud Carlos III (PI06/0867, PI09/00090, PI13/02187 and PI18/01142) cofunded by ERDF, "A way to make Europe", the Basque Department of Health (Projects 2005111093, 2009111069, 2013111089, 2015111065 and 2018111086), the Provincial Government of Gipuzkoa (Projects DFG06/002, DFG08/001 and DFG15/221) and annual agreements with the municipalities of the study area (Zumarraga, Urretxu, Legazpi, Azkoitia y Azepeitia and Beasain).

The INMA Menorca (MEN), Valencia (VAL) and Sabadell (SAB) cohorts were funded by grants from the Instituto de Salud Carlos III (CB06/02/0041, G03/176, FIS PI041436, PI081151, PI041705, and PS09/00432, FIS-FEDER 03/1615, 04/1509, 04/1112, 04/1931, 05/1079, 05/1052, 06/1213, 07/0314, and 09/02647), Fundació La Marató de TV3 (090430), Generalitat de Catalunya-CIRIT (1999SGR 00241), Conselleria de Sanitat Generalitat Valenciana, and Fundación Roger Torné. ISGlobal is a member of the CERCA Programme, Generalitat de Catalunya.

#### **IOW**

National Institutes of Health, USA, award numbers R01 AI091905 and R01 HL132321.

#### **LIFE-Child**

LIFE-Child is funded by the Leipzig Research Center for Civilization Diseases (LIFE). LIFE is an organizational unit affiliated to the Medical Faculty of the University of Leipzig. LIFE is funded by means of the European Union, by the European Regional Development Fund (ERDF) and by funds of the Free State of Saxony within the framework of the excellence initiative. The study was further supported by the German Research Foundation (DFG) for the CRC1052 ObesityMechanisms, project C05.

#### **Lifestyle—Immune System—Allergy Study and German Infant Study on the influence of Nutrition Intervention (LISA + GINI)**

The GINIplus study was mainly supported for the first 3 years of the Federal Ministry for Education, Science, Research and Technology (interventional arm) and Helmholtz Zentrum Munich (former GSF) (observational arm). The 4 year, 6 year, 10 year and 15 year follow-up examinations of the GINIplus study were covered from the respective budgets of the 5 study centres (Helmholtz Zentrum Munich (former GSF), Research Institute at Marien-Hospital Wesel, LMU Munich, TU Munich and from 6 years onwards also from IUF—Leibniz Research-Institute for Environmental Medicine at the University of Düsseldorf) and a grant from the Federal Ministry for Environment (IUF Düsseldorf,

FKZ20462296). Further, the 15 year follow-up examination of the GINIplus study was supported by the Commission of the European Communities, the 7th Framework Program: MeDALL project, and as well by the companies Mead Johnson and Nestlé. The LISA study was mainly supported by grants from the Federal Ministry for Education, Science, Research and Technology and in addition from Helmholtz Zentrum Munich (former GSF), Helmholtz Centre for Environmental Research—UFZ, Leipzig, Research Institute at Marien-Hospital Wesel, Pediatric Practice, Bad Honnef for the first 2 years. The 4 year, 6 year, 10 year and 15 year follow-up examinations of the LISA study were covered from the respective budgets of the involved partners (Helmholtz Zentrum Munich (former GSF), Helmholtz Centre for Environmental Research—UFZ, Leipzig, Research Institute at Marien-Hospital Wesel, Pediatric Practice, Bad Honnef, IUF—Leibniz-Research Institute for Environmental Medicine at the University of Düsseldorf) and in addition by a grant from the Federal Ministry for Environment (IUF Düsseldorf, FKZ 20462296). Further, the 15-year follow-up examination of the LISA study was supported by the Commission of the European Communities, the 7th Framework Program: MeDALL project.

#### **Norwegian Mother and Child (MoBa) Cohort Study (Pål R. Njølstad)**

MOBA: The Norwegian Mother and Child Cohort Study was supported by grants from the European Research Council (AdG #293574), the Bergen Research Foundation ("Utilizing the Mother and Child Cohort and the Medical Birth Registry for Better Health"), Stiftelsen Kristian Gerhard Jebsen (Translational Medical Center), the University of Bergen, the Research Council of Norway (FRIPRO grant #240413), the Western Norway Regional Health Authority (Strategic Fund "Personalized Medicine for Children and Adults", and Open Grants «Understanding infant weight biology through genomics and deep phenotyping» grant #912250), and the Norwegian Diabetes Foundation; the Research Council of Norway through its Centres of Excellence funding scheme (#262700), Better Health by Harvesting Biobanks (#229624).

#### **Norwegian Mother and Child (MoBa) Cohort Study (Bo Jacobsson)**

The analyses were supported by the Research Council of Norway (grant FRIMEDBIO ES547711 to Bo Jacobsson), the March of Dimes Foundation Prematurity Research Initiative Program (grant 21-FY16-121 to Bo Jacobsson).

#### **Northern Finnish Birth Cohort Study 1966 (NFBC66) and 1968 (NFBC86)**

The Northern Finland Birth Cohort (NFBC) Research program, including the study of the NFBC1966 and 1986 studies, received financial support from Academy of Finland (1114194, 24300796), NHLBI grant 5R01HL087679 through the STAMPEED program (1RL1MH083268-01), ENGAGE project and grant agreement HEALTH-F4-2007-201413, the Medical Research Council (grant G0500539, centre grant G0600705, PrevMetSyn), and the Wellcome Trust (project grant GR069224), UK. The program is currently being funded by the H2020-633595 DynaHEALTH action and Academy of Finland EGEA-project.

#### **The Physical Activity and Nutrition in Children (PANIC) Study**

The PANIC Study has financially been supported by grants from Ministry of Education and Culture of Finland, Ministry of Social Affairs and Health of Finland, Research Committee of the Kuopio University Hospital Catchment Area (State Research Funding), Finnish Innovation Fund Sitra, Social Insurance Institution of Finland, Finnish Cultural Foundation, Foundation for Paediatric Research, Diabetes Research Foundation in Finland, Finnish Foundation for Cardiovascular Research, Juho Vainio Foundation, Paavo Nurmi Foundation, Yrjö Jahnnson Foundation, and the city of Kuopio. The PhD students and post-doctoral researchers of The PANIC Study have financially been supported by personal grants from Program for Clinical Research and Program for Health Sciences of Doctoral School of University of Eastern Finland, Finnish Doctoral Programs in Public Health, Finnish Doctoral Program in Oral Sciences, European Commission Marie Curie Fellowship Programme, North Savo Regional Fund of Finnish Cultural Foundation, Juho Vainio Foundation, Päiviikki and Sakari Sohlberg Foundation, Foundation for Paediatric Research, Diabetes Research Foundation in Finland, Finnish Foundation for Cardiovascular Research, Emil Aaltonen Foundation, Paulo Foundation, Jalmary and Rauha Ahokas Foundation, Aarne and Aili Turunen Foundation, Finnish Medical Foundation, Jenny and Antti Wihuri Foundation, Orion Foundation, Urho Känkänen Foundation, Finnish Association for the Study of Pain, Finnish Dental Society Apollonia, Northern Savo Dental Association, Health Centre Dentist Association, Finnish Female Dentists Association, Kuopio Naturalists' Society, Olvi Foundation, and the city of Kuopio.

#### **Prevention and incidence of asthma and mite allergy birth cohort study—(PIAMA)**

The PIAMA study was funded by grants from the Dutch Asthma Foundation (grant 3.4.01.26, 3.2.06.022, 3.4.09.081 and 3.2.10.085CO), the ZON-MW

Netherlands Organization for Health Research and Development (grant 912-03-031), the Stichting Astmabestrijding and the Ministry of the Environment. Genome-wide genotyping was funded by the European Commission as part of GABRIEL (A multidisciplinary study to identify the genetic and environmental causes of asthma in the European Community) contract number 018996 under the Integrated Program LSH-2004-1.2.5-1 Post genomic approaches to understand the molecular basis of asthma aiming at a preventive or therapeutic control and a Grant from BBMRI-NL (CP 29).

#### Project Viva

This study was supported by the National Institutes of Health (UG3OD23286, R01 HD034568, and R01 AI102960).

#### The Raine Study

We are grateful to the Raine Study participants and their families and we thank the Raine Study and Lions Eye Institute research staff for cohort coordination and data collection. The Raine Study acknowledges the National Health and Medical Research Council (NHMRC) for their long term contribution to funding the study over the last 29 years. The core management of the Raine Study has been funded by the University of Western Australia, Curtin University, Telethon Kids Institute, Women and Infants Research Foundation, Edith Cowan University, Murdoch University, The University of Notre Dame Australia, and the Raine Medical Research Foundation. The collection of the GWAS data for Gen2 of the Raine Study participants were supported by the Canadian Institutes of Health Research—CIHR (MOP-82893) and the National Health and Medical Research Council of Australia [grant numbers 634445, 634509, and 1021105]. The authors gratefully acknowledge the assistance of the Western Australian DNA Bank (National Health and Medical Research Council of Australia National Enabling Facility). This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and Government of Western Australia.

#### SKOT (1 & 2)

The SKOT 1 and 2 study have their base at the Department of Human Nutrition, Faculty of Life Sciences, University of Copenhagen and was supported by grants from The Danish Directorate for Food, Fisheries and Agri Business as part of the ‘Complementary and young child feeding (CYCF)—impact on short- and long-term development and health’ project and partly by contributions from the research program ‘Governing Obesity’ by the University of Copenhagen Excellence Program for Interdisciplinary Research ([www.go.ku.dk](http://www.go.ku.dk)).

#### Special Turku Coronary Risk Factor Intervention Project (STRIP)

STRIP was supported by the Academy of Finland (grants 206374, 294834, 251360, and 275595), the Juho Vainio Foundation, the Finnish Foundation for Cardiovascular Research, the Finnish Ministry of Education and Culture, the Finnish Cultural Foundation, the Sigrid Juselius Foundation, Special Governmental grants for Health Sciences Research (Turku University Hospital), the Yrjö Jahnsson Foundation, and the Turku University Foundation.

#### TDCOB

TDCOB received support from the Danish Innovation Foundation and The Region Zealand Health and Medical Research Foundation. This study was partly supported by grants from the Program Committee for Individuals, Disease and Society of the Danish Innovation Foundation (grant numbers: 0603-00484B [TARGET] and 0603-00457B [BIOCHILD]). The Novo Nordisk Foundation Center for Basic Metabolic Research, an independent Research Center at the University of Copenhagen, is partially funded by an unrestricted donation from the Novo Nordisk Foundation.

#### TEENS of Attica: Genes and Environment Study (TEENAGE)

TEENAGE study has been co-financed by the European Union (European Social Fund—ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF)—Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund. This work was funded by the Wellcome Trust (098051).

#### Personal funding

EZ received funding from Wellcome Trust (098051). CSM was funded by the University of Southern Denmark. BSP is supported through funds of the Max Planck core society and the Simons Foundation (Award ID: 514787). NJT is a Wellcome Trust Investigator (202802/Z/16/Z), is the PI of the Avon Longitudinal Study of Parents and Children (MRC & WT 102215/Z/13/Z), is supported by the University of Bristol NIHR Biomedical Research Centre

(BRC-1215-20011), the MRC Integrative Epidemiology Unit (MC\_UU\_12013/3) and works within the CRUK Integrative Cancer Epidemiology Programme (C18281/A19169). BSTP is funded by Max Planck Society core funds and the Simons Foundation (514787). BK received personal funding by the European Research Council Advanced Grant META-GROWTH (ERC-2012-AdG—no. 322605). RNB and RMF are funded by a Wellcome Trust and Royal Society Sir Henry Dale Fellowship, grant 104150/Z/14/Z. VK was funded by Marie Curie CAPICE EU project (European Union's Horizon 2020 research and innovation programme under grant agreement no. 721567). SFAG was financially supported by an Institute Development Award from the Children's Hospital of Philadelphia, a Research Development Award from the Cotswold Foundation, the Daniel B. Burke Endowed Chair for Diabetes Research, the Children's Hospital of Philadelphia Endowed Chair in Genomic Research and NIH grant R01 HD056465. NFJ received funding from Basque Dpt. of Health (SAN2018111086). HHHA was supported by ZonMW grant number 916.19.151. MHZ was supported by BBMRI-NL (CP2013-50). TGMV was supported by ZonMW (TOP 40–00812–98–11010). XL was supported by the Nordic Center of Excellence in Health-Related e-Sciences. BF was supported by the Oak Foundation and the Novo Nordisk Foundation (12955). MFH is supported by ADA Pathways to Stop Diabetes award 1-15-ACE-26. VVWJ received funding from the Netherlands Organization for Health Research and Development (VIDI 016.136.361) and the European Research Council (ERC-2014-CoG-648916). MMcC is a Wellcome Senior Investigator and an NIHR Senior Investigator. Relevant funding support for this work comes from Wellcome (090532, 106130, 098381, 203141, 212259), NIDDK (U01-DK105535), and NIHR (NF-SI-0617-10090).

#### Availability of data and materials

GWAS summary data will be deposited at the EGG website (<https://egg-consortium.org/>) at publication. Individual study data are available from the corresponding author on reasonable request.

#### Declarations

##### Ethics approval and consent to participate

All studies were approved by the national research ethics committee. Study-specific details are shown in the acknowledgements and supplemental cohort information. Informed consent to participate was obtained for all participants. Detailed and study-specific information can be found in the acknowledgements and supplemental cohort information.

##### Consent for publication

Not applicable.

##### Competing interests

MMcC: The views expressed in this article are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. He serves on advisory panels for Pfizer, NovoNordisk, Zoe Global; has received honoraria from Merck, Pfizer, NovoNordisk and Eli Lilly; has stock options in Zoe Global; has received research funding from Abbvie, Astra Zeneca, Boehringer Ingelheim, Eli Lilly, Janssen, Merck, NovoNordisk, Pfizer, Roche, Sanofi Aventis, Servier & Takeda. Ms Receives funding from Pfizer Inc. for a project not related to this research. The other authors declare that they have no competing interests.

##### Author details

<sup>1</sup>The Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. <sup>2</sup>Department of Pediatrics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. <sup>3</sup>Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands. <sup>4</sup>Division of Human Genetics, Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA. <sup>5</sup>Quantinum Research LLC, San Diego, CA, USA. <sup>6</sup>Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. <sup>7</sup>Division of Endocrinology and Diabetes, The Children's Hospital of Philadelphia, Philadelphia, PA, USA. <sup>8</sup>Division of Human Genetics, Center for Spatial and Functional Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA.

Received: 20 August 2021 Accepted: 30 May 2022

Published online: 04 June 2022

## References

- Tisserand DJ, Bosma H, Van Boxtel MP, Jolles J. Head size and cognitive ability in nondemented older adults are related. *Neurology*. 2001;56(7):969–71.
- Gale CR, O'Callaghan FJ, Bredow M, Martyn CN, Avon Longitudinal Study of P, Children Study T. The influence of head growth in fetal life, infancy, and childhood on intelligence at the ages of 4 and 8 years. *Pediatrics*. 2006;118(4):1486–92.
- Bartholomeusz HH, Courchesne E, Karns CM. Relationship between head circumference and brain volume in healthy normal toddlers, children, and adults. *Neuropediatrics*. 2002;33(5):239–41.
- Kiserud T, Piaggio G, Carroli G, Widmer M, Carvalho J, Jensen LN, et al. Correction: The World Health Organization fetal growth charts: a multinational longitudinal study of ultrasound biometric measurements and estimated fetal weight. *PLoS Med*. 2021;18(1):e1003526.
- Rollins JD, Collins JS, Holden KR. United States head circumference growth reference charts: birth to 21 years. *J Pediatr*. 2010;156(6):907–13. e2.
- Smit DJ, Luciano M, Bartels M, van Beijsterveldt CE, Wright MJ, Hansell NK, et al. Heritability of head size in Dutch and Australian twin families at ages 0–50 years. *Twin Res Hum Genet*. 2010;13(4):370–80.
- Adams HH, Hibar DP, Chouraki V, Stein JL, Nyquist PA, Renteria ME, et al. Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat Neurosci*. 2016;19(12):1569–82.
- Haworth S, Shapland CY, Hayward C, Prins BP, Felix JF, Medina-Gomez C, et al. Low-frequency variation in TP53 has large effects on head circumference and intracranial volume. *Nat Commun*. 2019;10(1):357.
- Ikram MA, Fornage M, Smith AV, Seshadri S, Schmidt R, Debette S, et al. Common variants at 6q22 and 17q21 are associated with intracranial volume. *Nat Genet*. 2012;44(5):539–44.
- Jansen PR, Nagel M, Watanabe K, Wei Y, Savage JE, de Leeuw AC, et al. GWAS of brain volume on 54,407 individuals and cross-trait analysis with intelligence identifies shared genomic loci and genes. *bioRxiv*. 2019.
- Taal HR, St Pourcain B, Thiering E, Das S, Mook-Kanamori DO, Warrington NM, et al. Common variants at 12q15 and 12q24 are associated with infant head circumference. *Nat Genet*. 2012;44(5):532–8.
- Guo LH, Alexopoulos P, Wagenpfeil S, Kurz A, Perneckzy R, Alzheimer's Disease Neuroimaging I. Brain size and the compensation of Alzheimer's disease symptoms: a longitudinal cohort study. *Alzheimers Dement*. 2013;9(5):580–6.
- Perneckzy R, Wagenpfeil S, Lunetta KL, Cupples LA, Green RC, Decarli C, et al. Head circumference, atrophy, and cognition: implications for brain reserve in Alzheimer disease. *Neurology*. 2010;75(2):137–42.
- Brown AS, Gyllenberg D, Hinkka-Yli-Salomaki S, Sourander A, McKeague IW. Altered growth trajectory of head circumference during infancy and schizophrenia in a National Birth Cohort. *Schizophr Res*. 2017;182:115–9.
- Hazlett HC, Poe M, Gerig G, Smith RG, Provenzale J, Ross A, et al. Magnetic resonance imaging and head circumference study of brain size in autism: birth through age 2 years. *Arch Gen Psychiatry*. 2005;62(12):1366–76.
- Fukumoto A, Hashimoto T, Mori K, Tsuda Y, Arisawa K, Kagami S. Head circumference and body growth in autism spectrum disorders. *Brain Dev*. 2011;33(7):569–75.
- Courchesne E, Carper R, Akshoomoff N. Evidence of brain overgrowth in the first year of life in autism. *JAMA*. 2003;290(3):337–44.
- Bartley AJ, Jones DW, Weinberger DR. Genetic variability of human brain size and cortical gyral patterns. *Brain*. 1997;120(Pt 2):257–69.
- Wikland KA, Luo ZC, Niklasson A, Karlberg J. Swedish population-based longitudinal reference values from birth to 18 years of age for height, weight and head circumference. *Acta Paediatr*. 2002;91(7):739–54.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics*. 2010;26(17):2190–1.
- Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291–5.
- Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, 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):272–9.
- Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ATC, Replication DIG, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet*. 2012;44(4):369–75, S1–3.
- Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826.
- GTEX Consortium. Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348(6235):648–60.
- Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*. 2014;10(5):e1004383.
- UniProt C. UniProt: a hub for protein information. *Nucleic Acids Res*. 2015;43(Database issue):D204–12.
- da Huang W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37(1):1–13.
- Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47(11):1236–41.
- Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700,000 individuals of European ancestry. *Hum Mol Genet*. 2018;9:621.
- Dickinson ME, Flenniken AM, Ji X, Teboul L, Wong MD, White JK, et al. High-throughput discovery of novel developmental phenotypes. *Nature*. 2016;537(7621):508–14.
- Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet*. 2018;50(7):912–9.
- Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet*. 2019;51(3):404–13.
- Nagel M, Jansen PR, Stringer S, Watanabe K, de Leeuw CA, Bryois J, et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat Genet*. 2018;50(7):920–7.
- Lee JJ, Wedow R, Okbay A, Kong E, Maghziyan O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet*. 2018;50(8):1112–21.
- Vinas-Guasch N, Wu YJ. The role of the putamen in language: a meta-analytic connectivity modeling study. *Brain Struct Funct*. 2017;222(9):3991–4004.
- Yang Y, Wang JZ. From structure to behavior in basolateral amygdala-hippocampus circuits. *Front Neural Circuits*. 2017;11:86.
- Affi AK. The basal ganglia: a neural network with more than motor function. *Semin Pediatr Neurol*. 2003;10(1):3–10.
- Lockwood PL, Wittmann MK. Ventral anterior cingulate cortex and social decision-making. *Neurosci Biobehav Rev*. 2018;92:187–91.
- Chen S, Tang J, Huang L, Lin J. Expression and prognostic value of Mycl1 in gastric cancer. *Biochem Biophys Res Commun*. 2015;456(4):879–83.
- Chen S, Zheng Z, Tang J, Lin X, Wang X, Lin J. Association of polymorphisms and haplotype in the region of TRIT1, MYCL1 and MFS2A with the risk and clinicopathological features of gastric cancer in a southeast Chinese population. *Carcinogenesis*. 2013;34(5):1018–24.
- Yaylim-Eraltan I, Bozkurt N, Ergen A, Zeybek U, Ozturk O, Arikan S, et al. L-myc gene polymorphism and risk of thyroid cancer. *Exp Oncol*. 2008;30(2):117–20.
- Liu LM, Xiong DD, Lin P, Yang H, Dang YW, Chen G. DNA topoisomerase 1 and 2A function as oncogenes in liver cancer and may be direct targets of nitidine chloride. *Int J Oncol*. 2018;53(5):1897–912.
- Lv C, Liu X, Zheng Q, Chen H, Yang X, Zhong J, et al. Analysis of topoisomerase I expression and identification of predictive markers for efficacy of topotecan chemotherapy in small cell lung cancer. *Thorac Cancer*. 2018;9(9):1166–73.

45. Banne E, Atawneh O, Henneke M, Brockmann K, Gartner J, Elpeleg O, et al. West syndrome, microcephaly, grey matter heterotopia and hypoplasia of corpus callosum due to a novel ARFGEF2 mutation. *J Med Genet.* 2013;50(11):772–5.
46. Tanyalcin I, Verhelst H, Halley DJ, Vanderhasselt T, Villard L, Goizet C, et al. Elaborating the phenotypic spectrum associated with mutations in ARFGEF2: case study and literature review. *Eur J Paediatr Neurol.* 2013;17(6):666–70.
47. Smith GD, Ebrahim S. “Mendelian randomization”: can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol.* 2003;32(1):1–22.
48. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet.* 2014;23(R1):R89–9.
49. Harris SR. Measuring head circumference: update on infant microcephaly. *Can Fam Phys.* 2015;61(8):680–4.
50. Ivanovic DM, Leiva BP, Perez HT, Inzunza NB, Almagia AF, Toro TD, et al. Long-term effects of severe undernutrition during the first year of life on brain development and learning in Chilean high-school graduates. *Nutrition.* 2000;16(11–12):1056–63.
51. Wiles NJ, Peters TJ, Heron J, Gunnell D, Emond A, Lewis G. Fetal growth and childhood behavioral problems: results from the ALSPAC cohort. *Am J Epidemiol.* 2006;163(9):829–37.
52. Hagenaars SP, Harris SE, Davies G, Hill WD, Liewald DC, Ritchie SJ, et al. Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N = 112,151) and 24 GWAS consortia. *Mol Psychiatry.* 2016;21(11):1624–32.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

