

Supplementary Notes and Figures

A. Cohort Descriptions

1) MoBa

The Norwegian Mother, Father and Child Cohort Study (MoBa)^{1,2} is a population-based pregnancy cohort from across Norway consisting of approximately 114,500 children, 95,200 mothers and 75,200 fathers³. Participants were recruited through hospitals, firstly in Bergen starting in 1999, and then expanding to 50 of 52 Norway's hospitals with maternity units. All pregnant women were eligible for inclusion and 41% accepted of the 277,702 invited between 1999 and 2008 (N = 112,908). About 1,900 twin pairs were born in the MoBa study, and about 16,400 women participated in MoBa with multiple different pregnancies². Due to the wide inclusion criteria, there are numerous related families within the MoBa sample giving it a complex relatedness structure³.

Phenotype

The current study is based on version 12 of the quality-assured questionnaire data files released for research in January 2019. The phenotype and genetic data analysed were prepared on the 5th of December 2022. Information on gross motor skills was collected by parental questionnaire at 18 months and 36 months after birth. The parental questionnaire administered when children were 18 months asked 'How many months was your child when he/she took his/her first steps unaided (questionnaire version A and B; variable EE400) (N = 17,488) or "Can your child walk unaided? If yes, how old was your child when he/she could first walk unaided?" (questionnaire version C, D and E; variable EE987) (N = 56,715). If data was unavailable for a participant, including if independent walking had not yet been attained at 18 months, responses from a questionnaire administered when children were 36 months was used, being "How many months old was your child when he/she took his/her first steps unaided?" (all questionnaire versions – no edits made; variable GG27) (N = 5,594).

Phenotype descriptive statistics (genotyped participants only) for MoBa are provided in the Supplementary Table S1.

Genotype and Quality Control (QC)

Blood samples from infants were taken from the umbilical cord at birth⁴. Genotyping of the samples took place at different times as part of different projects over several years. Including mothers, fathers and children, 238,001 samples were genotyped in 24 batches (two of which were split into sub-batches). The arrays used for the different batches include: Illumina HumanCoreExome12v1.1, Illumina HumanCoreExome24v1.0, Illumina HumanOmniExpress24v1.0, Illumina HumanOmniExpress24v1.2, Illumina Global Screening Array MD v1.0, Illumina Global Screening Array MD v3.0., and Illumina InfiniumOmniExpress-24v1.2.

Of the 238,001 samples genotyped in 26 batches, 235,412 were successfully genotyped and 234,505 of those were retained in the cohort after excluding those who had not withdrawn consent. Based on the principal components analysis, approximately 95% of the sample were identified as having European ancestry. The publicly available European Haplotype Reference Consortium (HRC)⁵ release 1.1 was used as the reference panel for phasing and imputation. Full details of the genotyping and QC pipeline, including the list of researchers involved in this process, can be found here: ³. After post-imputation quality control, the dataset included 207,569 individuals (including 76,577 children) and 6,981,748 SNPs. Further post-imputation QC in MoBa as well as the primary and sex stratified Genome-Wide Association Study (GWAS) were conducted by Anja Hollowell and Anna Gui for children who had phenotype data concerning motor or behavioural traits (N = 67,407). At this stage, individuals were further excluded if they had genotyping rate < 95%. SNPs were excluded if they had minor allele frequency (MAF) < 0.5%, failed Hardy Weinberg Equilibrium (HWE) exact test at $p < 1 \times 10^{-6}$, or had call-rate < 98%. In addition, to exclude potential duplicates, for each of the 33 pairs samples that were genetically identical ($P(\text{identity by descent or IBD}=0) < 0.1$ & $P(\text{IBD}=1) < 0.1$, $\text{PI-HAT} > 0.8$), and also identical across 12 infant phenotype measures (age at onset of walking, two other motor measures, and 9 behavioural measures), one of the pair was excluded from the data at random.

The current GWAS included children with complete information on age at onset of walking, covariates (sex, year of birth, genotyping batch, imputation batch and the first 10 principal components, PCs), and genotype data ($N = 58,302$). Sex-stratified GWAS analyses were performed on 29,846 males, and 28,456 females, including year of birth, genotyping batch, imputation batch and the first 10 PCs as covariates in the model. The QQ-plot for the MoBa GWAS is shown in the Supplementary Fig. S1.

Additional checks for inflation in the MoBa samples

Since the Genomic Control (GC) factor lambda (λ_{GC}) for the MoBa GWAS was $\lambda_{GC} = 1.228$, we ran some additional checks to verify whether the inflation observed in the QQ-plot (Supplementary Fig. S1) was due to bias linked to the sample ancestry and relatedness, phenotypic outliers, covariate-phenotype interactions or variant allele frequencies, variant LD patterns or more likely reflected polygenicity. These checks are summarised in Table SM1 below.

The MoBa sample only included individuals of European ancestry and the GWAS model included 10 PCs to correct for residual population stratification. Additionally, the λ_{GC} for a GWAS on a different infant motor phenotype collected at the same age from a largely overlapping sample ($N = 55,227$), namely fine motor skills from parental report at 18 months was 1.040. This result strongly suggested that the observed inflation was not due to population stratification but was linked to the age at onset of walking phenotype specifically. Additionally, the QQ-plot by allele frequency (see Fig. S1) showed that the observed pattern was not changed by further filtering by allele frequency. Thus, we performed additional GWAS analyses and computed the λ_{GC} , to see whether inflation was caused by particular characteristics of the data.

To check whether the observed inflation pattern in the QQ-plot was due to the presence of sample outliers, we randomly split the sample in halves and computed the λ_{GC} for both samples ($N = 29,151$ each, $\lambda_{GC} = 1.117$ and 1.123). We then repeated this operation and computed the lambda GC for the new two-halves ($N = 29,151$ each, $\lambda_{GC} = 1.112$ and 1.131 respectively). The similarity in lambdas across the two halves suggest there is not a specific individual/subset of individuals driving the high total sample lambda. The lower lambdas for

the half-samples is consistent with the idea that inflation reflected polygenicity, as smaller sample sizes reduce power therefore the Chi-squared statistics are lower.

The age at onset of walking phenotype for MoBa ranged from 6 months to 33 months and was not normally distributed (skew=0.91). To verify whether inflation could be caused by the presence of phenotype outliers, we excluded individuals who walked later than 25 months of age, which corresponded to the latest age at onset of walking reported in the other cohorts (as summarised in the Supplementary Table S1). The λ_{GC} for this GWAS was 1.230, suggesting that extreme phenotype outliers were not causing genomic inflation in the primary MoBa GWAS. Further, to exclude that this effect was caused by skewness in the phenotype distribution, we excluded all phenotype values above 3 SD from the mean of the whole sample, that is those who learnt to walk later than 18 months. Again, we found the $\lambda_{GC} = 1.233$, indicating that inflation was not due to the presence of phenotype outliers.

We then tested whether the observed pattern of inflation in the MoBa GWAS was due to the inclusion of specific covariates in the GWAS model. When excluding genotyping and imputation batch covariates, the λ_{GC} was 1.228; when excluding the year of birth covariate, the λ_{GC} was 1.226; when excluding year of birth and batches from the model, leaving only sex and the 10 PCs as covariates, the λ_{GC} was 1.225. These analyses suggested inflation was not caused by the covariates included in the primary MoBa GWAS.

Since one of the most significant signals in the MoBa GWAS was on Chromosome 6 in the HLA region, we conducted an additional check by removing chromosome 6 from the GWAS to verify whether the LD pattern in that part of the genome caused inflation. However, λ_{GC} for the GWAS without chromosome 6 was 1.222, similar to the λ_{GC} for the primary MoBa GWAS.

Last, the LD Score Regression (LDSC) intercept for the primary MoBa GWAS was 1.008 (0.008), confirming that the inflation observed in the MoBa GWAS QQ-plot was likely due to polygenicity of the phenotype rather than confounding factors such as relatedness and population stratification⁶. Indeed, our MiXeR⁷ analyses confirmed that the age at onset of walking phenotype is highly polygenic, with 11.9K variants explaining 90% of the SNP heritability. Other highly polygenic traits had values of λ_{GC} higher than 1.2. For example, for ADHD⁸ the MiXeR polygenicity estimate was 7.3K variants explaining 90% of the SNP heritability and λ_{GC} was 1.307, for bipolar disorder⁹ the MiXeR polygenicity estimate was 8.6K and λ_{GC} was 1.38). Moreover, inflation was observed for GWAS of many polygenic

quantitative traits on the UK BioBank dataset, which is an ancestrally homogeneous sample where all related individuals had been excluded (for example, sitting height showed a λ_{GC} of 1.62 for SNPs with MAF > 0.5 and $\lambda_{GC} = 4.034$ when including all SNPs with MAF > 0.01, <https://pheweb.org/UKB-Neale/pheno/20015> , or heel bone mineral density with a λ_{GC} of 1.20 for MAF > 0.5 and $\lambda_{GC} = 1.605$ for MAF > 0.01 <https://pheweb.org/UKB-Neale/pheno/4125#qq> , see the NealeLab blog for a discussion regarding inflation for polygenic traits: <https://www.nealelab.is/blog/2017/9/11/details-and-considerations-of-the-uk-biobank-gwas>). Further, a similar QQ plot showing a “step-like” line linked to LD was observed in the recent GWAS of educational attainment¹⁰ for SNPs with MAF > 5% (Extended Fig.1 a, $\lambda_{GC} = 3.55$).

Based on these additional checks, and because all other samples had λ_{GC} below the 1.10 recommended threshold¹¹, we decided not to apply automatic correction for genomic control based on the λ_{GC} in the GWAS meta-analysis, to avoid removing meaningful signal.

Table SM1. Summary of the inflation factors for all the analyses performed to examine possible reasons for the high genomic inflation.

Possible reason	Performed analysis	Lambda GC
The inflation is due to population stratification or relatedness in the sample	GWAS of fine motor skills at 18 months in a largely overlapping sample	1.040
The inflation is due to a specific individual/subset of individual outliers	Randomly split the sample into two halves.	1.117
		1.123
	Repeat random split into two halves.	1.112
		1.131
The inflation is due to phenotype outliers	Exclude individuals who walked later than 25 months (upper limit in NSHD and NTR cohorts)	1.230
	Exclude individuals with phenotype values exceeding 3 standard deviations from the sample mean	1.233

The inflation is due to the interaction between covariates and the phenotype	Exclude genotyping and imputation batch covariates	1.228
	Exclude the year of birth covariate	1.226
	Exclude the year of birth, genotyping and imputation batch covariates	1.225
The inflation is due to the LD pattern in the HLA region	Exclude chromosome 6 from the GWAS	1.222

Ethics

MoBa and the related data collection was authorised by a licence from the Norwegian Data Protection Agency and an approval from the The Regional Committees for Medical and Health Research Ethics (REK). MoBa is regulated by the Norwegian Health Registry Act. The current study was approved by REK (2016/1702).

2) Netherlands Twin Register (NTR)

The Netherlands Twin Register (NTR)^{20,21} is a population-based cohort of over 200,000 people from across the Netherlands. It consists of twin-families, i.e. twins, their parents, spouses and siblings aged between 0 and 99 years at recruitment. NTR started around 1987 with new-born twins and adolescent and adult twins. Full details have been reported previously²². The current sample includes monozygotic and dizygotic twins and triplets who were recruited shortly after birth.

Phenotype

Information on age of walking without support was obtained from NTR survey YS-2, which was mailed to the parents after the second birthday of the children (mean age 2.34 years, SD 0.25). The parents (usually the mother of twins) were asked “With how many months could your twin walk without support for the first time? in half months?”. With the first survey that

was sent to parents, typically a few weeks/months after registration we informed them that they would be asked for this information in the next survey and included an example of the answer sheet. The procedure for collection of the information was validated in a large study^{18,19}. The answers on age of walking without support ranged from 9.5 to 21 months (mean 15.13, SD 2.36, median 15, interquartile range 3).

Phenotype descriptive statistics for the NTR sample included in the present GWAS are provided in the Supplementary Table S1.

Genotyping and QC

Buccal cells or blood samples for DNA isolation were collected in multiple NTR projects, thus genotype data was obtained from participants at different time points. Genotyping in NTR was done on multiple platforms over time, including Perlegen-Affymetrix, Affymetrix 6.0, Affymetrix Axiom, Illumina Human Quad Bead 660, Illumina Omni 1M and Illumina GSA, in total on 29,943 individuals. On each platform, genotyping was performed following manufacturers protocols, using the appropriate calling software. For each genotype platform, samples were removed if DNA sex did not match the expected phenotype, if the PLINK heterozygosity F statistic was < -0.10 or > 0.10 , or if the genotyping call rate was < 0.90 . DNA IBD estimations were made for each platform and samples were removed if they did not match the expected relationships (PLINK¹⁹, KING²³). SNPs were removed if the MAF $< 1 \times 10^{-6}$, if the HWE p-value $< 1 \times 10^{-6}$, and if call rate < 0.95 .

Subsequently, for each platform, genotype data were aligned with 1000 Genomes reference¹⁵ panel using the HRC and 1000 Genomes checking tool, which tests and filters for SNPs with allele frequency differences larger than 0.20 as compared to the CEU population, palindromic SNPs and DNA strand and position issues. The data of the 6 platforms were then merged into a single data set keeping all QC-ed SNPs of each platform. For each individual, only one platform was chosen. Based on the ~10.8K SNPs that all platforms have in common, IBD state was then estimated for all individual pairs using the PLINK and KING programs. These estimates were then compared to the expected familial relations, and any individuals whose genetic relatedness estimate did not match their expected familial relations were removed.

Principal components were calculated for each platform after genotype QC, with additional LD SNP pruning ($VIF \leq 2$), in order to exclude CEU population outliers, based on 1000 genomes population projection with the SMARTPCA software²⁴. Then, per platform, data were phased using EAGLE²⁵ and then imputed to 1000 Genomes¹⁵ with Minimac-3²⁶. Post-imputation, the resulting separate platform VCFs were merged into a single VCF file per chromosome using BCFTOOLS for all SNPs present on each of the 6 platforms.

The current GWAS included children with complete information on age at onset of walking without support, covariates (age at parents' report, sex and platform), and genotype data ($N = 6,251$, born between 1986-2016). Sex-stratified analyses were performed in 2,852 males, and 3,399 females. In this case 2 platform covariates, minus 1 to avoid collinearity were added, depending on the individuals with phenotypes used in the study. The described QC steps and GWAS were conducted by Jouke-Jan Hottenga and Veronika Odintsova. The QQ-plot from the primary NTR GWAS is shown on Supplementary Fig. S2.

Ethics

Informed consent was obtained from parents or guardians. The study was approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre, Amsterdam, an Institutional Review Board certified by the U.S. Office of Human Research Protections (IRB number IRB00002991 under Federal-wide Assurance-FWA00017598; IRB/institute codes, NTR 03-180).

3) The Lifelines Cohort Study (Lifelines)

Lifelines is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviours of 167,729 persons living in the North of the Netherlands. It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioural, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on multi-morbidity and complex genetics^{12,13}. The study was established in 2006 and asked all the general medical practitioners operating in the provinces of Friesland, Groningen and

Drenthe to invite their patients aged between 25 and 50 to take part in the study unless the patient met one of five exclusion criteria (as determined by the practitioner): a) having a severe psychiatric or physical illness (such that the individual was not fully capable to make rational decisions), b) life expectancy of less than 5 years, c) being unable to complete a Dutch language questionnaire, d) a lack of ability in the Dutch language, or e) not being able to visit their medical practitioner (<http://wiki-lifelines.web.rug.nl/doku.php?id=cohort>). Those who were interested in participating in the study were required to give written consent and then complete a baseline questionnaire. Parents of children could also volunteer their children for participation. Residents of the relevant provinces were also able to join the study through the study website^{13,14}.

Phenotype

On admission, Lifelines study participants completed questionnaires which, for parents of children taking part in the study, included questions on the child's development and behaviour. During the first assessment of participants, known as the baseline assessment, parents were provided with questionnaire 1A which, amongst other things, asked "When did your child start walking (3-5 steps without help)?" (N ~12,000). We refer to Supplementary Table S1 for information about the phenotype coding and the descriptives for age at onset of walking phenotype for the Lifelines sample included in the present GWAS as providing both phenotype and good quality genotype data (see below).

Genotype and QC

During a visit to one of the twelve research sites set up in their local areas, participants provided blood samples. The blood samples were analysed at the Lifelines laboratory and genetic information extracted for sub-samples of Lifelines.

Genetic data was collected as part of two subsequent studies of which only the second collected data from children. This study was called the University Medical Centre in Groningen (UMCG) Genetics Lifelines Initiative (UGLI) as part of which genotyped 38,030 Lifelines participants.

The UGLI data from approximately 38,000 participants (which includes all the genotyped children) was genotyped using the Infinium Global Screening Array (GSA) MultiEthnic Disease Version 1.0. This additional assessment selected 38,030 Lifelines participants with the criteria that they were of European ancestry and an adequate DNA sample for the participant was available. The UGLI participants' data was genotyped in 31 different batches over a period of 5 months in two laboratories. The quality control steps carried out on the entire UGLI sample are described in the UGLI 1.1 Quality Control report and summarised in the Supplementary Table S2.

Following these QC procedures on the total (adult and child) UGLI dataset, 36,339 samples and 571,420 markers remained on autosomal and X chromosomes. Imputation was carried out using the Sanger imputation service and the HRC panel⁵. These pre-processing steps were curated by Serena Sanna, Patrick Deleen, Raul Aguirre-Gamboa, Gerben van der Vries, Ilja Nolte and Esteban Lopera-Maya.

Further post-imputation QC in Lifelines and the GWAS were conducted by Anja Hollowell and Anna Gui. SNPs were filtered based on MAF ($< 0.5\%$), HWE exact test ($p < 1 \times 10^{-6}$) and had call-rate ($< 98\%$). Individuals were further excluded if they had genotyping rate $< 95\%$ and if they did not provide phenotype information about early motor and temperament traits, leaving a sample of 3,483 individuals.

X-chromosome data were available only for the Lifelines dataset and was used for sex checks. Post-imputation QC removal of variants and individuals was conducted on X-chromosome data using the same criteria as described above. The pseudo-autosomal region of the X chromosome was removed during QC. X chromosome data were used to identify any individuals for whom there was a discrepancy between phenotypic and genetic sex ($N < 10$), who were then excluded from analyses.

To ensure each sample was ancestrally homogeneous and aligned with a European reference panel, PCs were calculated, and pairwise plots of PCs along with 1000 Genomes¹⁵ phase 3 reference panel data were visually inspected to identify any individuals to be excluded from analyses. Less than ten individuals for whom either PC1 or PC2 fell outside of 4 standard deviations (SDs) of the centroid of the 1000 Genomes mean were removed.

Finally, IBD was calculated using a pruned set of SNPs and all samples with PI-HAT >0.2 were examined using a relatedness plot¹⁶. For each pair of samples deemed to be genetically identical ($P(\text{IBD}=0) < 0.1$ & $P(\text{IBD}=1) < 0.1$, PI-HAT > 0.8), phenotype data were inspected. For those genetically identical pairs that also had identical phenotype across three measures (age at first steps and two early temperament measures), one sample of the pair was excluded at random as a plausible duplicate. As a result of this process, 10 samples were removed.

From the UGLI dataset, 3,415 individuals (N males = 1,647, N females = 1,768) with phenotype information on age at onset of walking and covariates (sex, child's age at parental report, batch and the first 10 PCs) were included in the primary and sex stratified GWAS analyses for the present study. The Supplementary Fig. S3 shows the QQ-plot from the primary Lifelines GWAS of age at onset of walking.

Ethics

Participants in Lifelines gave written consent prior to physical examination. The study is conducted according to the principles of the Declaration of Helsinki and in accordance with the UMCG research code and is approved by the medical ethical committee of UMCG (document number METC UMCG METc 2007/152).

4) MRC National Study for Health and Development (NSHD)

The NSHD^{17,18} is a population maternity cohort where all women who gave birth in a single week in March 1946 in England, Wales or Scotland were targeted for enrolment. From these 16,695 births, 13,687 were participants in the first stage of the study. Later stages of the study only included subsets of this initial sample.

Phenotype

The assessment when the children in the cohort were aged 1 to 4 years between 1947 and 1950 included 5,362 participants who were selected with the criteria that the births were single births, and the mothers were married. Data was collected by a health visitor in the

mother's home at age 2 years (N = 4,689). One of the items from the questionnaire administered at this time-point was "how many months old was baby when he walked several steps without support".

A follow up assessment when the birth cohort were aged 53 years, where 2,592 participants (N males = 1,297, N females = 1,295) with phenotype information on the date at age at the onset of walking were included in the GWAS analyses. Supplementary Table S1 provides the descriptives for age at onset of walking phenotype for the entire NSHD sample and by sex.

Genotype and QC

Blood samples were collected. Illumina DrugDev Consortium Array was used to genotype the biological samples and QC analyses were performed using PLINK¹⁹ 1.9. Sample QC removed individuals with call rates <95%, extreme heterozygosity ($\mu \pm 3$ SDs), sex mismatches, relatedness and duplicates (PI-HAT>0.2), and PC outliers. All participants were of European ancestry. Genotyped SNPs were excluded on the basis of the following parameters: call rate < 95%, MAF < 0.01 or HWE p-value < 1×10^{-4} . The genotyped SNPs were used to impute information on missing common variants using the HRC⁵ v1.1 reference panel, accessed via the Michigan Imputation Server. QC of imputed data led to SNPs being excluded with INFO score < 0.3, MAF < 0.5%, call rate > 98%, HWE p-value < 1×10^{-6} .

Pre- and post-imputation QC was carried out by Valerie Kuan.

The primary and sex stratified GWAS analyses were conducted by Andrew Wong. Sex and the first 10 ancestry PCs were added as covariates in the primary GWAS model (see Supplementary Fig. S4 for the QQ-plot). Only the first 10 PCs were used as covariates in the sex stratified GWAS model.

Ethics

The collection of blood samples and DNA information from the participants was approved by ethical approval reference MREC no. 98/2/121.

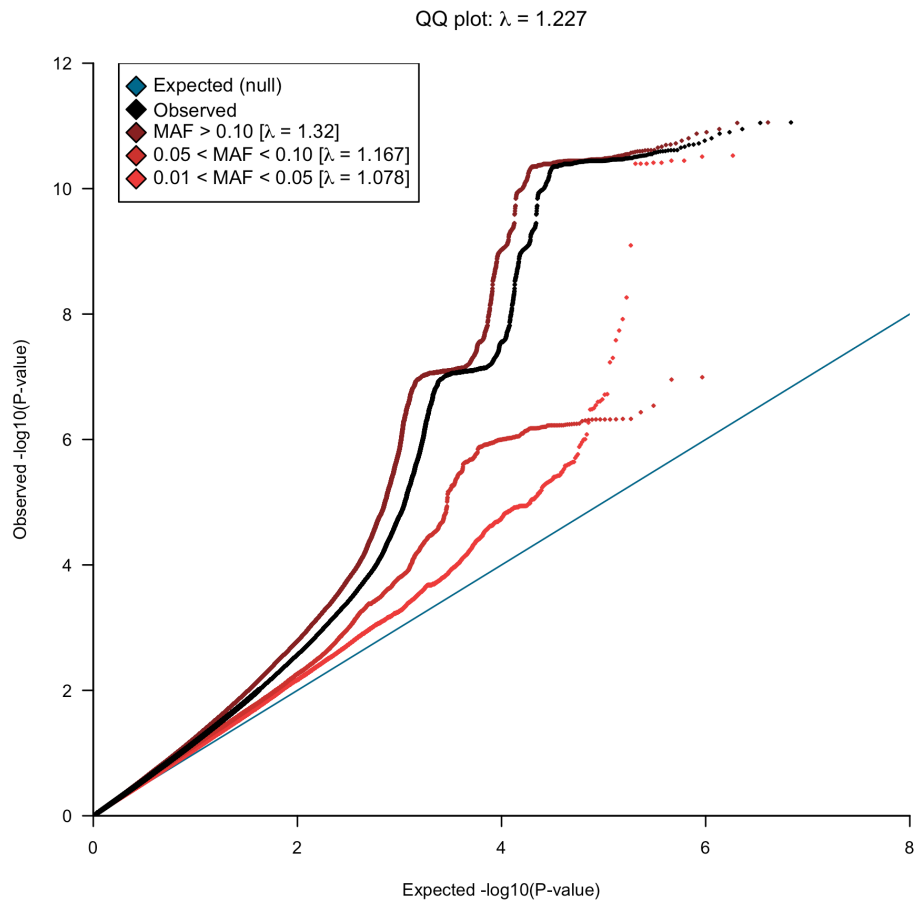
B. Study Ethics

This study and the related secondary data analysis were approved by the Departmental Ethics Committee of the Psychological Science Department of Birkbeck, University of London on 27th October 2020 (reference number 2021007).

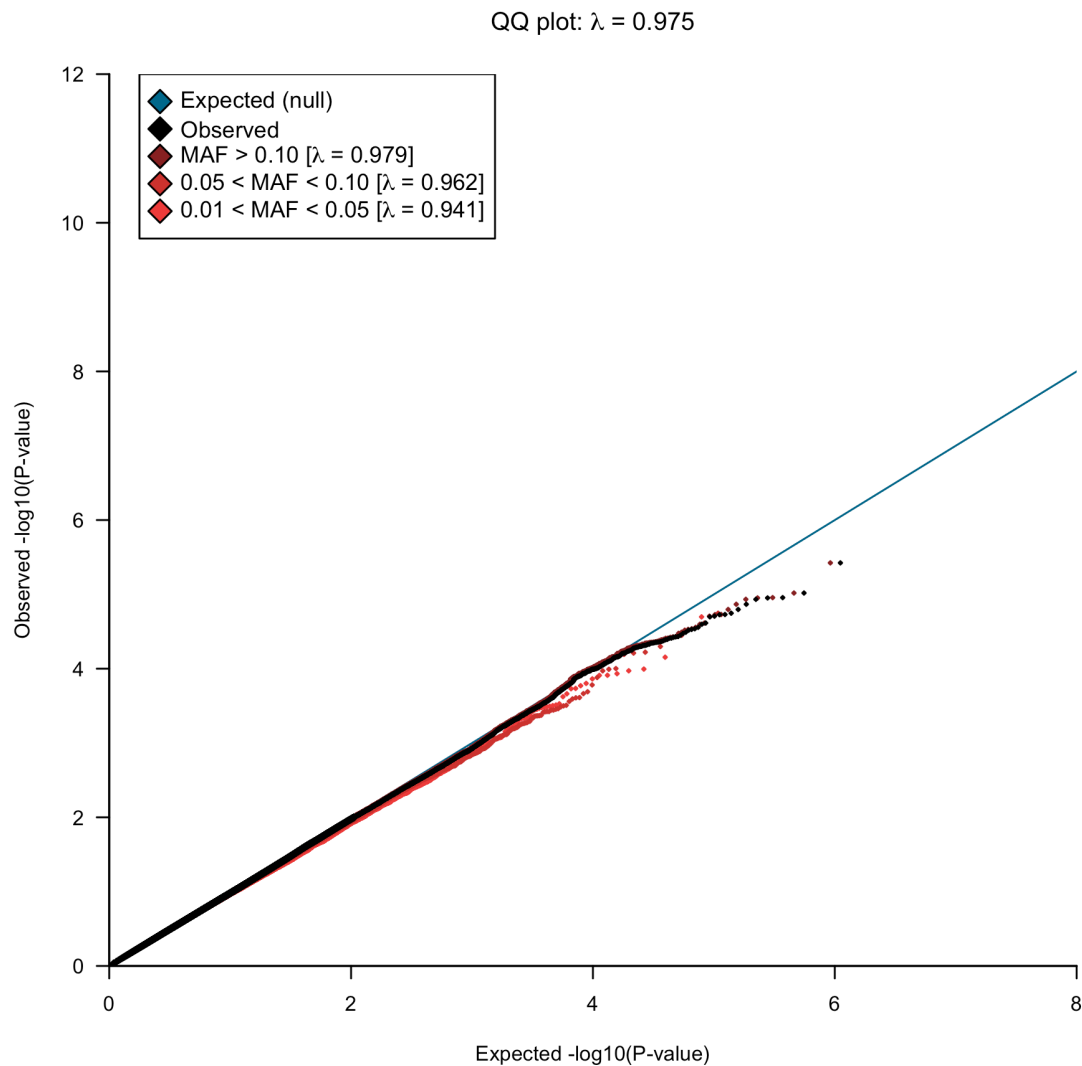
C. Genetic correlation between cohorts

The SNP-based heritability (h^2_{SNP}) estimated using LDSC⁶ for the MoBa sample was $h^2_{\text{SNP}} = 25.11\%$ (SE = 1.34%) and for the NTR sample $h^2_{\text{SNP}} = 19.09\%$ (SE = 7.42%). Lower h^2_{SNP} estimates and larger standard errors were obtained for the smaller samples, namely Lifelines ($h^2_{\text{SNP}} = 9.52\%$, SE = 12.98%) and NSHD ($h^2_{\text{SNP}} = -3.02\%$, SE = 17.17%). Genetic correlation (r_g) between MoBa and NTR was $r_g = 0.89$ (SE = 0.17, $p = 1.803 \times 10^{-7}$) and between NTR and Lifelines was $r_g = 0.46$ (SE = 0.55, $p = 0.404$). Other genetic correlations were out of bound (MoBa-Lifelines $r_g = 1.168$, SE = 0.72, $p = 0.103$) or non-estimable (MoBa-NHSD r_g , Lifelines-NHSD r_g and NTR-NHSD $r_g = \text{nan}$) due to low power, as indicated by the large SNP-based heritability standard errors obtained for the smaller cohorts.

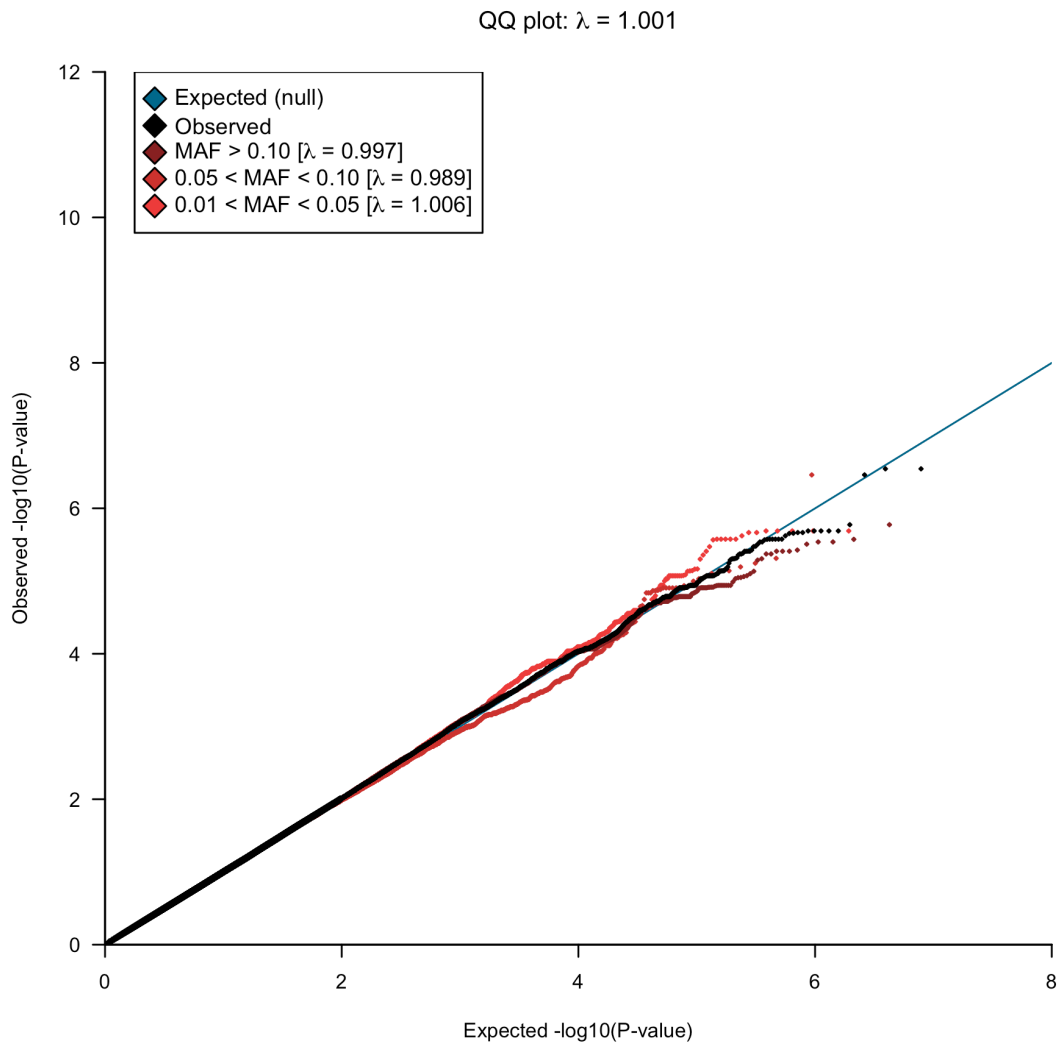
Supplementary Figures



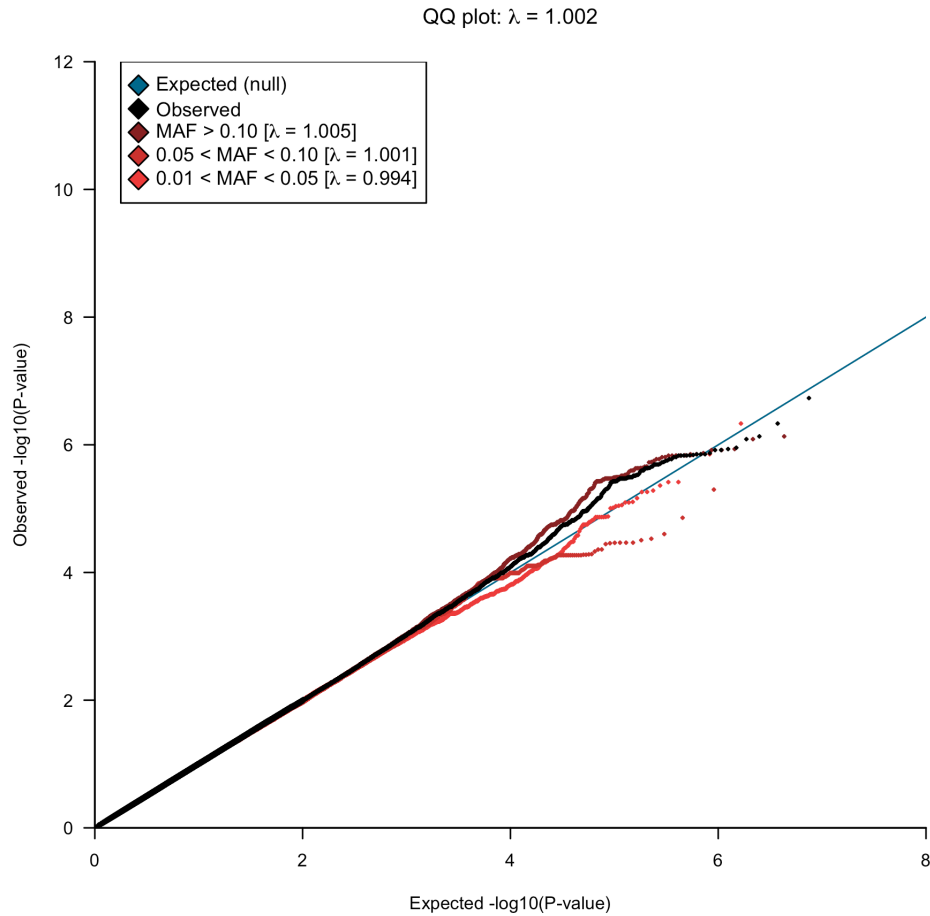
Supplementary Fig. S1. Quantile-quantile plot of the observed $-\log_{10}[\text{p-values}]$ from MoBa GWAS for age at onset of walking against the expected $-\log_{10}[\text{two-sided p-value}]$ (blue line). Results are presented by allele frequency, brown dots representing signals for the SNPs with minor allele frequency (MAF) > 10%, red for SNPs with MAF between 5 and 10%, orange circles for SNPs with MAF between 1 and 5% and black circles for all SNPs included in the GWAS (MAF > 1%).



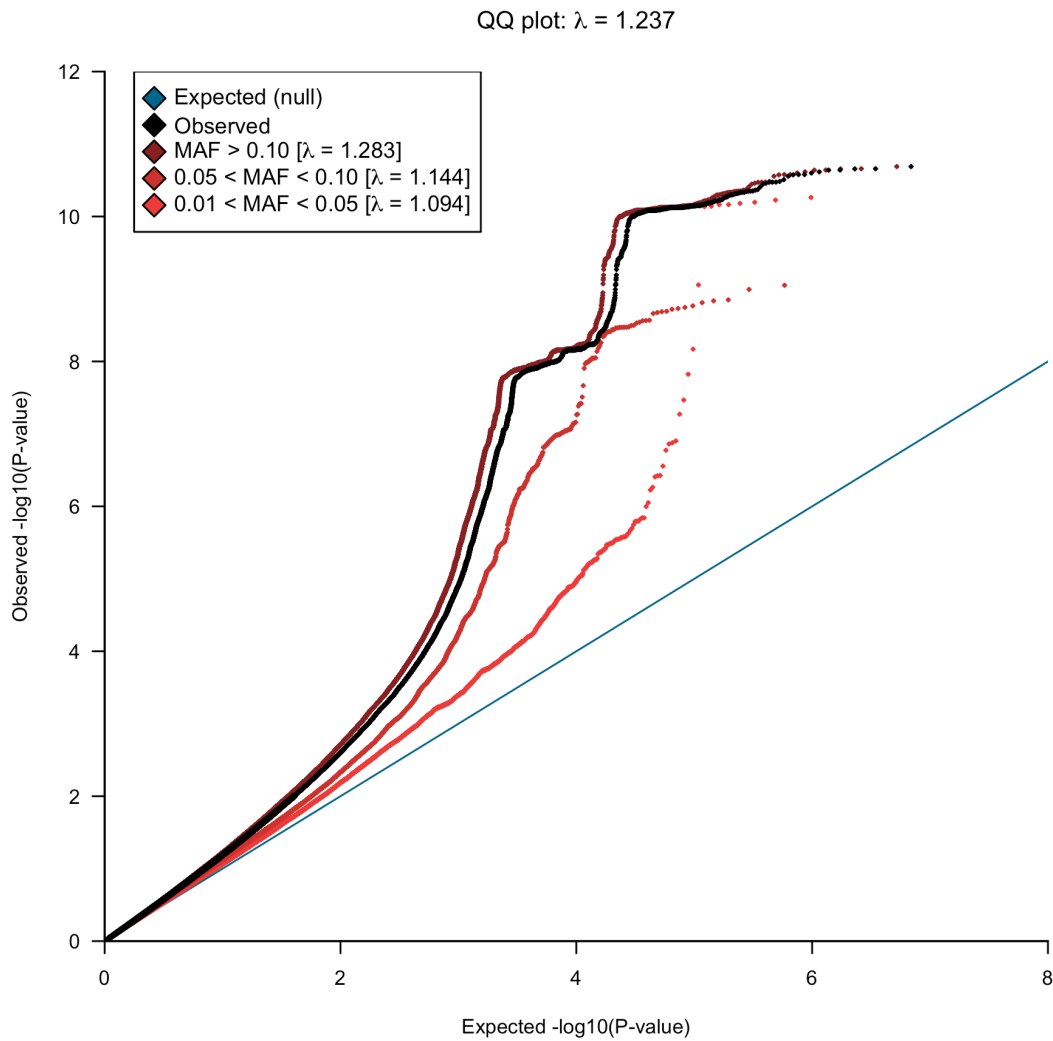
Supplementary Fig. S2. Quantile-quantile plot of the observed $-\log_{10}[\text{p-values}]$ from Netherlands Twin Register (NTR) GWAS for age at onset of walking against the expected $-\log_{10}[\text{two-sided p-value}]$ (blue line). Results are presented by allele frequency, brown dots representing signals for the SNPs with minor allele frequency (MAF) > 10%, red for SNPs with MAF between 5 and 10%, orange circles for SNPs with MAF between 1 and 5% and black circles for all SNPs included in the GWAS (MAF > 1%).



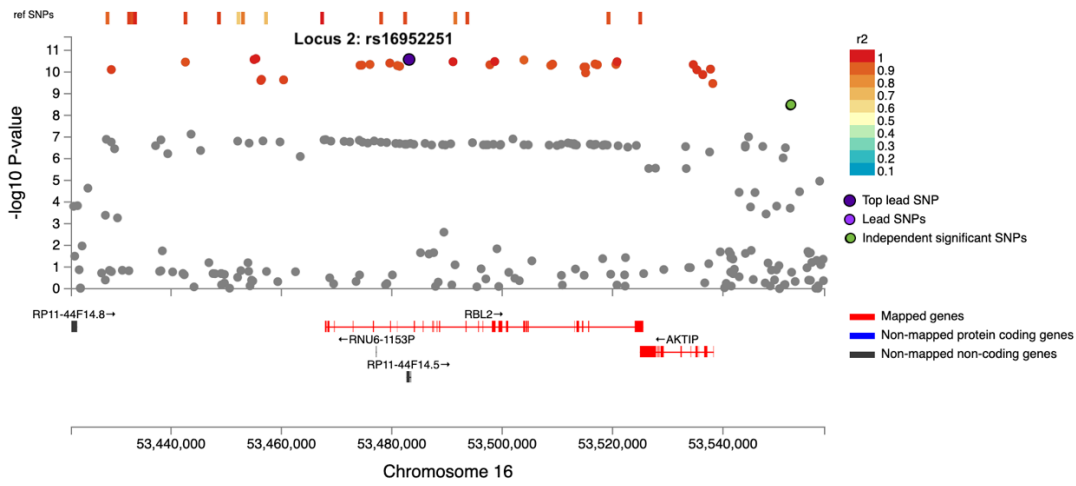
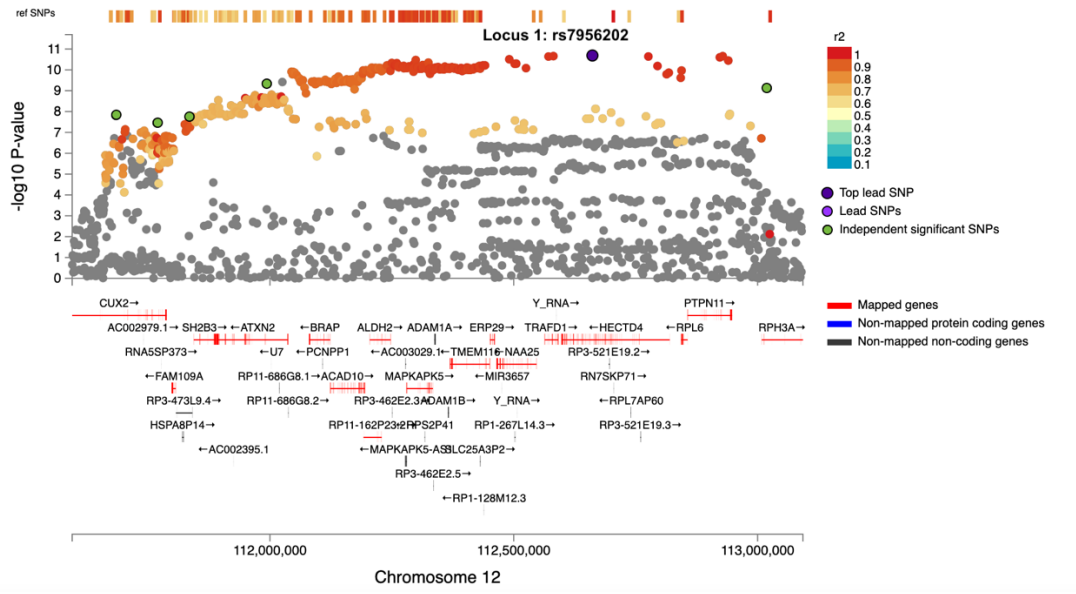
Supplementary Fig. S3. Quantile-quantile plot of the observed $-\log_{10}[\text{p-values}]$ from Lifelines GWAS for age at onset of walking against the expected $-\log_{10}[\text{two-sided p-value}]$ (blue line). Results are presented by allele frequency, brown dots representing signals for the SNPs with minor allele frequency (MAF) > 10%, red for SNPs with MAF between 5 and 10%, orange circles for SNPs with MAF between 1 and 5% and black circles for all SNPs included in the GWAS (MAF > 1%).

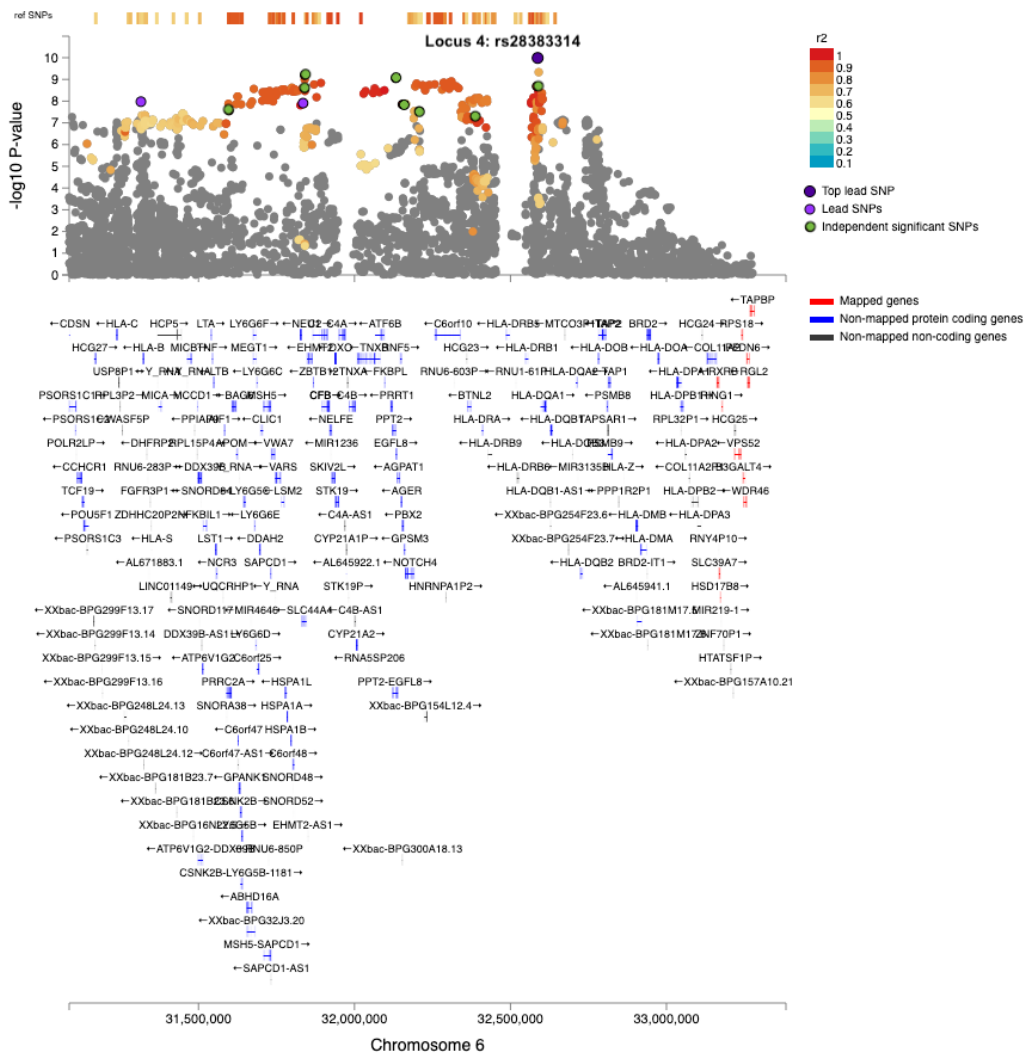
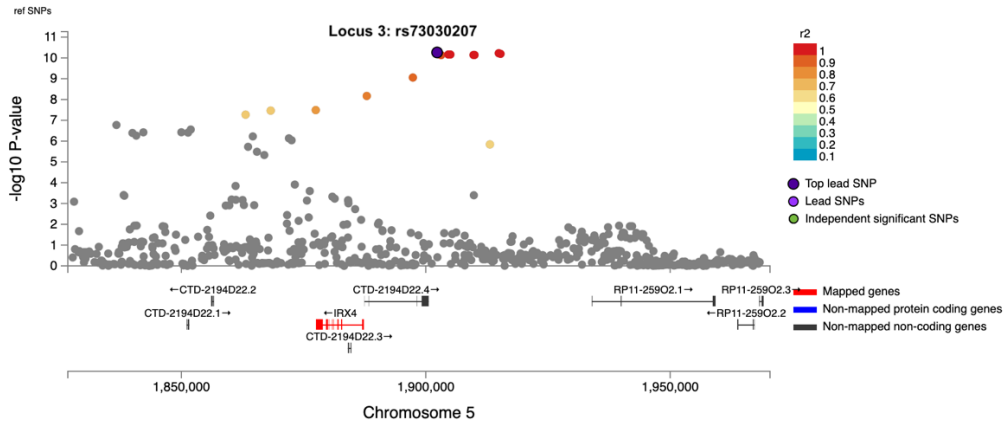


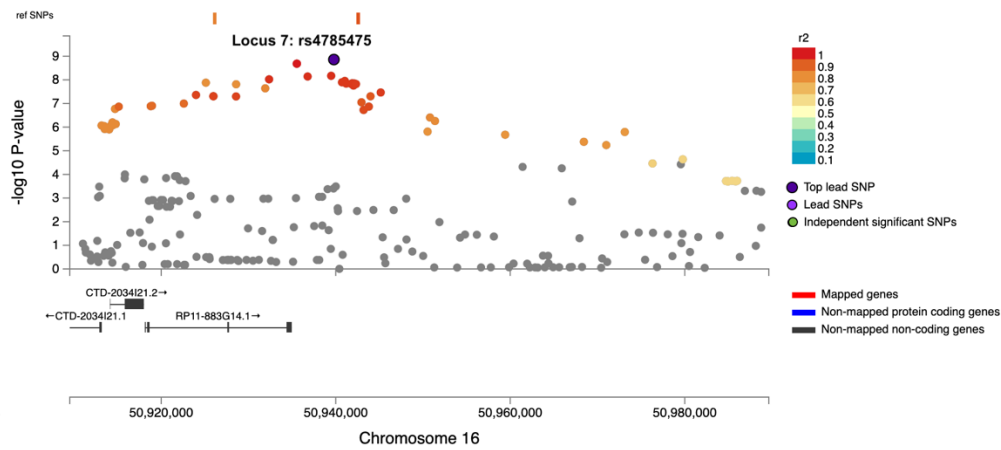
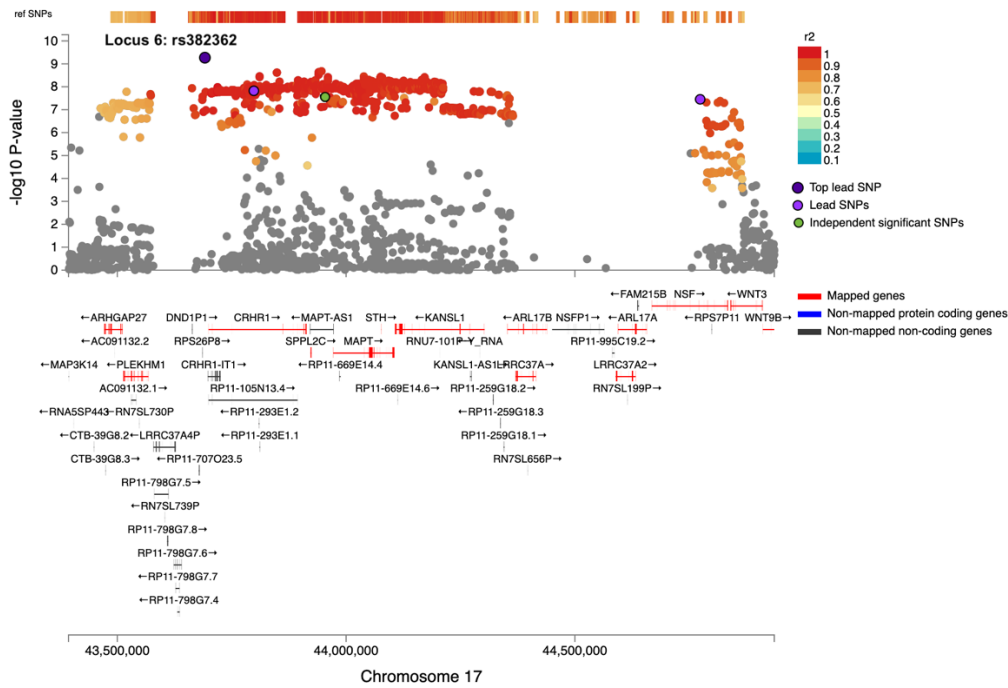
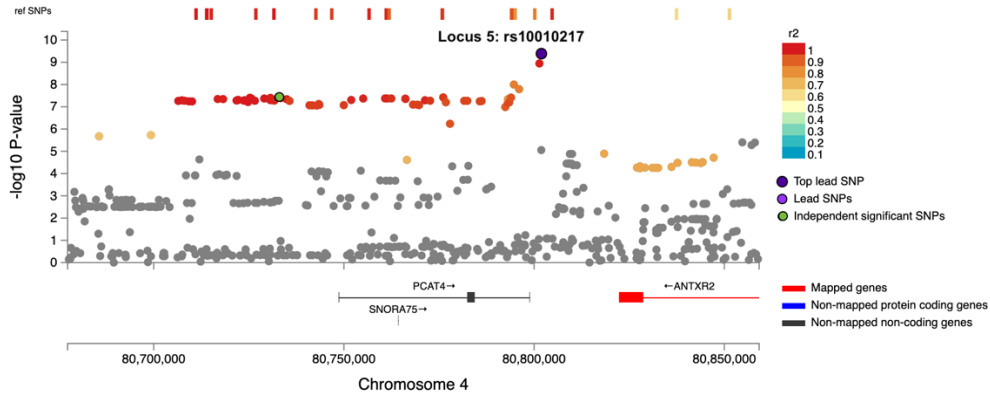
Supplementary Fig. S4. Quantile-quantile plot of the observed $-\log_{10}[\text{p-values}]$ from MRC National Study of Health and Development (NSHD) GWAS for age at onset of walking against the expected $-\log_{10}[\text{two-sided p-value}]$ (blue line). Results are presented by allele frequency, brown dots representing signals for the SNPs with minor allele frequency (MAF) $> 10\%$, red for SNPs with MAF between 5 and 10%, orange circles for SNPs with MAF between 1 and 5% and black circles for all SNPs included in the GWAS (MAF $> 1\%$).

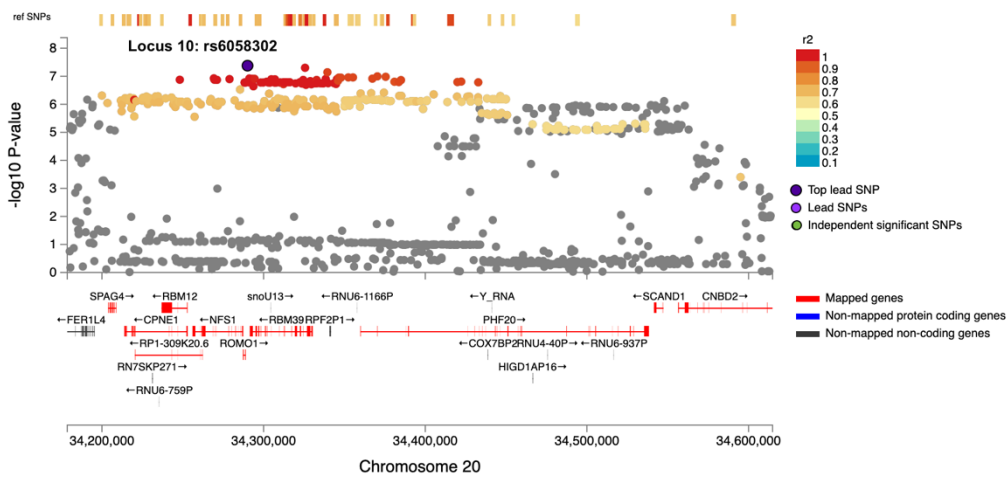
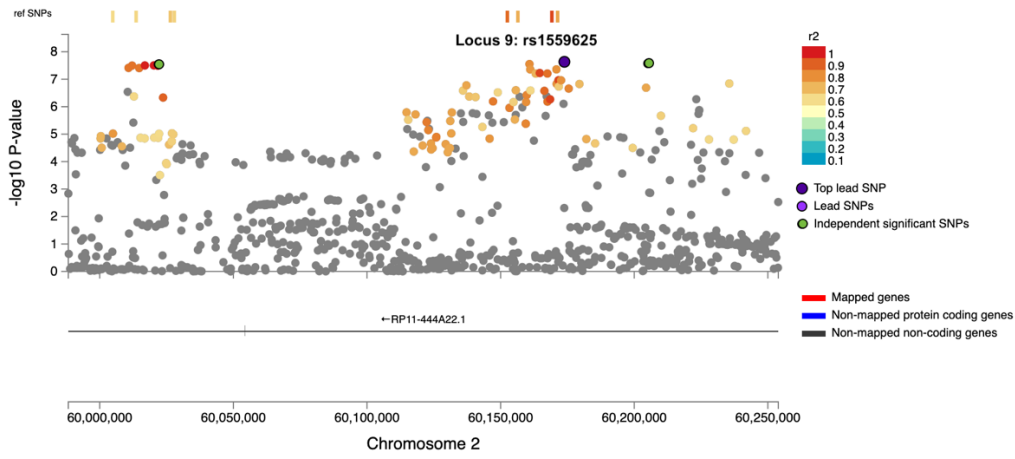
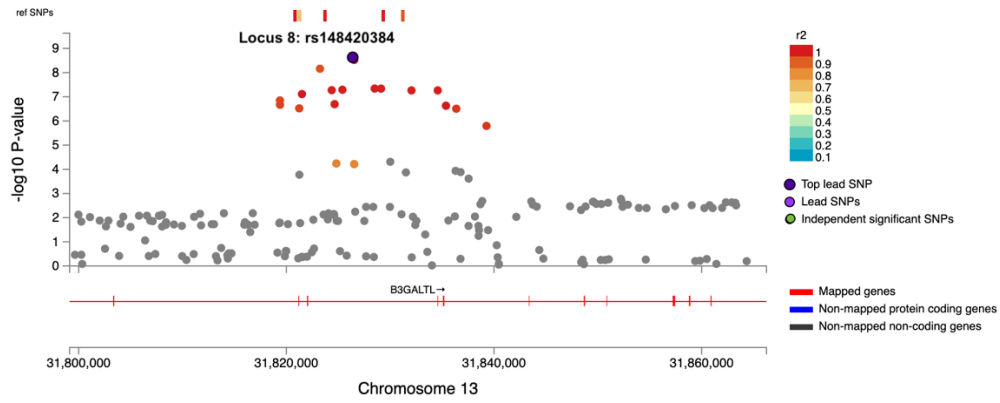


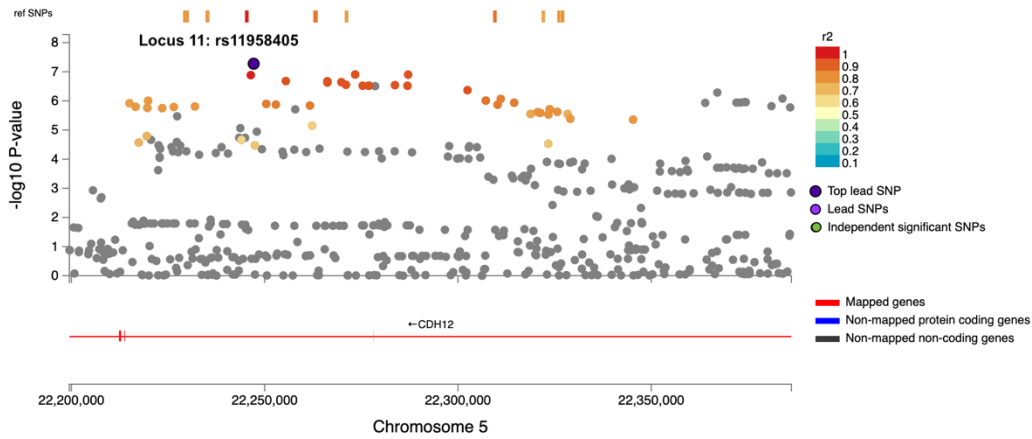
Supplementary Fig. S5. Quantile-quantile plot of the observed $-\log_{10}[\text{p-values}]$ from GWAS meta-analysis for age at onset of walking ($N=70,560$) against the expected $-\log_{10}[\text{two-sided p-value}]$ (blue line). Results are presented by allele frequency, brown dots representing signals for the SNPs with minor allele frequency (MAF) > 10%, red for SNPs with MAF between 5 and 10%, orange circles for SNPs with MAF between 1 and 5% and black circles for all SNPs included in the GWAS (MAF > 1%).



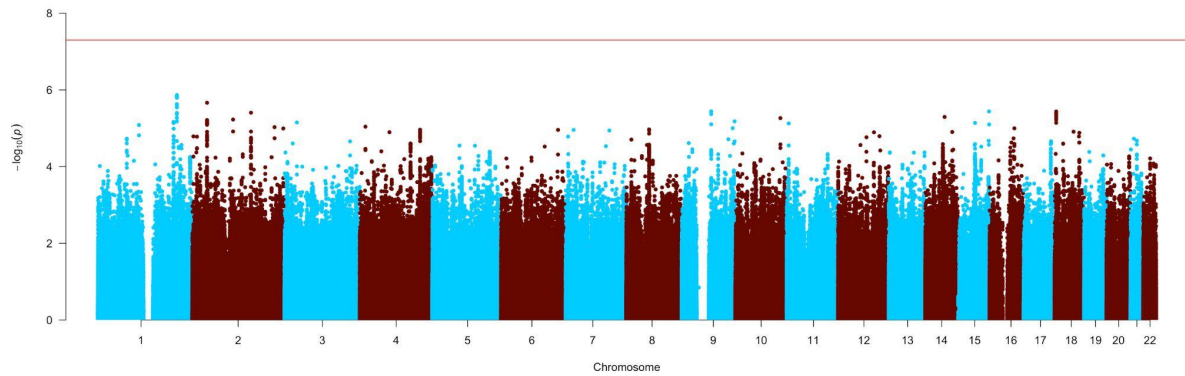




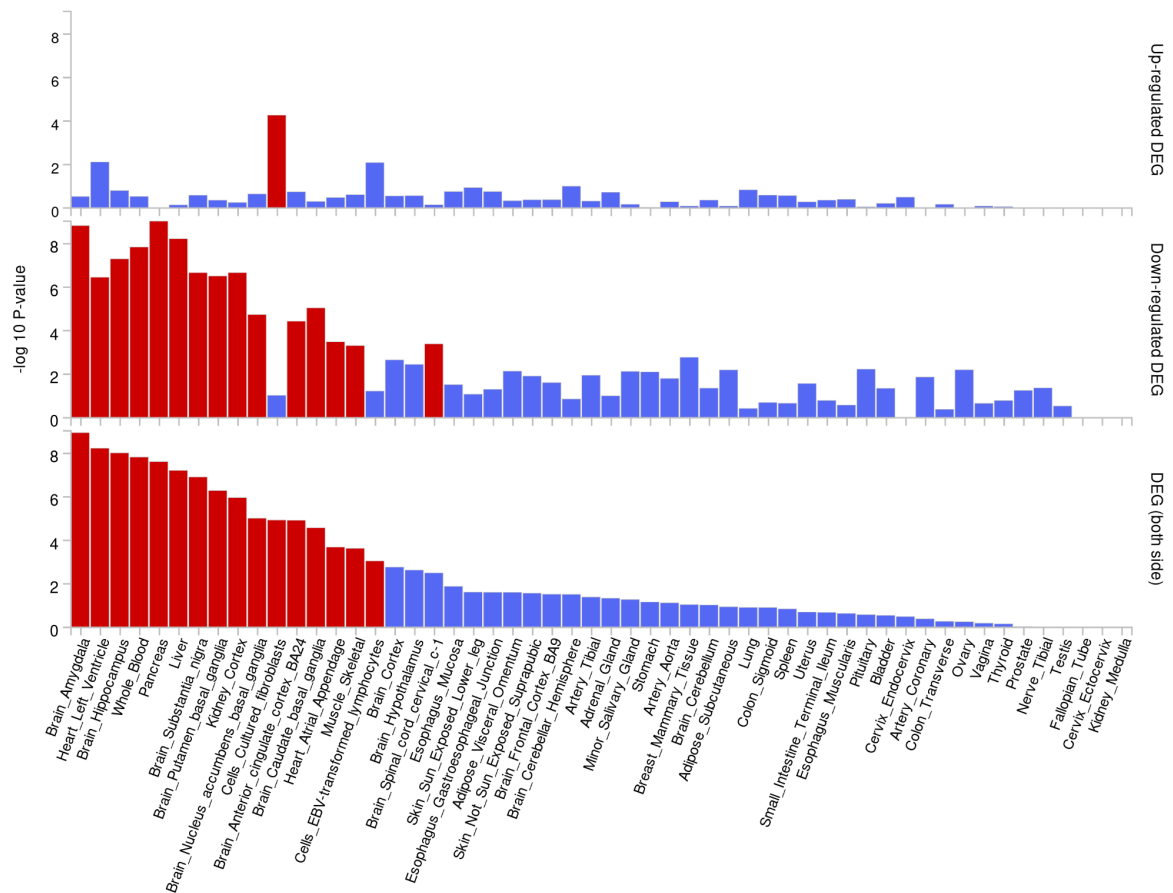




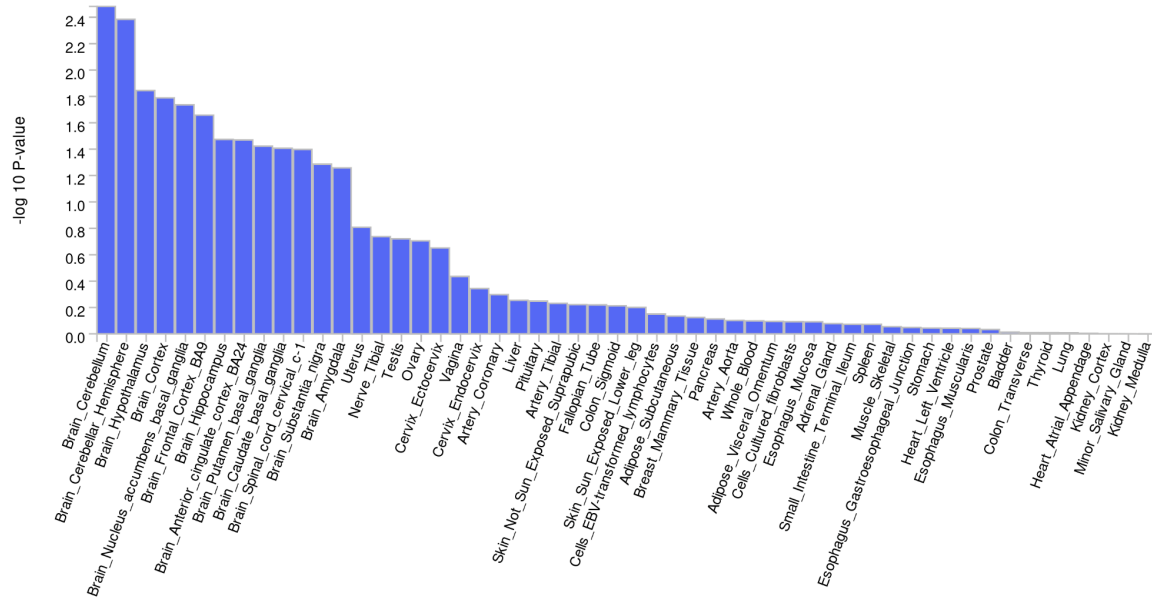
Supplementary Fig. S6. Regional plots for the top SNPs in the eleven genomic loci associated with age at onset of walking, realised with FUMA²⁷. Association p-values on the $-\log_{10}$ scale are shown on the y axis, and the chromosomal position is indicated along the x axis. The top lead SNP within the locus is indexed by a blue dot and annotated with the genomic locus number and SNP ID. Other independent (LD $r^2 < 0.6$) significant SNPs are indexed in green. SNPs in LD with the lead variant are color-coded based on LD r^2 estimated on the 1000 Genomes¹⁵ European population. SNPs that are not in LD with any of the significant independent lead SNPs in the region are colored grey. At the bottom, genes mapped by positional mapping are presented in red.



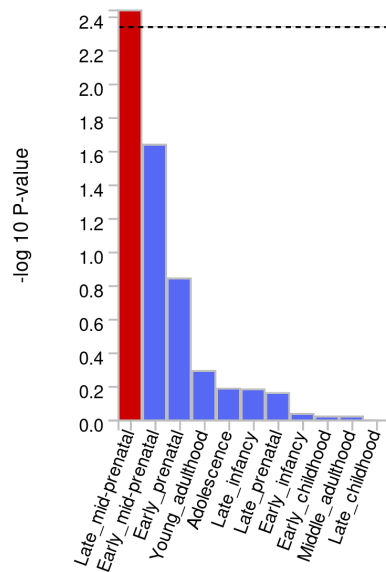
Supplementary Fig. S7. Heterogeneity Manhattan plot. The y-axis represents the $-\log_{10}$ [two-sided p-value] of an omnibus test for heterogeneity across cohorts (I^2 statistics) for the age at onset of walking meta-GWAS. The x-axis represents genomic position (chromosomes 1-22). The horizontal red line indicates the p-value threshold for genome-wide statistical significance ($p = 5 \times 10^{-8}$).



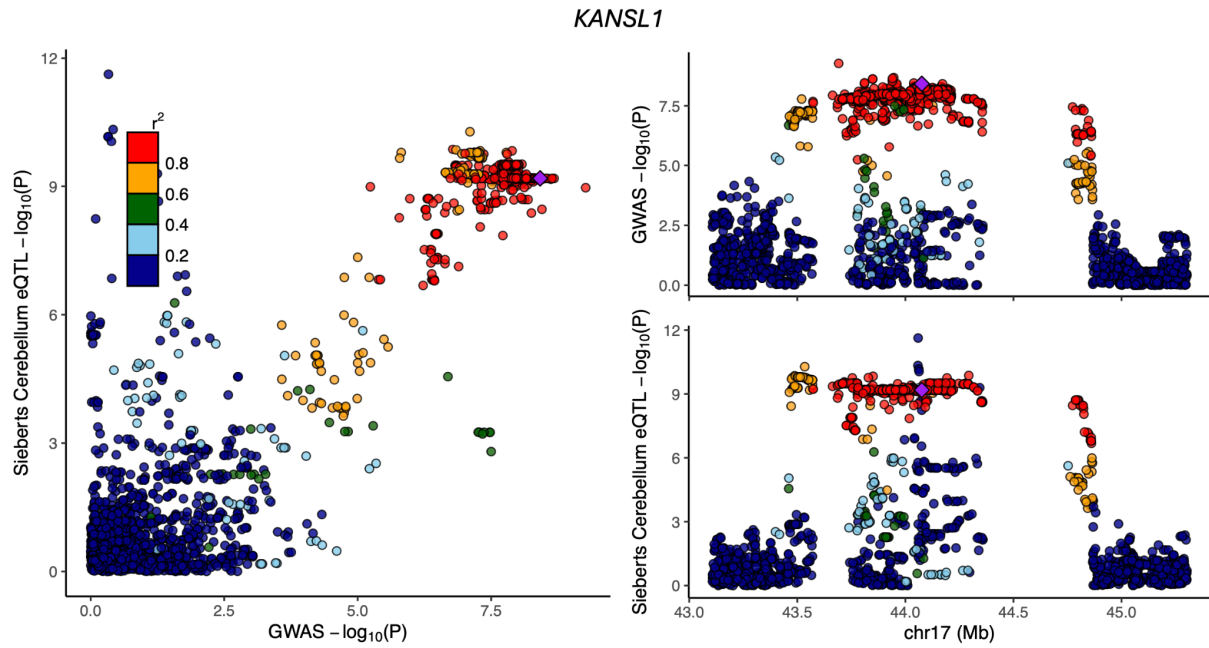
Supplementary Fig. S8. Enrichment of up-regulated and down-regulated differentially expressed genes (DEGs) performed for 54 GTEx²⁹ v8 tissue datasets available in FUMA²⁷ (on the x axis). The y axis represents the $-\log_{10}[\text{one-sided p-value}]$ of the probability hypergeometric test performed in FUMA (GENE2FUNC). Red bars indicate the tissue for which the DEGs are enriched at a Bonferroni corrected p-value < 0.05 .



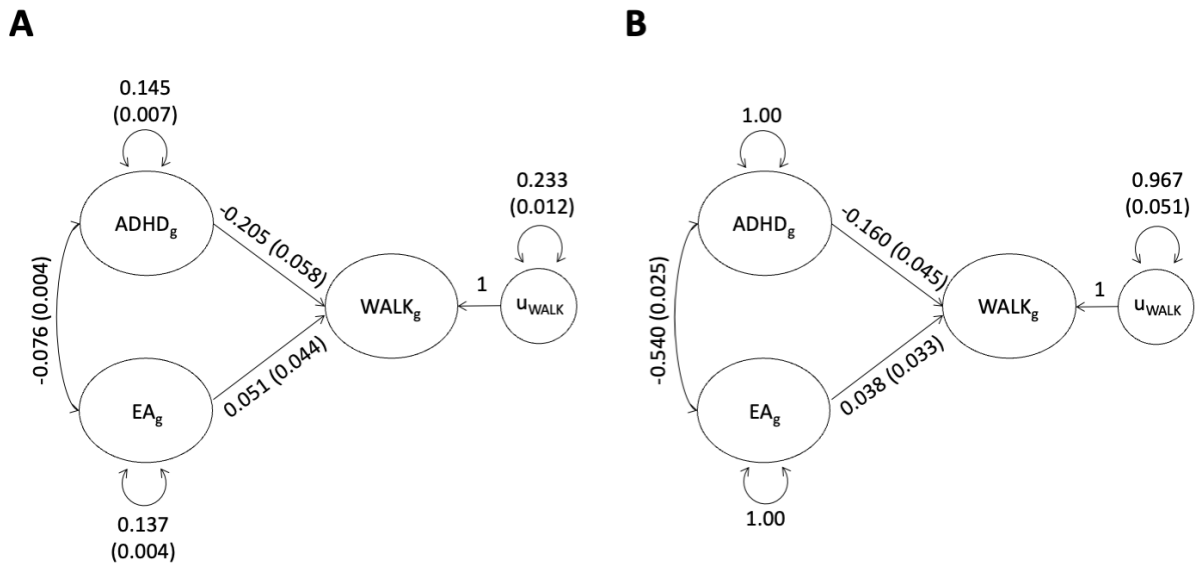
Supplementary Fig. S9. Results of the MAGMA³⁰ gene property analysis performed for 53 GTEx²⁹ v8 tissue types (on the x axis). The y axis represents the $-\log_{10}$ [one-sided p-value] testing the positive relationship between tissue specificity and genetic association of genes with age at onset of walking.



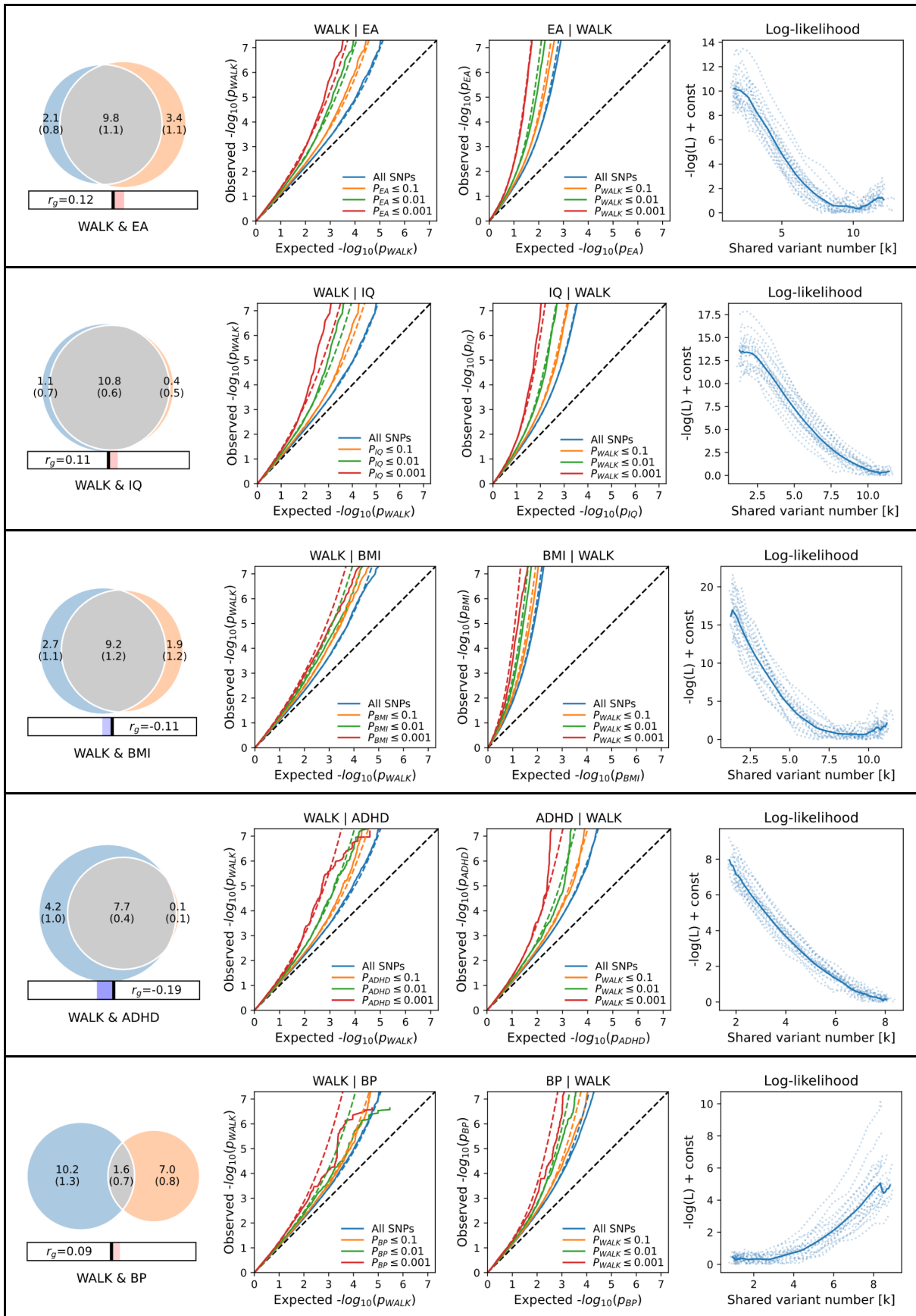
Supplementary Fig. S10. Results of the MAGMA³⁰ gene property analysis performed for the BrainSpan developmental stages dataset²⁸ available in FUMA²⁷ (on the x axis). The y axis represents the $-\log_{10}[\text{one-sided p-value}]$ testing the positive relationship between expression in the BrainSpan tissues and genetic association of genes with age at onset of walking. The dashed line represents the Bonferroni-corrected α level, and the red bars indicate that the beta-value of the regression between gene expression in the BrainSpan tissues and genetic association with age at onset of walking is significantly higher than 0.

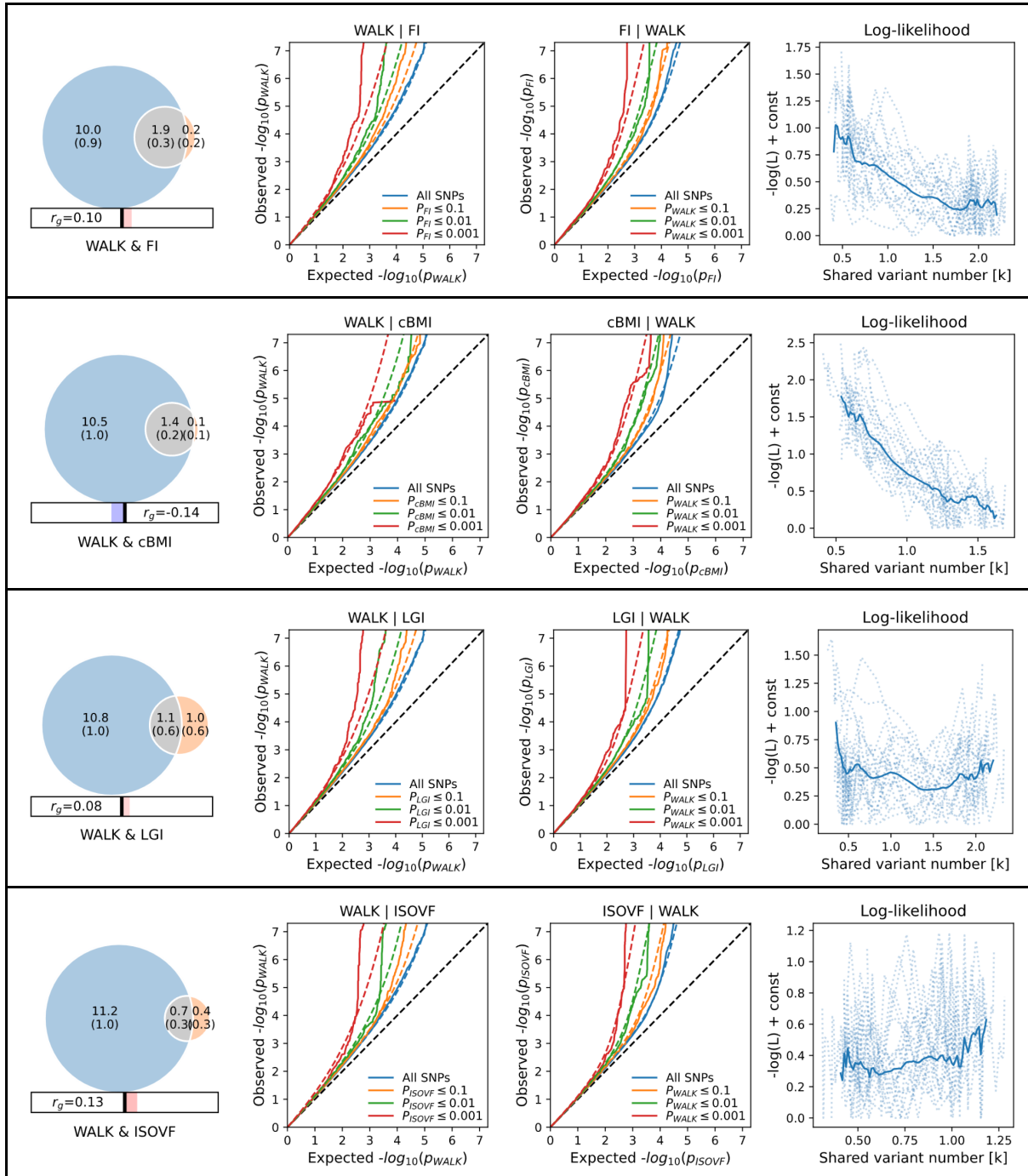


Supplementary Fig. S11. GWAS evidence for association with age at onset of walking ($-\log_{10}[\text{p-value}]$, x -axis) plotted against the statistical evidence of being an expression Quantitative Trait Locus (eQTL) for gene *KANSL1* within the genomic locus 6 in human adult cerebellum³¹ ($-\log_{10}[\text{p-value}]$), y -axis) for each SNP (points) within a 1Mb window around the gene. Points are colored by LD correlation with the lead SNP (purple diamonds), estimated in individuals with European ancestry in 1000 Genomes^{15,32}.



Supplementary Fig. S12. Unstandardized (A) and standardized (B) results from Genomic SEM³³ indicating the genetic covariance and genetic correlation estimates, respectively, and associated sampling covariance matrices, from a multivariate LD score regression of the age at learning to walk (WALK) on ADHD⁸ and educational attainment (EA)³⁴.





Supplementary Fig. S13. MiXeR⁷ bivariate model figures including negative log-likelihood plots. WALK = age at onset of walking; IQ = Intelligence; BMI = Body Mass Index; ADHD = Attention Deficit / Hyperactivity Disorder; BP = Bipolar Disorder; FI = Folding Index; cBMI = childhood BMI; LGI = Local Gyrfication Index; ISOVF = Isotropic Volume Fractionation.

Supplementary References

1. Magnus, P. *et al.* Cohort profile: The Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol* **35**, 1146–1150 (2006).
2. Magnus, P. *et al.* Cohort Profile Update: The Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol* **45**, 382–388 (2016).
3. Corfield, E. C. *et al.* The Norwegian Mother, Father, and Child cohort study (MoBa) genotyping data resource: MoBaPsychGen pipeline v.1. 2022 **bioRxiv**,.
4. Paltiel, L. *et al.* The biobank of the Norwegian Mother and Child Cohort Study – present status. *Norsk Epidemiologi* **24**, (2014).
5. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279–1283 (2016).
6. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291–295 (2015).
7. Frei, O. *et al.* Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nat Commun* **10**, 2417 (2019).
8. Demontis, D. *et al.* Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nat Genet* **55**, 198–208 (2023).
9. Mullins, N. *et al.* Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat Genet* **53**, 817–829 (2021).
10. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
11. Winkler, T. W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* **9**, 1192–1212 (2014).
12. Stolk, R. P. *et al.* Universal risk factors for multifactorial diseases. *Eur J Epidemiol* **23**, 67–74 (2008).
13. Scholtens, S. *et al.* Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol* **44**, 1172–1180 (2015).
14. Sijtsma, A. *et al.* Cohort Profile Update: Lifelines, a three-generation cohort study and biobank. *Int J Epidemiol* **51**, e295–e302 (2022).
15. Campbell, C. L. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
16. Marees, A. T. *et al.* A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int J Methods Psychiatr Res* **27**, 1–10 (2018).
17. Wadsworth, M., Kuh, D., Richards, M. & Hardy, R. Cohort profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development). *Int J Epidemiol* **35**, 49–54 (2006).
18. Rousseau, K. *et al.* MUC7 haplotype analysis: results from a longitudinal birth cohort support protective effect of the MUC7*5 allele on respiratory function. *Ann Hum Genet* **70**, 417–27 (2006).
19. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575 (2007).
20. Boomsma, D. I. *et al.* Netherlands Twin Register: A Focus on Longitudinal Research. *Twin Research* **5**, 401–406 (2002).
21. Willemsen, G. *et al.* The Netherlands twin register biobank: A resource for genetic epidemiological studies. *Twin Research and Human Genetics* **13**, 231–245 (2010).

22. Ligthart, L. *et al.* The Netherlands Twin Register: Longitudinal Research Based on Twin and Twin-Family Designs. *Twin Research and Human Genetics* **22**, 623–636 (2019).
23. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–73 (2010).
24. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet* **2**, e190 (2006).
25. Loh, P.-R., Palamara, P. F. & Price, A. L. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat Genet* **48**, 811–6 (2016).
26. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284–1287 (2016).
27. Watanabe, K., Taskesen, E., Van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1–10 (2017).
28. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–9 (2011).
29. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science (1979)* **369**, 1318–1330 (2020).
30. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput Biol* **11**, 1–19 (2015).
31. Sieberts, S. K. *et al.* Large eQTL meta-analysis reveals differing patterns between cerebral cortical and cerebellar brain regions. *Sci Data* **7**, 340 (2020).
32. Liu, B., Gloudemans, M. J., Rao, A. S., Ingelsson, E. & Montgomery, S. B. Abundant associations with gene expression complicate GWAS follow-up. *Nat Genet* **51**, 768–769 (2019).
33. Grotzinger, A. D. *et al.* Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat Hum Behav* **3**, 513–525 (2019).
34. Okbay, A. *et al.* Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals. *Nat Genet* **54**, 437–449 (2022).