Linking obesity-associated genotype to child language development: the role of early-life neurology-related proteomics and brain myelination

Jian Huang,[a,b,c,d,*] Jinyi Che,[a,b,e] Michelle Z. L. Kee,[a] Ai Peng Tan,[a,f,g] Evelyn C. Law,[a,e,h] Patricia Pelufo Silveira,[e,i] Irina Pokhvisneva,[i] Sachin Patel,[i] Keith M. Godfrey,[j] Lourdes Mary Daniel,[k] Kok Hian Tan,[l] Yap Seng Chong,[a,m] Shiao-Yng Chan,[a,c,m] Johan G. Eriksson,[a,c,m,n,o] Dennis Wang,[a,b,c,e,p,q] and Jonathan Yinhao Huang[r,s]

[a]Institute for Human Development and Potential (IHDP), Agency for Science, Technology and Research (A*STAR), Singapore, Republic of Singapore
[b]Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Singapore, Republic of Singapore
[c]Human Potential Translational Research Program, Yong Loo Lin School of Medicine, National University of Singapore, Republic of Singapore
[d]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, Norfolk Place, London, UK
[e]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Republic of Singapore
[f]Department of Diagnostic Imaging, National University Hospital, Singapore, Republic of Singapore
[g]Department of Diagnostic Radiology, Yong Loo Lin School of Medicine, NUS, Singapore, Republic of Singapore
[h]Department of Paediatrics, Khoo Teck Puat-National University Children's Medical Institute, National University Hospital, Singapore, Republic of Singapore
[i]Department of Psychiatry, Faculty of Medicine and Ludmer Centre for Neuroinformatics and Mental Health, Douglas Hospital Research Centre, McGill University, Montreal, Canada
[j]MRC Lifecourse Epidemiology Centre and NIHR Southampton Biomedical Research Centre, University of Southampton & University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom
[k]Department of Child Development, KK Women's and Children's Hospital, Singapore, Republic of Singapore
[l]Department of Maternal Fetal Medicine, KK Women's and Children's Hospital, Singapore, Republic of Singapore
[m]Department of Obstetrics & Gynaecology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Republic of Singapore
[n]Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland
[o]Folkhälsan Research Center, Helsinki, Finland
[p]National Heart and Lung Institute, Imperial College London, London, UK
[q]Department of Computer Science, University of Sheffield, Sheffield, UK
[r]Centre for Quantitative Medicine, Duke-NUS Medical School, Singapore, Republic of Singapore
[s]Thompson School of Social Work & Public Health, Office of Public Health Studies, University of Hawai'i at Mānoa, Honolulu, Hawaii, USA

eBioMedicine
2025;113: 105579
Published Online xxx
https://doi.org/10.1016/j.ebiom.2025.105579

Summary

Background The association between childhood obesity and language development may be confounded by socio-environmental factors and attributed to comorbid pathways.

Methods In a longitudinal Singaporean mother-offspring cohort, we leveraged trans-ancestry polygenic predictions of body mass index (BMI) to interrogate the causal effects of early-life BMI on child language development and its effects on molecular and neuroimaging measures. Leveraging large genome-wide association studies, we examined whether the link between obesity and language development is causal or due to a shared genetic basis.

Findings We found an inverse association between polygenic risk for obesity, which is less susceptible to confounding, and language ability assessed at age 9. Our findings suggested a shared genetic basis between obesity and language development rather than a causal effect of obesity on language development. Interrogating early-life mechanisms including neurology-related proteomics and language-related white matter microstructure, we found that EFNA4 and VWC2 expressions were associated with language ability as well as fractional anisotropy of language-related white matter tracts, suggesting a role in brain myelination. Additionally, the expression of the EPH-Ephrin signalling pathway in the hippocampus might contribute to language development. Polygenic risk for

*Corresponding author. Institute for Human Development and Potential (IHDP), Brenner Centre for Molecular Medicine, 30 Medical Drive, 117609, Republic of Singapore.
E-mail address: huang_jian@sics.a-star.edu.sg (J. Huang).

obesity was nominally associated with EFNA4 and VWC2 expression. However, we did not find support for mediating mechanisms via these proteins.

Interpretation This study demonstrates the potential of examining early-life proteomics in conjunction with deep genotyping and phenotyping and provides biological insights into the shared genomic links between obesity and language development.

Funding Singapore National Research Foundation and Agency for Science, Technology and Research.

Keywords: Polygenic risk score; Obesity; Language development; Neurology-related protein

Research in context

Evidence before this study

Language development is a cornerstone of cognitive growth. The development of childhood weight and obesity is plausibly linked to suboptimal neurodevelopment but is highly confounded by socioenvironmental factors. However, identifying the extent and mechanisms by which early weight status affects language development may lead to promising early targets for mitigating disparities in lifelong health and well-being.

Added value of this study

We leveraged trans-ancestry polygenic predictions of BMI to interrogate the causal effects of early-life BMI on child language development and its effects on molecular and neuroimaging measures. We did not find evidence for a direct causal influence of genetically predicted obesity-related traits on language-related skills using Mendelian randomisation. Instead, we found evidence of a shared genetic basis between obesity-related traits and language-related skills. We also investigated the role of neurology-related proteomics and found that higher polygenic risk for obesity was nominally and positively associated with EFNA4 and VWC2 expression, both of which were inversely associated with the WIAT-III language-related composite score. However, mediating mechanisms via EFNA4 or VWC2 were not supported.

Implications of all the available evidence

These findings support that genetic variants associated with childhood obesity may be implicated in language development, explaining previous observations. However, while our further investigation uncovered mechanisms of interest for early language development, evidence points to shared genomic pathways rather than a direct causal pathway from obesity development to language development. Although mediating mechanisms via individual proteins were not identified, our investigation suggests a more comprehensive early-life proteomic profiling warrants further investigation. This has the potential to unravel the link between cardiometabolic health and neurodevelopment at the molecular level and to provide insight into how human development lays the foundation for future health trajectories.

Introduction

Language is the primary medium for conveying thoughts and emotions. It is deeply intertwined with the way individuals process and interpret information. This interconnection highlights the critical role of language development as a cornerstone of cognitive growth, setting a foundation for future health and well-being.[1] Previous research has established associations of early-life language skills with executive function and performance in school assessments among primary-aged children.[2,3] Furthermore, lower early-life language ability has been associated with mental disorders and psychosocial disabilities in adulthood,[4] and poorer cognitive function and an elevated risk of Alzheimer's disease in later years.[5,6] These findings highlight the enduring effects of early language abilities. The underlying biological processes that mediate language development, such as brain myelination,[7] can be influenced by a variety of factors, including nutrition, physical activity, and obesity,[8] which supports the observed association between childhood obesity and poorer neurocognitive functioning.[9] In particular, the inhibitory role of obesity in myelination may contribute to its association with cognitive development and educational outcomes.[10,11] Since obesity and language development are both heavily socially patterned,[12,13] early obesity management may be a modifiable pathway to mitigate early language disparities. However, shared socio-environmental factors render the effects of obesity difficult to study. In contrast, recent studies using genetically predicted body mass index (BMI) have shown promise in understanding the interplay between environmentally and genetically influenced obesity and the causal effects of long-term exposure to obesity by mitigating environmental confounding compared to conventional observational studies.[14]

In this study, we leveraged genetic discovery from large genome-wide association studies (GWAS) and the deep phenotyping and longitudinal design of the Growing Up in Singapore Towards healthy Outcomes (GUSTO) cohort for a comprehensive investigation of the potential mechanisms underlying the relationship between childhood obesity and language development through a triangulation method. Genetic correlations were observed between childhood obesity-related traits and language-related skills using summary-level data. We further incorporated extensive molecular and neuroimaging phenotyping in primary-aged children with cutting-edge analytic methods in the GUSTO cohort. Specifically, we constructed a trans-ancestry polygenic risk score (PRS) integrating genetic architecture knowledge from GWAS of BMI in both European and Asian populations.[15] Utilising this PRS, we were able to interrogate the unconfounded effects of genetically-predicted obesity on language ability at ages 4 and 9 years, magnetic resonance imaging (MRI)-based white matter microstructure at 7.5 years, and neurology-related proteomics at 8 years. We found that genetic predisposition to obesity, as indicated by the PRS, was associated with lower language scores. However, causality was not established based on either one-sample or two-sample Mendelian randomisation (MR). Nevertheless, to offer biological insight into language development, our investigation in the GUSTO children extended to examining the involvement of neurology-related proteins and fractional anisotropy (FA) of language-related white matter tracts, which are rarely studied in children. Specifically, we adopted a Bayesian framework and accounted for protein-protein interactions to identify proteins relevant to language development and examine if common neurology-related proteins contribute to the link between genetic predisposition to obesity and language development. The associations between candidate proteins in plasma and FA were investigated to shed light on the potential role of myelination. Mediation analysis was conducted to assess the potential mediating mechanisms underlying the effect of polygenic risk for obesity on language development via candidate proteins. To further elucidate the role of a candidate protein, ephrin-A4 (EFNA4), we constructed pathway- and brain-region-specific expression-based PRS (ePRS) focusing on the EPH-Ephrin signalling pathway.[16] Fig. 1 shows the overarching framework of this study.

Methods

GWAS summary-level data

We obtained the genetic associations of childhood BMI and obesity from GWAS meta-analyses led by the Early Growth Genetics (EGG) Consortium (childhood BMI: N = 39,620, aged 2–10 years; childhood obesity: N_{case} = 8613, $N_{control}$ = 12,696, aged 2–19 years).[17,18] Childhood obesity was defined as ≥95th percentile of BMI, and the control group maintained a BMI <50th percentile throughout childhood.[18] We obtained the genetic associations of self-reported perceived body size at age 10 from a GWAS in the UK Biobank participants (N = 453,169).[19] Childhood BMI, obesity, and self-reported perceived body size have a high genetic correlation (r_g >0.9). To achieve a larger sample size for constructing a trans-ancestry PRS in our GUSTO cohort, we obtained an East Asian GWAS of BMI from BioBank Japan (N = 163,835) and a European GWAS of BMI from a meta-analysis of GIANT and UK BioBank (N = 681,275).[20,21] Genetic associations for language-related skills were obtained from GWAS meta-analyses led by the GenLang Consortium (N = 12,828–27,180, aged 5–26 years).[1] Specifically, five language-related traits were investigated: word reading, nonword reading, spelling, phoneme awareness, and nonword repetition. These five traits were genetically correlated with each other (r_g >0.5), except that the correlation is weaker between nonword repetition and phoneme awareness (r_g = 0.30).[1] Details about the GWAS are summarised in Supplementary Table S1.

GUSTO cohort

Cohort characteristics

The GUSTO cohort is a parent-offspring prospective cohort that recruited pregnant women aged 18 and above at the National University Hospital (NUH) and KK Women's and Children's Hospital (KKH) in Singapore between June 2009 and September 2010. The cohort has been described in detail previously.[22] In brief, the original GUSTO cohort consisted of 1450 pregnancies. Excluding 246 lost to follow-up before delivery, 96 conceived by *in vitro* fertilisation, ten pairs of twins, two children without information on date of birth and sex lost to follow-up within the first week, and one death within the first week, our present study is based on follow-up of 1095 live, singleton births. Biological sex was determined based on medical records and parents' reports. The GUSTO cohort has been extensively phenotyped through numerous visits both during pregnancy and after birth. These follow-ups focused on assessing maternal health as well as monitoring the growth and development of the child. We gathered sociodemographic data via self-administered questionnaires completed by participants during pregnancy and collected obstetric information from medical records. Maternal pre-pregnancy BMI was estimated based on self-reported pre-pregnancy weight and height measured by trained staff at the 26th gestational weeks. The height and weight of the GUSTO children were measured during the annual follow-ups starting from age 3 to calculate body mass index (BMI, kg/m^2). Fat percentage was estimated using Quantitative Magnetic Resonance (QMR) during the annual follow-ups starting from age 5. The BMI and fat percentage distributions

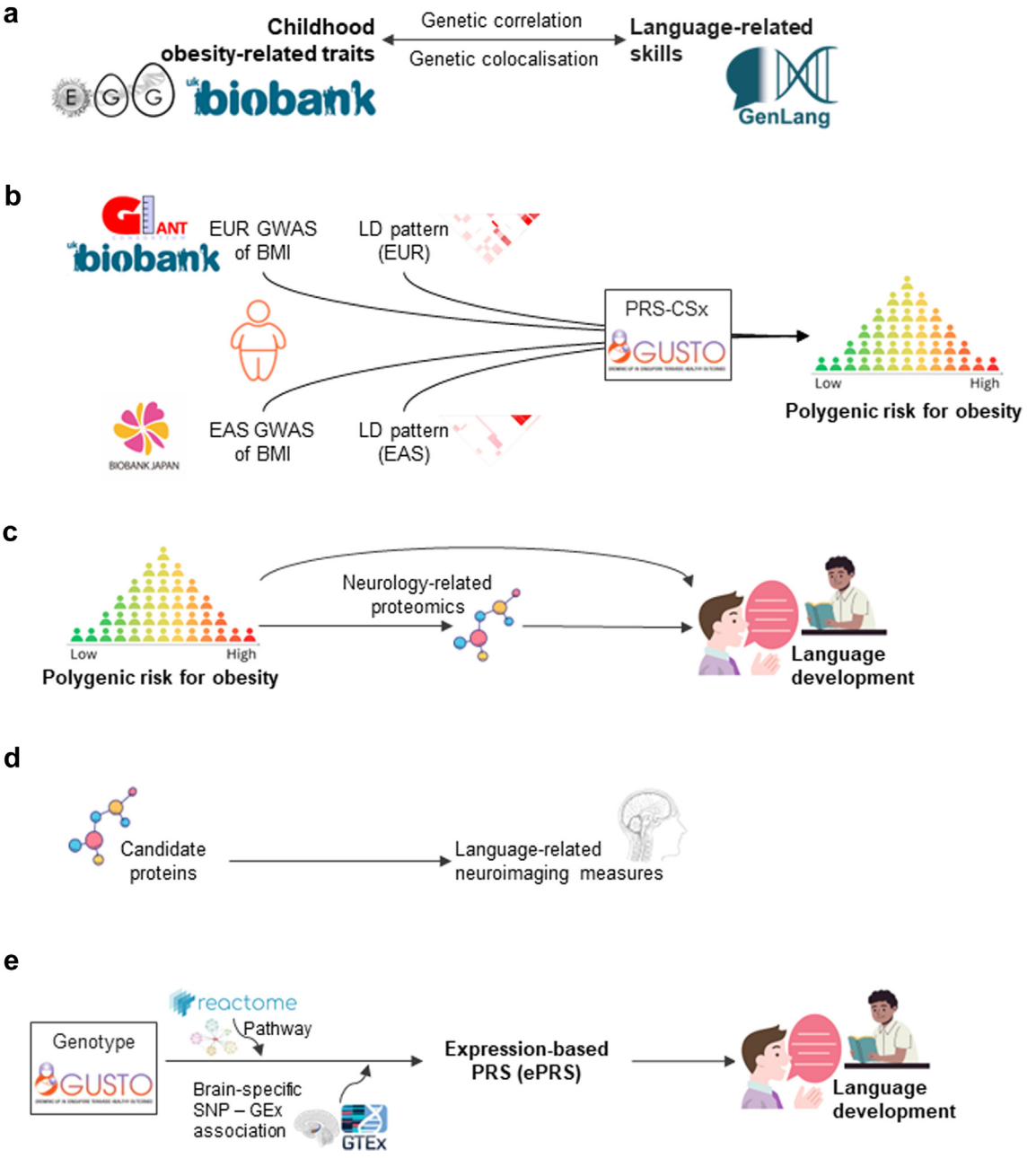

Fig. 1: Overarching framework for investigating the relationships between obesity-associated genotype to child language development. a) Genetic correlation and colocalisation for obesity-related traits and language-related skills. Childhood obesity-related traits include childhood BMI and childhood obesity from the Early Growth Genetics (EGG) Consortium and self-reported childhood body size from the UK Biobank. Language-related skills include word reading, nonword reading, spelling, phoneme awareness, and nonword repetition. b) Construction of polygenic risk score (PRS) for obesity based on genome-wide association studies (GWAS) of body mass index (BMI) in European (EUR) and East Asian (EAS) populations using a trans-ancestry Bayesian polygenic method, PRS-CSx; c) Longitudinal analysis of the associations of child PRS for obesity and language development assessed by Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4, age 4) and the Wechsler Individual Achievement Test, Third Edition (WIAT-III, age 9), and mediation analysis via protein expression. d) Cross-sectional analysis of the associations between candidate proteins (age 8) and language-related neuroimaging measures (age 7.5). Candidate proteins were prioritised based on their association with child PRS for obesity (MSR1) and WIAT-III scores (CNTN5, EFNA4, and VWC2). For neuroimaging measures, we focused on fractional anisotropy (FA) of language-related white matter tracts; e) Construction of expression-based PRS (ePRS) for the EPH-Ephrin signalling pathway. Autosomal genes on the pathway were obtained from the REACTOME database, and the brain-region-specific genetic associations of gene expression (GEx) were obtained from the Genotype-Tissue Expression (GTEx) project. Longitudinal analysis was performed for the associations between ePRS for the EPH-Ephrin signalling pathway and WIAT-III scores (age 9).

among the GUSTO children from ages 3 to 9 are shown in Supplementary Figure S1. In this study, we performed the analyses by including all participants with available data. Given that insufficient sample size increases the variability of analysis, we calculated the coefficient of variation as $\frac{Standard\ Error}{|\beta|} \times 100\%$ to assess the precision of estimates.

Genotyping and imputation

For genotyping the cohort participants and their parents, we used cord tissue from the children, maternal blood samples collected during the 26th week of pregnancy, and paternal buccal swab samples obtained approximately two to three years after the child's birth, respectively. GUSTO child and parental genotyping was performed using the Infinium OmniExpressExome array. We removed genetic variants with minor allele frequencies (MAF) <0.05, call rates <95%, and P-value for Hardy–Weinberg equilibrium $\leq 10^{-6}$. Allele frequencies were compared to those in the 1000 Genomes Project reference panel (East Asian population for Chinese and Malay; South Asian population for Indian). Genetic variants with an allele frequency differing more than 0.2 for Chinese and Indian or 0.3 for Malay were excluded. The resulting data were pre-phased using SHAPEIT v2.837 with family trio information and then imputed using the Sanger Imputation Service. Imputed data with an INFO score >0.8 were retained. The final genotyping data were available for 1025 mother-child dyads and 689 fathers after quality control (out of the 1095 live, singleton births). Principal component analysis was applied to child genotyped data, and the first three principal components were generated using PLINK 2.0 and used to account for genetic ancestry in the downstream analyses, given that the GUSTO families are of Chinese, Malay, or Indian ethnicity with homogeneous ethnic backgrounds.[23]

Language development

Language development was assessed using the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4) at age 4 (N = 702) and the Wechsler Individual Achievement Test, Third Edition (WIAT-III) at age 9 (N = 365) in the GUSTO cohort. Both tests were conducted in English, the official language used in the Singaporean school system. In brief, the PPVT-4 is an achievement test of receptive vocabulary conducted by letting the child choose one out of the four pictures corresponding to the word the experimenter said.[24] We used the WIAT-III to assess the children's academic ability in listening, speaking, and writing.[25] Specifically, a language-related composite score based on their performance in three WIAT-III tasks: pseudoword decoding, oral reading, and spelling. Both the PPVT-4 and WIAT-III scores were standardised by study sample standard deviation in the following analyses. Distributions of PPVT-4 and WIAT-III raw scores among the GUSTO children are shown in Supplementary Figure S2.

Neurology-related proteomics

Proteomic analysis was performed using the Olink Target 96 Neurology Panel.[26] This panel measures the expression levels of proteins that are implicated in neurobiological processes such as neural development and synaptic function, neurological diseases, and broader roles in processes such as cellular regulation, immunology, development and metabolism. In GUSTO, children's fasting blood samples were collected during the follow-up at age ~8 years in tubes containing ethylenediamine tetraacetic acid (EDTA). Among the 1095 children who remained in the cohort at delivery, EDTA plasma samples were available for 497 and were sent for proteomic analysis. Olink's internal control is spiked into the samples at the same concentration. Their levels are monitored for each sample and compared against the median level of internal control in the corresponding plate. If the controls differ by more than ±0.3 NPX, the sample is considered to fail the quality control. In our study, thirty-five samples failed Olink's internal quality control and were excluded from this analysis. Among the 92 neurology-related proteins being investigated (Supplementary Table S2), the expressions of five proteins (LXN (Q9BS40), GDNF (P39905), LAIR-2 (Q6ISS4), TMPRSS5 (Q9H3S3), HAGH (Q16775)) was below the limit of detection (LOD) for 0.2%–1.5% of the samples and the expression of two proteins (Beta-NGF (P01138) and MAPT (P10636)) was below the LOD for over 90% of the samples. Olink Panel measures the protein expression in a relative quantification unit, Normalized Protein eXpression (NPX), in a log2 scale, thus one NPX difference represents a doubling of protein expression. For our analysis, we did not exclude measurements below the LOD.

MRI-based white matter microstructure measures

A total of 425 participants underwent an MRI session of the brain at age ~7.5 years utilising a 3-T scanner (Magnetom Skyra; Siemens, Germany). Individuals with a head motion parameter over 3 mm were excluded from the analysis. White matter tract masks defined from AutoPtx were then applied to the fractional anisotropy (FA) map to extract mean white matter FA along the white matter tracts of interest. FA indicates the directional coherence of water diffusivity and is a marker of axonal myelination. Increasing axonal myelination across development, indexed by a higher FA, is a key process of brain maturation as it increases signal transmission efficiency along white matter tracts.[27] In this analysis, the selection of regions of interest (ROIs) was based on reviews and meta-analyses conducted on the white matter language network.[28–30] Both the dorsal stream, namely the superior longitudinal fasciculus, and the ventral stream, consisting of the inferior longitudinal fasciculus, inferior fronto-occipital fasciculus, and uncinate fasciculus, were included. Medial lemniscus

was included as a negative control, given that it primarily transports sensory information of conscious proprioception, vibration, fine touch, and two-point discrimination and has no language-related function. We conducted separate investigations on the right and left hemispheres.

Statistics

Summary-level genetic correlation and colocalisation

We performed linkage disequilibrium score regression (LD Score regression) to assess the genetic correlation between child obesity-related traits and language-related skills,[31] and genetic colocalisation analysis to assess whether two traits share a genetic causal variant within a given genomic region.[32] We reported the genetic correlation (r_g) between each obesity-related trait and each language-related skill. For genetic colocalisation, we extracted 392 independent loci reported by the three GWAS of obesity-related traits and focused on the genomic region ±100 kb from each independent locus. We investigated colocalisation between each obesity-related trait and each language-related skill under a single causal variant assumption. One locus (rs146980124) was excluded from this analysis, given that no overlapping SNPs between GWAS of obesity-related traits and language-related skills were found within ±100 kb from this locus. A posterior probability of sharing a common causal variant (PP.H4) greater than 0.80 was considered the evidence for colocalisation.

GUSTO cohort characteristics

We tabulated descriptive statistics for covariates and compared their means and proportions in the full GUSTO cohort with those in the sub-samples with both child PRS and language development outcomes under investigation in this study. We used a Mann–Whitney *U*-test for continuous covariates (maternal pre-pregnancy BMI, age at delivery, and gestational age at birth) given that the normality assumption was not fulfilled (see Supplementary Figure S3 for Q–Q plots) and a chi-squared test for categorical covariates (child sex, ethnicity, maternal highest educational level household monthly income, parity, and maternal status of GDM). In our analyses, we included covariates that are a cause of the exposure, or of the outcome, or of both, according to the disjunctive cause criterion (see Supplementary Figure S4 for the directed acyclic graphs).[33]

Trans-ancestry polygenic risk score (PRS) for obesity

We employed a trans-ancestry Bayesian PRS method, PRS-CSx, to enhance the estimation of genetic effect sizes across different populations.[15] In brief, a shared continuous shrinkage (CS) prior was used to infer posterior SNP effect sizes using the GWAS summary statistics of different ancestries and the ancestry-matched LD reference panels. In this study, we applied the "auto" feature of PRS-CSx, which is designed to automatically determine the global shrinkage parameter from the discovery summary statistics. Population-specific posterior SNP effect sizes were then combined using an inverse-variance-weighted meta-analysis. Finally, PRS was calculated based on the meta-analysed posterior SNP effect sizes using PRSice-2.[34] Given that PRS-CSx adopts a Bayesian polygenic framework, LD clumping and P-value thresholding are not performed. Specifically, we constructed PRS for obesity for the GUSTO children and parents using East Asian and European GWAS of BMI along with the 1000 Genomes Project EAS and EUR LD reference panels to estimate the cross-population effect sizes. Details of the GWAS are summarised in Supplementary Table S1. Ultimately, the meta-analysed posterior effect sizes of 862,759 SNPs were estimated and used for the final PRS construction. PRS are standardised by study sample standard deviation for downstream analyses.

To assess the association between PRS and child obesity, we applied linear Generalized Estimating Equation (GEE) models to estimate the average change in the BMI (age 3–9 years) or fat percentage (age 5–9 years) per unit higher in child PRS for obesity.[35] In brief, GEE accommodates longitudinal data by specifying a correlation structure for the repeated measurements, thereby adjusting for within-subject correlation. In this study, we used GEE models with a cluster-robust variance estimator and the first-order autoregressive structure (AR1), assuming correlations to be highest between adjacent time points. We adjusted for child age at the time of BMI and fat percentage measure and covariates, including child sex, genetic ancestry, maternal highest educational level, household monthly income, gestational age at birth, maternal age at delivery, maternal pre-pregnancy BMI, parity, maternal status of GDM. We also adjusted for maternal and paternal PRS for obesity to account for the effects of parental genomes (genetic nurture). We estimated marginal R^2 to assess the variance explained by a GEE model, including child PRS, sex, and age at the time of BMI and fat percentage measure.[35]

Unravelling the link between obesity and language development with triangulation

In GUSTO, we investigated the relationships of child PRS for obesity and childhood BMI with language development assessed by PPVT-4 and WIAT-III using multiple linear regressions. For the analysis of child PRS for obesity, we adjusted for age at language assessment, child sex, genetic ancestry, maternal highest educational level, household monthly income, pregnancy-related factors (i.e., gestational age at birth, maternal age at delivery, maternal pre-pregnancy BMI, parity, and maternal status of GDM), and parental PRS for obesity. To examine the influence of socio-economic status and other early-life factors on the phenotypic

relationship between childhood BMI and language development, we adjusted for child sex, age at the time of height and weight measure, and age at language assessment in Model 1 and additionally adjusted for genetic ancestry, maternal highest educational level, household monthly income, and pregnancy-related factors in Model 2. Additionally, we assessed the causal relationship of genetically predicted BMI and fat percentage with PPVT-4 and WIAT-III scores using a two-stage least-square (2SLS) regression, i.e., one-sample MR analysis. Specifically, we constructed a Clumping and Thresholding (C + T) PRS for obesity using only the independent, genome-wide significant SNPs (P-value<5×10^{-8} and R^2<0.001, Supplementary Table S3) from the same East Asian GWAS of BMI from BioBank Japan as we used for the trans-ancestry PRS.[36] We used the conventional clumping and thresholding PRS as the genetic instrumental variable to minimise the risk of pleiotropy in the MR analysis. In the 2SLS regression, we adjusted for child sex, age at language assessment, age at BMI and fat percentage measurements, genetic ancestry, and parental PRS for obesity. Sex-specific analyses were performed by fitting the same model only in boys and only in girls separately. We applied a stringent Bonferroni correction with a P-value threshold of 0.05/(2 × 3) for the analyses of language development to account for multiple comparisons of two language scores and three sex groups (Overall (Boys and Girls), Boys, and Girls).

Leveraging genetic associations from large GWAS in individuals of European ancestry, we performed univariable two-sample MR analyses to investigate the causal relationships between childhood obesity-related traits and language-related skills. For each MR analysis, we selected single-nucleotide polymorphisms (SNPs) associated with an obesity-related trait at a P-value<5×10^{-8} and with an F-statistics>10. We kept only the independent SNPs ($r^2 < 0.001$) by performing linkage disequilibrium (LD) clumping with a window size 10,000 kb and removed correlated SNPs that were less strongly associated with the exposure based on their P-values. To minimise the risk of reverse causation, we performed Steiger filtering to remove SNPs with a stronger association with the outcome than with the exposure.[37] Variant harmonisation was performed, and non-inferable palindromic SNPs were dropped. These criteria ensured that SNPs selected as instrumental variables for MR analyses met the relevance assumption.[36] We listed the SNPs used for each MR analysis in Supplementary Table S4. SNP-specific effects were estimated by dividing the SNP-outcome association by the SNP-exposure association (Wald estimator) for each SNP selected as the instrumental variable.[36] We then used the inverse-variance weighted (IVW) random effects model to pool the SNP-specific Wald estimates, weighted by the inverse variance of the SNP-outcome association. MR analysis also assumes no confounders between the instrumental variables and the outcome (i.e., exchangeability), and the instrumental variables only influence the outcome via the exposure (i.e., exclusion restriction).[36] These two assumptions are not verifiable but will be assessed through sensitivity analyses in this study. We applied three sensitivity methods (weighted median (WM), MR-Egger regression, and the Joint Analysis of Marginal algorithm for MR (JAM-MR)) to test the robustness of MR causal estimates. In brief, the WM method is consistent when up to 50% of the SNPs were invalid instrumental variables,[38] the MR-Egger regression allows for pleiotropic effects, assuming that the magnitude of pleiotropic effects is independent of the strength of the SNP-exposure association,[39] and JAM-MR is a Bayesian inference method that downweighs SNPs with pleiotropic effects.[40] Additionally, we applied MR-Radial to detect outlier SNPs that substantially influence the analysis and are likely to contribute to the violation of MR assumptions.[41] Sex-specific analyses were not considered given that sex-specific GWAS were not available. We accounted for multiple comparisons using Bonferroni correction with a P-value threshold of 0.05/(3 × 5) for three obesity-related traits and five language-related skills. We reported the MR results following the STROBE-MR guideline.[42]

Linking polygenic risk for obesity, neurology-related proteins, language-related white matter tract formations, and language development

In GUSTO, we adopted a clustering and mixture method to identify neurology-related proteins implicated in language development. Protein–Protein interaction clusters were identified based on the STRING database (Version 12.0)[43] and using the edge-betweenness clustering method.[44] Following this, we performed Bayesian kernel machine regression (BKMR) with hierarchical variable selection to identify protein clusters and individual proteins relevant to language development.[45] To establish the temporal relationship, we only considered the WIAT-III score in this analysis, as WIAT-III was assessed at a later time point than the blood sample collection for proteomic analysis. Detailed methods are described in our previous work investigating blood biomarker mixtures.[46] In brief, we performed 500,000 Markov chain Monte Carlo (MCMC) iterations (50% burn-in) with 10 independent chains. We estimated posterior inclusion probabilities (PIPs) for each protein cluster (Cluster PIP) and the PIPs for each protein, given that its cluster was selected into the model (Conditional PIP). Sex-specific BKMR analysis was not performed as the restricted sample sizes led to instability and non-convergence in the MCMC iteration process.[47] We considered protein clusters with a Cluster PIP greater than 0.50 to be relevant to the WIAT-III score. Within each cluster, we evaluate the relevance of individual proteins based on their Conditional PIP.

We validated the direction of associations using linear regression models for each protein and performed sex-specific analysis for candidate proteins. Considering the exploratory nature of this analysis, we chose a parsimonious model adjusting for child sex, genetic ancestry, age at WIAT-III assessment and age at blood draw for proteomic analysis, plate ID for proteomic analysis, mother's highest educational level, and household monthly income.

We also investigated the sex-specific associations of child PRS for obesity with neurology-related protein expression. We fitted multiple linear regressions for each of the 92 proteins separately, adjusting for age at blood draw for proteomic analysis, child sex, genetic ancestry, maternal highest educational level, household monthly income, pregnancy-related factors, parental PRS for obesity, and the plate ID for proteomic analysis as a technical covariate. We applied a stringent Bonferroni correction with a P-value threshold of $0.05/(92 \times 3)$ as our level of formal significance to account for multiple comparisons of 92 proteins and three sex groups (Overall (Boys and Girls), Boys, Girls). As an exploratory study, we also indicated associations that are nominally significant with a P-value <0.05.

We further investigated the associations between candidate neurology-related proteins and FA in white matter tracts using multiple linear regression adjusting for children's sex, genetic ancestry, age at MRI scan and age at blood draw for proteomic analysis, plate ID for proteomic analysis, mother's highest educational level, and household monthly income. As an exploratory analysis, we highlighted associations with P-values smaller than 0.05 and 0.01.

Mediation analysis was performed using a regression-based counterfactual approach, which decomposes the causal effect into pure and total (including interaction) direct effects and pure and total (including mediated-interaction) indirect effects.[48] Specifically, we assessed the indirect effect of child PRS for obesity on WIAT-III language score via two candidate neurology-related proteins, i.e., EFNA4 and VWC2, based on the findings in the aforementioned analyses. We adjusted for age at language assessment, age at blood draw for proteomic analysis, child sex, genetic ancestry, maternal highest educational level, household monthly income, parental PRS for obesity, gestational age at birth, maternal age at delivery, maternal pre-pregnancy BMI, parity, maternal status of GDM, and plate ID for proteomic analysis. Additionally, we examined the genetic correlation between BMI and expression of EFNA4 and VWC2 using GWAS of adult populations of European ancestry (Supplementary Table S1).

Brain-region-specific predicted expression (ePRS) analyses

Among the candidate neurology-related proteins for the WIAT-III score identified in our analysis, EFNA4 was more strongly associated with PRS for obesity as well as FA in multiple white matter tracts. In addition, EFNA4 has been reported to regulate myelination, which is relevant to language ability.[49] Thus, we further examined the influence of the expression of the EPH-Ephrin signalling pathway on the WIAT-III score. We constructed pathway- and brain-region-specific ePRS based on a method we developed and applied in previous studies.[16,50] A total of 90 autosomal genes were identified for the EPH-Ephrin signalling pathway (R-HSA-2682334.3) using the REACTOME database (release 88). SNPs from these genes were annotated, and their associations (effect sizes and P-values) with brain-region-specific gene expression were obtained from the Genotype-Tissue Expression (GTEx) Project V7.[51] ePRS for the GUSTO children were then constructed by summing the effect alleles of the uncorrelated SNPs (r^2 >0.2 across 500 kb region) weighted by their effect sizes for gene expression. In brief, the pathway-specific ePRS represent individual variations in the expression of the EPH-Ephrin signalling pathway in the five brain tissues under investigation in this study, namely the amygdala, caudate, hippocampus, putamen, and prefrontal cortex.

Software and packages

Trans-ancestry PRS were constructed using *PRS-CSx* and *PRSice-2 v2.3.5*. GEE was performed using R package *geepack 1.3.11*. Bayesian kernel machine regression was performed using R package *bkmr 0.2.2*. MR analyses were performed using R packages *TwoSampleMR 0.6.7*, *ieugwasr 1.0.1*, *R2BGLiMS*, and *RadialMR*. Genetic correlation was estimated using *ldsc* v1.0.1. Genetic colocalisation analysis was performed using *coloc* v5.2.2. 2SLS regression was performed using the R package *ivreg 0.6–3*. ePRS were generated using PRSoS.[52] Mediation analysis was performed using the R package *regmedint 1.0.1*. All statistical analyses were performed using R version 4.2.3.

Ethics

Ethics approval was obtained from Centralised Institutional Review Board of SingHealth (2018/2767) and the Domain Specific Review Board of Singapore National Healthcare Group (D/2009/021, B/2014/00414). Informed consent was obtained from the parents and from the parents on behalf of their children.

Role of funders

None of the funders were directly involved in the design, data collection, analysis, or interpretation, or writing of this report.

Results

Genetic correlation and colocalisation for obesity and language development

Childhood obesity was inversely correlated with spelling (r_g = –0.23, P-value = 2.0×10^{-4}, 95% CI –0.36 to –0.11),

word reading (r_g = –0.15, P-value = 0.0087, 95% CI –0.27 to –0.04), and nonword reading (r_g = –0.18, P-value = 0.011, 95% CI –0.32 to –0.04) based on LD Score regression (Table 1). Colocalisation was only identified for child body size and word reading within ±100 kb from rs7503580 (PP.H4 = 0.82, Table 1 and Supplementary Table S5), and the four SNPs on *BAIAP2* contributed the highest SNP-specific PP.H4 conditional on H4 being true (PP.H4 = 0.41 for rs12449515, PP.H4 = 0.24 for rs12944344, PP.H4 = 0.18 for rs12949063, and PP.H4 = 0.17 for rs4072587).

Trans-ancestry PRS for obesity in GUSTO

Among the 1095 GUSTO children, 733 had both genotyping data for PRS construction and at least one language development outcome under investigation in this study. We did not observe substantial differences between the entire cohort and the sub-sample regarding important demographic characteristics, such as maternal ethnicity and household monthly income, or clinical covariates, such as maternal pre-pregnancy BMI (Table 2).

The marginal R^2 for BMI and fat percentage were 0.11 (Girls = 0.10, Boys = 0.11) and 0.12 (Girls = 0.070, Boys = 0.10), respectively, based on the GEE models including child PRS, sex, and age at BMI and fat percentage measurement. Additionally adjusting for genetic ancestry, maternal highest educational level, household monthly income, pregnancy-related covariates and parental PRS, child PRS for obesity remained positively associated with childhood BMI (GEE model: $\beta = 0.93$, P-value = 2.1×10^{-13}, 95% CI 0.68–1.18) and fat percentage (GEE model: β = 2.63, P-value = 7.8×10^{-8}, 95% CI 1.67–3.59), as shown in Supplementary Table S6. These associations were consistent in both boys and girls. Expectedly, the C + T PRS for obesity using only the independent, genome-wide significant SNPs had a lower marginal R^2 of 0.045 for BMI (Girls = 0.040, Boys = 0.045) and of 0.10 for fat percentage (Girls = 0.048, Boys = 0.053) and was less strongly associated with childhood BMI and fat percentage (Supplementary Table S6).

Causation not identified between obesity and language development

Fig. 2 and Table 3 show that child PRS for obesity was inversely associated with child language scores as assessed by PPVT-4 (age 4) and WIAT-III (age 9) in the GUSTO cohort. These associations were stronger in boys compared to girls. After accounting for multiple comparisons, we found that a higher child PRS for obesity remained associated with a lower WIAT-III language score in the overall group (linear regression: β = –0.37, 95% CI –0.55 to –0.19, P-value = 6.5×10^{-5}) and in boys (linear regression: β = –0.56, 95% CI –0.81 to –0.31, P-value = 2.5×10^{-5}). The association in girls was less evident (linear regression: β = –0.27, 95% CI –0.53 to 9.8×10^{-5} P-value = 0.050). However, phenotypic childhood BMI was not associated with PPVT-4 or WIAT-III language scores (Supplementary Table S7) and an attenuation in the effect sizes was observed upon further adjustment for genetic ancestry, maternal highest educational level, household monthly income, and pregnancy-related factors. This attenuation was more pronounced for the associations between BMI and WIAT-III language score in boys. Using one-sample

Obesity-related trait	Language-related skill	Genetic correlation			Genetic colocalisation	
		r_g	P-value[a]	95% CI	Lead SNP	PP.H4
Childhood BMI	Nonword Reading	3.8×10^{-3}	0.96	−0.13 to 0.14	rs10182458	0.24
Childhood BMI	Nonword Repetition	−0.064	0.50	−0.25 to 0.12	rs35926495	0.60
Childhood BMI	Phoneme Awareness	−0.025	0.74	−0.17 to 0.12	rs11865086	0.09
Childhood BMI	Spelling	−0.10	0.13	−0.23 to 0.029	rs13107325	0.21
Childhood BMI	Word Reading	−0.029	0.65	−0.15 to 0.094	rs62037365	0.21
Childhood Body Size	Nonword Reading	−0.025	0.53	−0.10 to 0.054	rs6931604	0.42
Childhood Body Size	Nonword Repetition	−0.067	0.20	−0.17 to 0.036	rs35926495	0.68
Childhood Body Size	PhonemeAwareness	−0.071	0.11	−0.16 to 0.017	rs12798028	0.46
Childhood Body Size	Spelling	−0.062	0.14	−0.14 to 0.021	rs1199333	0.54
Childhood Body Size	Word Reading	−0.032	0.37	−0.10 to 0.038	rs7503580	0.82
Childhood Obesity	Nonword Reading	−0.18	0.01	−0.32 to −0.042	rs10182458	0.17
Childhood Obesity	Nonword Repetition	−0.10	0.28	−0.28 to 0.081	rs3735781	0.33
Childhood Obesity	Phoneme Awareness	−0.13	0.111	−0.28 to 0.029	rs74583214	0.10
Childhood Obesity	Spelling	−0.23	2.0×10^{-4}	−0.36 to −0.11	rs141458884	0.32
Childhood Obesity	Word Reading	−0.15	0.0087	−0.27 to −0.039	rs62037365	0.42

Abbreviations: confidence interval (CI) inverse-variance weighted (IVW), Mendelian randomisation (MR), single nucleotide polymorphisms (SNP), posterior probability of both traits are associated and share a single causal variant (PP.H4). [a]P-value obtained from Linkage Disequilibrium (LD) Score regression.

Table 1: **Genetic correlation and colocalisation for obesity-related traits and language-related skills based on summary-level data.**

	GUSTO cohort (N = 1095)	Sub-sample (N = 733)	P-value[a]
	Mean (SD)	Mean (SD)	
Maternal pre-pregnancy BMI (kg/m^2)	22.7 (4.42)	22.8 (4.35)	0.71
Mother's age at delivery (years)	30.9 (5.13)	31.2 (5.02)	0.30
Gestational age at birth (weeks)	38.7 (1.57)	38.9 (1.27)	0.32
	N (%)	**N (%)**	
Sex			
Boys	572 (52.2%)	371 (50.6%)	0.53
Girls	523 (47.8%)	362 (49.4%)	
Ethnicity			
Chinese	598 (54.7%)	412 (56.2%)	0.71
Indian	205 (18.7%)	127 (17.3%)	
Malay	291 (26.6%)	194 (26.5%)	
Household monthly income (S$)			
0–1999	163 (15.9%)	111 (16.0%)	0.86
2000–3999	324 (31.6%)	208 (30.0%)	
4000–5999	255 (24.9%)	172 (24.8%)	
≥6000	282 (27.5%)	202 (29.1%)	
Mother's highest education level			
Mother has a university degree	355 (32.8%)	251 (34.5%)	0.50
Mother has no university degree	726 (67.2%)	477 (65.5%)	
Parity status			
Parous	627 (57.3%)	418 (57.0%)	0.96
Nulliparous	468 (42.7%)	315 (43.0%)	
Gestational diabetes mellitus			
Yes	188 (18.0%)	120 (17.0%)	0.64
No	854 (82.0%)	584 (83.0%)	

Abbreviations: body mass index (BMI); standard deviation (SD). [a]P-values were from a Mann–Whitney *U*-test for continuous covariates and from a chi-squared test for categorical covariates.

Table 2: **Cohort characteristics for 1095 children and the subset of 733 children who have both genotyped data for PRS and at least one language development outcome in the GUSTO cohort.**

MR analysis in GUSTO, we found inverse associations between Bayesian PRS predicted BMI and WIAT-III scores (Supplementary Table S8). However, such associations were not replicated for BMI predicted by the C + T PRS (Supplementary Table S7). Consistently, we did not find evidence supporting a causal relationship between childhood obesity-related traits and language-related skills using two-sample MR analyses based on European GWAS (Supplementary Figure S5). Sensitivity analyses excluding outlier SNPs identified by MR-Radial remained consistent. MR-Egger did not suggest potential pleiotropic effects (Supplementary Tables S9–S11).

Neurology-related proteins linking polygenic risk for obesity, language development, and language-related white matter tracts

Among the 92 neurology-related proteins under investigation, we identified 28 clusters based on their protein–protein interactions (Fig. 3 and Supplementary Table S12). Using the BKMR mixture method, contactin-5 (CNTN5; Cluster PIP = 0.79) was the single protein that was most relevant to the WIAT-III score, followed by brain-specific chordin-like protein, also called VWC2 (Cluster PIP = 0.56). Protein cluster 4 was also found to be associated with the WIAT-III score (Cluster PIP = 0.51), and EFNA4 (Conditional PIP = 0.48) was the most relevant protein within this cluster. These findings based on BKMR were consistent with those based on linear regression (Fig. 4 and Supplementary Table S13), where EFNA4 and VWC2 were inversely associated with WIAT-III score (EFNA4: β = –0.90, P-value = 0.0010, 95% CI –1.44 to –0.36; VWC2: β = –0.41, P-value = 0.014, 95% CI –0.74 to –0.08), and CNTN5 was positively associated with WIAT-III score (β = 0.45, P-value = 0.0070, 95% CI 0.13–0.77). Supplementary Table S13 suggested sex-specific associations between these proteins and WIAT-III score, with EFNA4 (P-value = 0.0029) and VWC2 (P-value = 0.0030) more relevant for boys and CNTN5 more relevant for girls (P-value = 0.037). Surprisingly, EFNA4 was positively associated with FA of language-related white matter tracts in girls but not in boys, with more substantial effects in the inferior longitudinal fasciculus and inferior fronto-occipital fasciculus (Supplementary Figure S6). Similarly, VWC2 was positively associated with FA of the left inferior longitudinal fasciculus and left inferior fronto-occipital fasciculus only in girls (Supplementary Figure S7). EFNA4 and VWC2 were robust to confounding, with no observed associations with FA of the medial lemniscus (negative control). On the contrary, CNTN5 was associated with FA of the medial lemniscus (negative control) but not with FA of other language-related white matter tracts (Supplementary Figure S8). Detailed results on the association of each protein with each functional brain region are presented in Supplementary Table S14.

On the other hand, macrophage scavenger receptor 1 (MSR1) was the only protein that remained associated with child PRS for obesity after accounting for multiple comparisons (linear regression: Overall: β = 0.15, P-value = 1.0×10^{-4}, 95% CI 0.08–0.23; Boys: β = 0.24, P-value = 2.7×10^{-5}, 95% CI 0.13–0.35). The association between child PRS and obesity with MSR1 in girls was not observed in our analysis (linear regression: β = 0.05, P-value = 0.36, 95% CI –0.06 to 0.17). Despite not being associated with the WIAT-III score, MSR1 was positively associated with FA of the right inferior longitudinal fasciculus and the left inferior fronto-occipital fasciculus (Supplementary Figure S9). Child PRS for obesity was only weakly associated with EFNA4 and VWC2 in boys (linear regression: EFNA4: β = 0.08, P-value = 0.023, 95% CI 0.01–0.15; VWC2: β = 0.10, P-value = 0.042, 95% CI 0.004–0.20; Supplementary Table S15).

Nevertheless, mediation analysis does not support the mediating mechanisms via EFNA4 or VWC2 expression for the effect of higher polygenic risk for obesity on a lower WIAT-III language score (Fig. 5 and Supplementary Table S16). Instead, we found strong

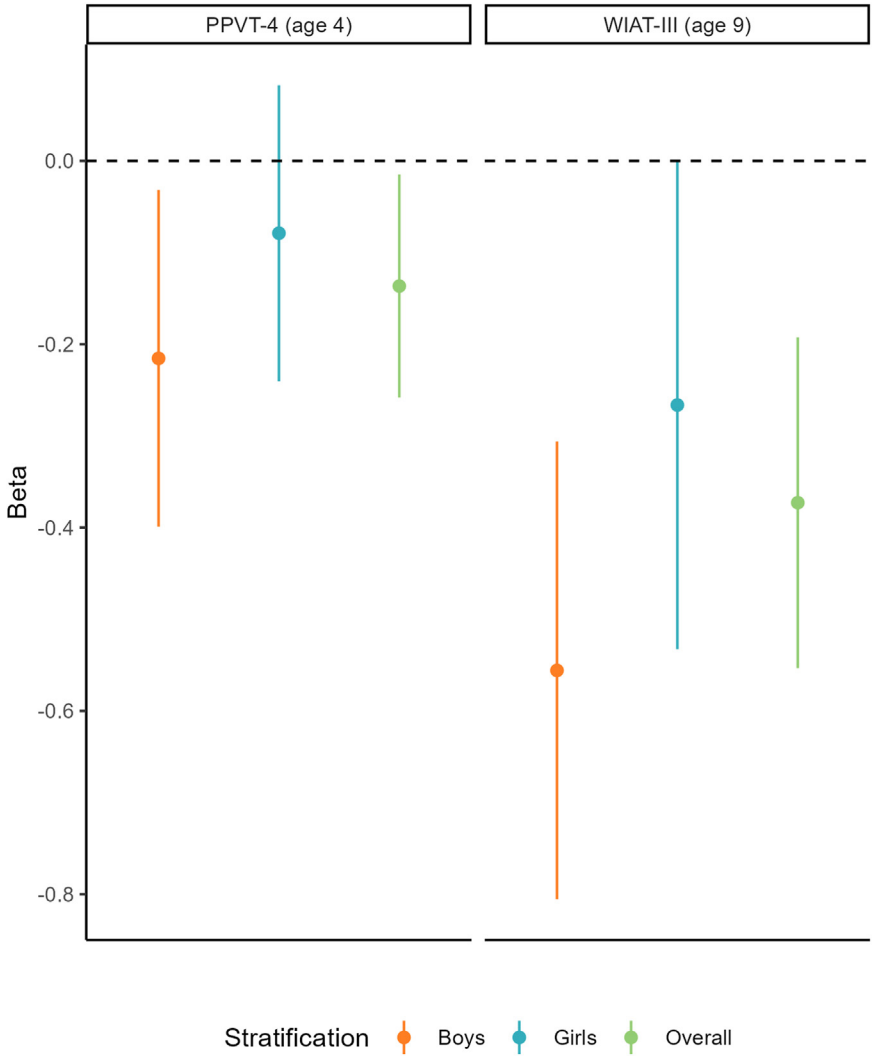

Fig. 2: The association of child polygenic risk score (PRS) for obesity with child language scores assessed by PPVT-4 and WIAT-III from multiple linear regression (PPVT-4: $N_{Overall}$ = 448, N_{Boys} = 224, N_{Girls} = 224; WIAT-III: $N_{Overall}$ = 235, N_{Boys} = 121, N_{Girls} = 114). Line segments indicate 95% confidence intervals. (PPVT-4: Peabody Picture Vocabulary Test, Fourth Edition; WIAT-III: Wechsler Individual Achievement Test, Third Edition; P-values were obtained from multiple linear regression).

Outcome	Stratification	Sample size	Beta	P-value[a]	Adjusted P-value[a]	95% CI
PPVT-4 (YR4)	Overall	448	−0.14	0.028	0.17	−0.26 to −0.015
PPVT-4 (YR4)	Girls	224	−0.08	0.34	1.0	−0.24 to 0.082
PPVT-4 (YR4)	Boys	224	−0.22	0.022	0.13	−0.40 to −0.032
WIAT-III (YR9)	Overall	235	−0.37	6.5×10^{-5}	3.9×10^{-4}	−0.55 to −0.19
WIAT-III (YR9)	Girls	114	−0.27	0.050	0.30	−0.53 to -9.8×10^{-5}
WIAT-III (YR9)	Boys	121	−0.56	2.5×10^{-5}	1.5×10^{-4}	−0.81 to −0.31

Abbreviations: standard error (SE), confidence interval (CI). [a]P-value obtained from linear regression. Adjusted P-value accounted for multiple comparisons of two language scores and three sex groups.

Table 3: **Associations of child polygenic risk score (PRS) for obesity and language-related skills (the GUSTO cohort).**

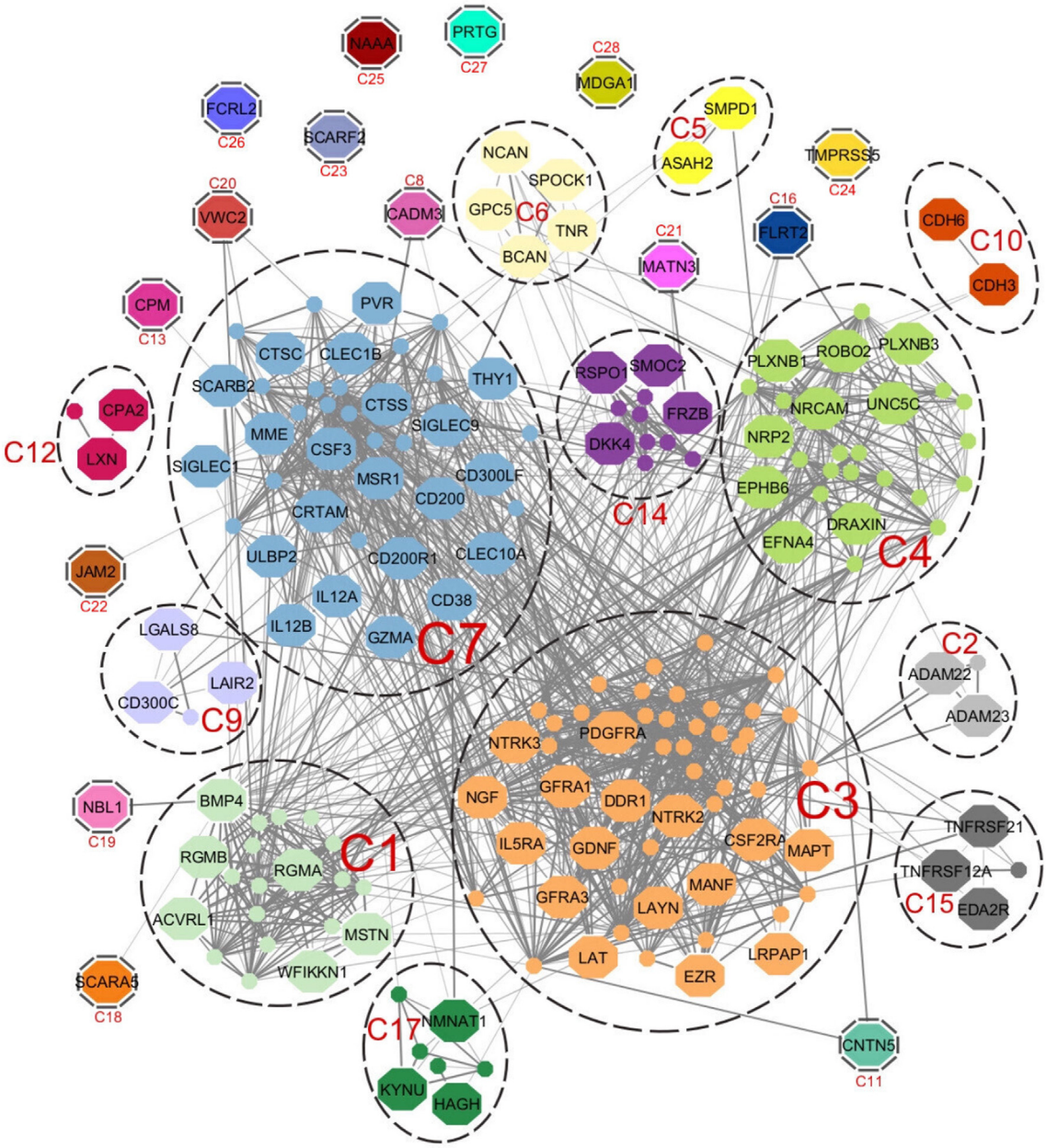

Fig. 3: Clusters of protein-coding genes based on protein–protein interaction. Larger nodes represent coding genes of neurology-related proteins measured in the GUSTO cohort. Smaller nodes represent coding genes of proteins that are associated with the neurology-related proteins based on the STRING database. Each edge represents the interaction between two proteins (only interactions with a STRING combined score ≥0.4 are presented).

genetic correlations between BMI and these two proteins (Supplementary Table S17). Given the more evident findings in our analyses and previously reported regulatory role in myelination of EFNA4, we further investigated brain-region-specific predicted expression related to EPH-Ephrin signalling. Consistent with the proteomic analysis, a higher ePRS of EPH-Ephrin signalling in the hippocampus was associated with a lower WIAT-III score in boys (Supplementary Table S18, linear regression: β = –0.21, P-value = 0.0062, 95% CI –0.35 to –0.06).

Discussion

In this study, we found evidence of a genetic correlation between obesity-related traits and language-related skills

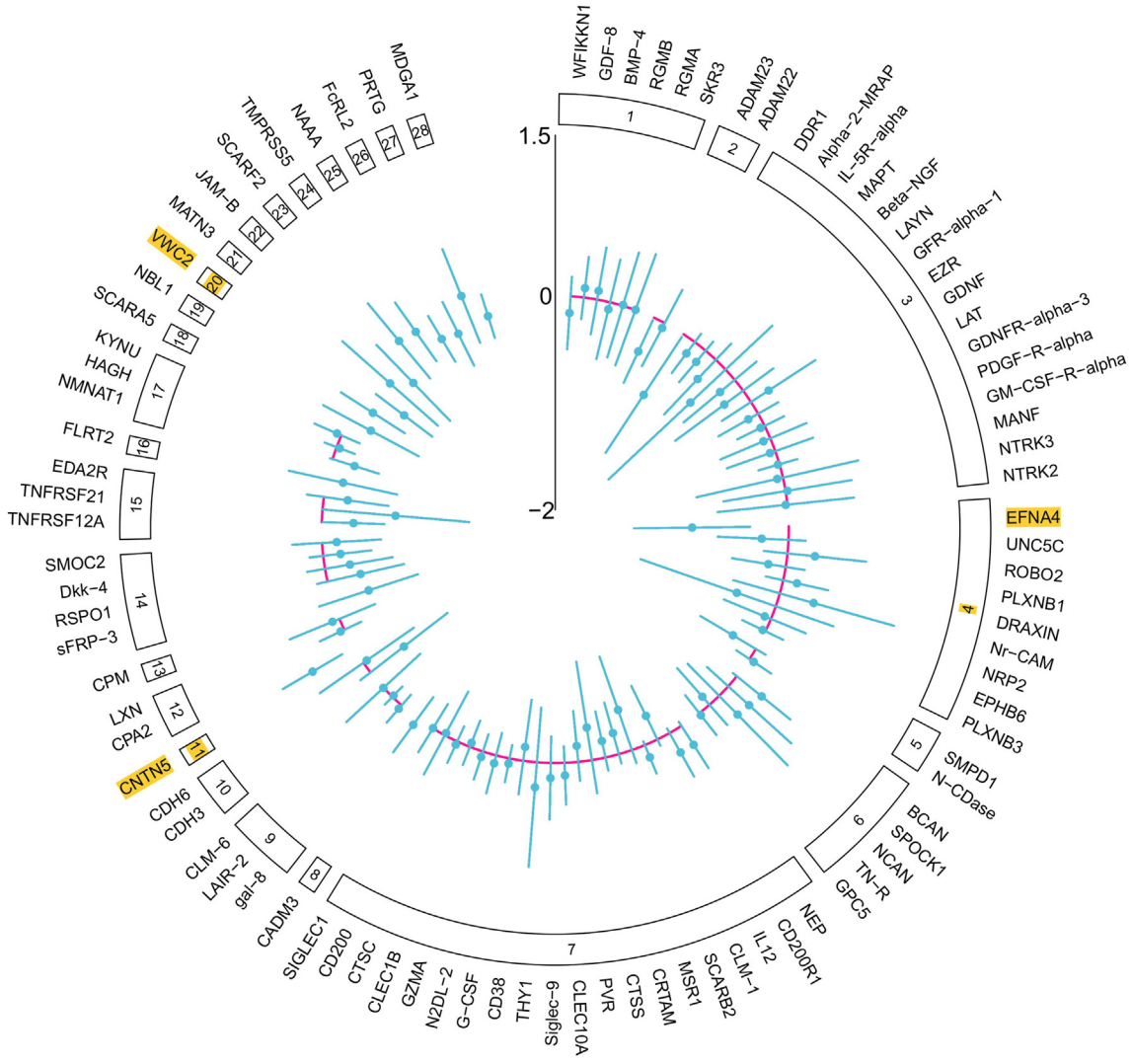

Fig. 4: Cluster-specific forest plot for the associations of neurology-related proteins with WIAT-III language-related composite score (N = 215). The clusters with a posterior inclusion probability (PIP) higher than 0.5 and the protein with the highest conditional PIP within those clusters are highlighted in yellow. The dots and segments indicate the point estimates and 95% confidence intervals from the linear regression model.

leveraging large GWAS summary statistics. In the GUSTO cohort, child PRS for obesity, which is less susceptible to socio-environmental confounding, was associated with both lower PPVT-4 and WIAT-III scores, particularly in boys. However, one-sample MR in the GUSTO cohort and two-sample MR based on summary-level GWAS data did not support a causal relationship between obesity and language development. Particularly, the contradicting findings of the one-sample MR analyses using Bayesian PRS and C + T PRS suggested that polygenic risk for obesity was associated with language development due to non-BMI-related mechanisms. Our examination of the neurology-related proteins and FA of language-related white matter tracts provided biological insight into language development. We measured a range of circulating neurology-related proteins in the GUSTO children and found sex-specific associations of EFNA4, VWC2, and CNTN5 with the WIAT-III score. Particularly, higher polygenic risk for obesity was nominally and positively associated with EFNA4 and VWC2 expression, both of which were inversely associated with the WIAT-III language-related composite score. However, mediating mechanisms via EFNA4 or VWC2 were not supported. The inverse association between EFNA4 and WIAT-III score was more pronounced in boys, which is supported by the consistent association between gene expression of the EPH-Ephrin signalling pathway in the hippocampus and

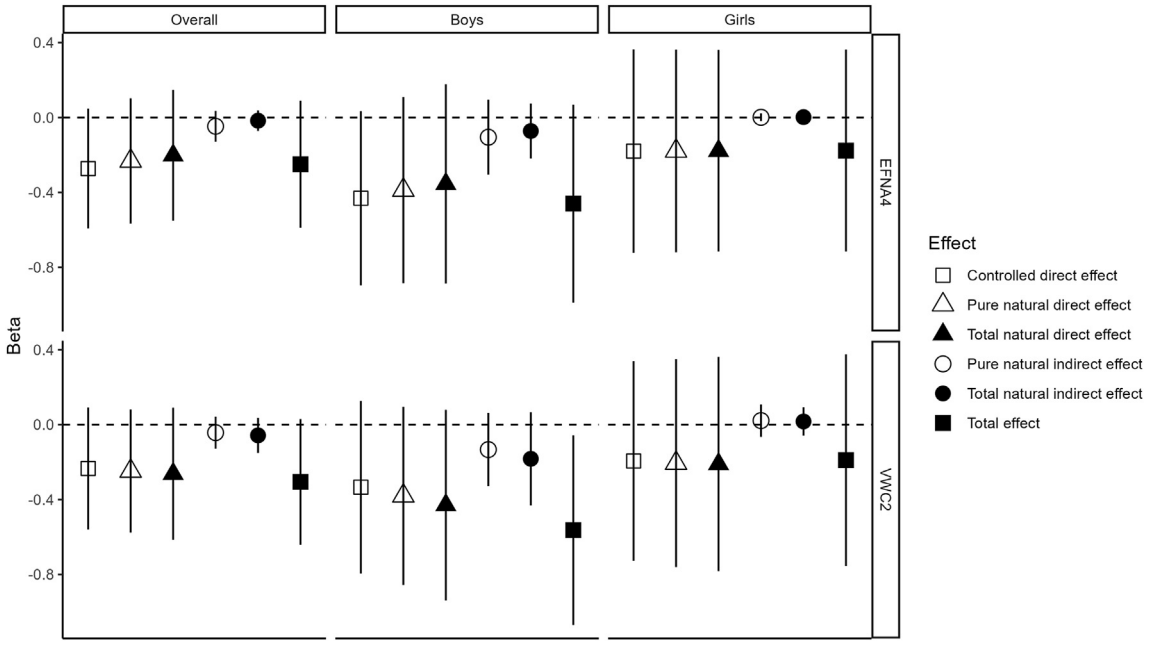

Fig. 5: Mediation analysis for child PRS for obesity, neurology-related protein (YR8), and WIAT-III language score (YR9). Line segments indicate 95% confidence intervals. ($N_{Overall}$ = 150, N_{Boys} = 77, N_{Girls} = 73).

WIAT-III score, as well as the existing knowledge about the regulatory role of Ephrin A4 receptor in myelination.[53]

EFNA4 belongs to the ephrin family, and it binds to various Ephrin A receptors. The Ephrin A4 receptor, coded by the *EPHA4* gene, is a primary receptor that binds to EFNA4, and it negatively regulates myelination in both the central and peripheral nervous systems.[53] It has also been found to modulate synaptic plasticity and neurogenesis.[54,55] Myelin deposition in the language-related brain regions improves children's vocabulary development.[7,49] More recently, the in–home language environment has been associated with brain myelination in early development.[56] In our study, we observed a positive association of EFNA4 with FA in the inferior longitudinal fasciculus and inferior fronto-occipital fasciculus in girls, which is an indirect marker of myelination. The inferior longitudinal fasciculus and inferior fronto-occipital fasciculus contribute to language comprehension and semantic processing, playing critical roles in the understanding and interpretation of language.[57,58] The associations of EFNA4 with FA in these white matter tracts in boys, though not significant, were in the opposite direction to those in girls. Another protein in the ephrin family, EFNA1, has shown sex-dimorphic gene expression patterns in Parkinson's disease, with over-expression in male patients but under-expression in female patients compared to the corresponding sex-specific control groups.[59] Due to the binding preference of EFNA4, its interactions with Ephrin A receptors can vary by cellular context and developmental stage, which may explain the observed sex-dimorphic associations between EFNA4 and FA in the white matter tracts. Our study further revealed that the EPH-Ephrin signalling pathway, rather than the isolated expression of the EFNA4 protein, may play a role in language development. On the other hand, CNTN5 plays a pivotal role in maintaining neural circuitry within the auditory system and may be responsible for delayed auditory responses in children diagnosed with ASD.[60] VWC2 functions as an antagonist to bone morphogenetic proteins (BMPs), which are essential in the development of the central nervous system.[61] However, it is less clear how VWC2 may be implicated in early-life language development. The associations of CNTN5 and VWC2 with FA in the white matter tracts were not as strong as those observed for EFNA4, suggesting different biological mechanisms may be involved. Our investigation did not support mediating mechanisms underlying the effect of polygenic risk for obesity on language development via EFNA4; however, polygenic risk for obesity may influence multiple proteins relevant to the EPH-Ephrin signalling pathway, which was associated with language score in this study. This suggests a more comprehensive early-life proteomic profiling warrants further investigation to unravel the link between cardiometabolic health and neurodevelopment at the molecular level.

Recent studies have highlighted the potential of large-scale plasma proteomics in elucidating the

complexities of human health and ageing.[62,63] Early-life proteomics can provide insight into human development which lay a foundation for future health trajectories. Specifically, EFNA4 and VWC2 exhibit inverse associations with general fluid cognitive ability in later life, whereas CNTN5 is positively associated with a general factor of FA derived from multiple white matter tracts.[64] These observations in the older population align with our findings in the younger population, suggesting the relevance of these proteins in early-life neurodevelopment and lifelong cognitive health. In this study, we also found that genetic risk for obesity exhibited a stronger association with MSR1, which is involved in regulating immune response and inflammatory response.[65] It also mediates the uptake of modified low-density lipoprotein (LDL), particularly oxidised LDL,[66] which may have implications for cognitive decline and neurodegeneration.[67,68] Although we did not find an association between MSR1 and language development, it has been associated with lower general fluid cognitive ability and white matter hyperintensity volume, which is correlated with cognitive impairment, in later life.

Limitations exist in this study of complex traits. First, language development in the GUSTO children was assessed using English-based tests. We were not able to take into account the mother tongue languages in the multi-ethnic setting in Singapore, e.g., Chinese, Malay, and Tamil/Non-Tamil Indian languages. However, functional magnetic resonance imaging (fMRI) investigation showed that the neural architecture of the language network is consistent across speakers of 45 different languages.[69] This may suggest that our findings are also applicable to speakers of other languages. Second, while the language tests we used focused on assessing vocabulary and reading abilities, it is important to recognise that language development involves other domains. For instance, pragmatic development is crucial for facilitating effective communication in various social contexts,[70] which may be intricately linked with issues related to obesity and self-esteem. Third, the sample size in the GUSTO cohort is small compared to population-based adult cohorts and biobanks. Thus, we are not able to split the GUSTO cohort into training and testing sets for PRS construction. However, we showed that our Bayesian PRS is highly associated with childhood BMI and fat percentage. The small sample sizes in this study also reduced the precision of our estimates as reflected by coefficients of variation exceeding 20%, particularly in the sex-specific analyses. The high variability limited the statistical power to detect associations of interest given the expected small effect sizes in this study. Fourth, GWAS meta-analysis of language-related skills may be susceptible to heterogeneity due to the different age ranges and assessments across cohorts. Nevertheless, participants from most cohorts were below 18 years of age, and the choice of assessment tools was aimed at maintaining consistency among the different cohorts. Fifth, GWAS of childhood body size relies on self-reporting during adulthood, which may introduce bias, especially since adults with larger body sizes might recall their early-life body size differently compared to thinner adults. By investigating three obesity-related traits, our MR analysis provided consistent results. Sixth, given that large-scale genetic studies of child obesity-related traits and language-related skills are only available for the European population, generalisation of these results to other populations requires cautious interpretation. Nevertheless, one-sample MR analysis in the GUSTO cohort showed consistent results. Last but not least, unmeasured confounding may exist in the association analyses for PRS for obesity and language outcomes and the mediation analyses.

Our study revealed that the genotypes associated with obesity are linked to early-life language development in children. This association is likely due to a shared genetic basis between obesity and language-related skills rather than a causal effect of obesity. Although mediating mechanisms underlying the effect of polygenic risk for obesity on language development via EFNA4 were not established, our findings suggest EPH-Ephrin signalling expression may be responsible for language development. Mediating mechanisms involving multiple proteins in the EPH-Ephrin signalling pathway warrant further investigation. This study demonstrates the potential of examining early-life proteomics in conjunction with deep genotyping and phenotyping. Comprehensive early-life proteomic profiling is needed to enhance the understanding of the link between cardiometabolic health and neurodevelopment at the molecular level. Other studies in similar cohorts across other sociodemographic contexts are needed to support and expand upon our findings.

Contributors

Jian Huang contributed to conceptualisation, data curation, formal analysis, investigation, methodology, verification of the underlying data, data interpretation, visualisation, drafting, reviewing and editing the manuscript.

Jinyi Che contributed to data curation, formal analysis, investigation, verification of the underlying data, data interpretation, visualisation, reviewing and editing the manuscript.

Michelle Z.L Kee contributed to data interpretation, reviewing and editing the manuscript.

Ai Peng Tan contributed to data curation, data interpretation, reviewing and editing the manuscript.

Evelyn Law contributed to data curation, data interpretation, funding acquisition, and reviewing and editing the manuscript.

Patricia Pelufo Silveira contributed to data interpretation, and reviewing and editing the manuscript.

Irina Pokhvisneva and Sachin Patel contributed data analysis, data interpretation, and reviewing and editing the manuscript.

Keith M Godfrey contributed to reviewing and editing the manuscript.

Lourdes Mary Daniel, Kok Hian Tan, Yap Seng Chong contributed to funding acquisition.

Shiao-Yng Chan, and Johan G. Eriksson contributed to funding acquisition, and reviewing and editing the manuscript.

Dennis Wang contributed to methodology, data interpretation, and reviewing and editing the manuscript.

Jonathan Huang contributed to methodology, data interpretation, and reviewing and editing the manuscript.

All authors reviewed and approved the final version of the manuscript. Jian Huang and Jinyi Che have accessed and verified the underlying data.

The GUSTO study group includes: Airu Chia, Andrea Cremaschi, Anna Magdalena Fogel, Anne Eng Neo Goh, Anne Rifkin-Graboi, Anqi Qiu, Arijit Biswas, Bee Wah Lee, Birit Froukje Philipp Broekman, Candida Vaz, Chai Kiat Chng, Chan Shi Yu, Choon Looi Bong, Daniel Yam Thiam Goh, Dawn Xin Ping Koh, Dennis Wang, Desiree Y. Phua, E Shyong Tai, Elaine Kwang Hsia Tham, Elaine Phaik Ling Quah, Elizabeth Huiwen Tham, Evelyn Chung Ning Law, Evelyn Keet Wai Lau, Evelyn Xiu Ling Loo, Fabian Kok Peng Yap, Falk Müller-Riemenschneider, Franzolini Beatrice, George Seow Heong Yeo, Gerard Chung Siew Keong, Hannah Ee Juen Yong, Helen Yu Chen, Hong Pan, Huang Jian, Huang Pei, Hugo P S van Bever, Hui Min Tan, Iliana Magiati, Inez Bik Yun Wong, Ives Lim Yubin, Ivy Yee-Man Lau, Jacqueline Chin Siew Roong, Jadegoud Yaligar, Jerry Kok Yen Chan, Jia Xu, Johan Gunnar Eriksson, Jonathan Tze Liang Choo, Jonathan Y. Bernard, Jonathan Yinhao Huang, Joshua J. Gooley, Jun Shi Lai, Karen Mei Ling Tan, Keith M. Godfrey, Keri McCrickerd, Kok Hian Tan, Kothandaraman Narasimhan, Krishnamoorthy Naiduvaje, Kuan Jin Lee, Li Chen, Lieng Hsi Ling, Lin Lin Su, Ling-Wei Chen, Lourdes Mary Daniel, Lynette Pei-Chi Shek, Maria De Iorio, Marielle V. Fortier, Mary Foong-Fong Chong, Mary Wlodek, Mei Chien Chua, Melvin Khee-Shing Leow, Michael J. Meaney, Michelle Zhi Ling Kee, Min Gong, Mya Thway Tint, Navin Michael, Neerja Karnani, Ngee Lek, Noor Hidayatul Aini Bte Suaini, Oon Hoe Teoh, Peter David Gluckman, Priti Mishra, Queenie Ling Jun Li, Sambasivam Sendhil Velan, Seang Mei Saw, See Ling Loy, Seng Bin Ang, Shang Chee Chong, Shiao-Yng Chan, Shirong Cai, Shu-E Soh, Stephen Chin-Ying Hsu, Suresh Anand Sadananthan, Swee Chye Quek, Tan Ai Peng, Varsha Gupta, Victor Samuel Rajadurai, Wee Meng Han, Wei Wei Pang, Yap Seng Chong, Yin Bun Cheung, Yiong Huak Chan, Yung Seng Lee, Zhang Han.

Data sharing statement

GWAS summary statistics are publicly available (see Supplementary Table S1 for details). Deidentified participant data from the GUSTO cohort will be made available on request. Data requests and research proposals can be submitted following the guidelines here: https://gustodatavault.sg/about/request-for-data.

Declaration of interests

KMG and SC are part of an academic consortium that has received research funding from Société Des Produits Nestlé S.A., and are co-inventors on patent filings by Nestlé S.A. outside the submitted work. KMG has received reimbursement for speaking at conferences sponsored by companies selling nutritional products. SC has received reimbursement from the Expert Group on Inositol in Basic and Clinical Research (EGOI; a not-for-profit academic organization) and Nestlé Nutrition Institute for speaking at conferences. The other authors declare no competing interests.

Acknowledgements

The study is supported by the National Research Foundation (NRF) under the Open Fund-Large Collaborative Grant (OF-LCG; MOH-000504) administered by the Singapore Ministry of Health's National Medical Research Council (NMRC) and the Agency for Science, Technology and Research (A*STAR). In RIE2025, the study is supported by funding from the NRF's Human Health and Potential (HHP) Domain, under the Human Potential Programme. JH is supported by the NMRC Open Fund - Young Individual Research Grant (MOH-001148) and the A*STAR Human Health and Potential–Prenatal/Early Childhood Grant (H24P2M0002). APT is supported by the NMRC Transition Award (MOH-001273-00).

KMG is supported by the National Institute for Health and Care Research (NIHR Senior Investigator (NF-SI-0515-10042) and NIHR Southampton Biomedical Research Centre (NIHR203319)) and Alzheimer's Research UK (ARUK-PG2022A-008).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2025.105579.

References

1 Eising E, Mirza-Schreiber N, de Zeeuw EL, et al. Genome-wide analyses of individual differences in quantitatively assessed reading- and language-related skills in up to 34,000 people. *Proc Natl Acad Sci U S A*. 2022;119(35):e2202764119.
2 Gibson JL, Newbury DF, Durkin K, Pickles A, Conti-Ramsden G, Toseeb U. Pathways from the early language and communication environment to literacy outcomes at the end of primary school; the roles of language development and social development. *Oxf Rev Educ*. 2021;47(2):260–283.
3 Romeo RR, Flournoy JC, McLaughlin KA, Lengua LJ. Language development as a mechanism linking socioeconomic status to executive functioning development in preschool. *Dev Sci*. 2022;25(5):e13227.
4 Schoon I, Parsons S, Rush R, Law J. Children's language ability and psychosocial development: a 29-year follow-up study. *Pediatrics*. 2010;126(1):e73–e80.
5 Snowdon DA, Greiner LH, Markesbery WR. Linguistic ability in early life and the neuropathology of Alzheimer's disease and cerebrovascular disease. Findings from the Nun Study. *Ann N Y Acad Sci*. 2000;903:34–38.
6 Snowdon DA, Kemper SJ, Mortimer JA, Greiner LH, Wekstein DR, Markesbery WR. Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study. *JAMA*. 1996;275(7):528–532.
7 Pujol J, Soriano-Mas C, Ortiz H, Sebastián-Gallés N, Losilla JM, Deus J. Myelination of language-related areas in the developing brain. *Neurology*. 2006;66(3):339–343.
8 Forbes TA, Gallo V. All wrapped up: environmental effects on myelination. *Trends Neurosci*. 2017;40(9):572–587.
9 Liang J, Matheson BE, Kaye WH, Boutelle KN. Neurocognitive correlates of obesity and obesity-related behaviors in children and adolescents. *Int J Obes*. 2014;38(4):494–506.
10 Black N, Johnston DW, Peeters A. Childhood obesity and cognitive achievement. *Health Econ*. 2015;24(9):1082–1100.
11 Li N, Yolton K, Lanphear BP, Chen A, Kalkwarf HJ, Braun JM. Impact of early-life weight status on cognitive abilities in children. *Obesity*. 2018;26(6):1088–1095.
12 Hemmingsson E. Early childhood obesity risk factors: socioeconomic adversity, family dysfunction, offspring distress, and junk food self-medication. *Curr Obes Rep*. 2018;7(2):204–209.
13 Noble KG, Houston SM, Brito NH, et al. Family income, parental education and brain structure in children and adolescents. *Nat Neurosci*. 2015;18(5):773–778.
14 Ojalehto E, Zhan Y, Jylhävä J, Reynolds CA, Dahl Aslan AK, Karlsson IK. Genetically and environmentally predicted obesity in relation to cardiovascular disease: a nationwide cohort study. *eClinicalMedicine*. 2023;58:101943.
15 Ruan Y, Lin YF, Feng YCA, et al. Improving polygenic prediction in ancestrally diverse populations. *Nat Genet*. 2022;54(5):573–580.
16 Hari Dass SA, McCracken K, Pokhvisneva I, et al. A biologically-informed polygenic score identifies endophenotypes and clinical conditions associated with the insulin receptor function on specific brain regions. *eBioMedicine*. 2019;42:188–202.
17 Vogelezang S, Bradfield JP, Ahluwalia TS, et al. Novel loci for childhood body mass index and shared heritability with adult cardiometabolic traits. *PLoS Genet*. 2020;16(10):e1008718.
18 Bradfield JP, Vogelezang S, Felix JF, et al. A trans-ancestral meta-analysis of genome-wide association studies reveals loci associated with childhood obesity. *Hum Mol Genet*. 2019;28(19):3327–3338.
19 Richardson TG, Sanderson E, Elsworth B, Tilling K, Davey Smith G. Use of genetic variation to separate the effects of early and later life adiposity on disease risk: mendelian randomisation study. *BMJ*. 2020;369:m1203.
20 Sakaue S, Kanai M, Tanigawa Y, et al. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet*. 2021;53(10):1415–1424.

21 Yengo L, Sidorenko J, Kemper KE, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet*. 2018;27(20):3641–3649.
22 Soh SE, Tint MT, Gluckman PD, et al. Cohort profile: Growing up in Singapore Towards healthy Outcomes (GUSTO) birth cohort study. *Int J Epidemiol*. 2014;43(5):1401–1409.
23 Teh AL, Pan H, Chen L, et al. The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. *Genome Res*. 2014;24(7):1064–1074.
24 Dunn LM, Dunn DM. Peabody picture vocabulary test–fourth edition [cited 2024 Jan 7]. Available from: http://doi.apa.org/getdoi.cfm?doi=10.1037/t15144-000; 2012.
25 Burns TG. Wechsler individual achievement test-III: what is the "gold standard" for measuring academic achievement? *Appl Neuropsychol*. 2010;17(3):234–236.
26 Assarsson E, Lundberg M, Holmquist G, et al. Homogenous 96-Plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One*. 2014;9(4):e95192.
27 Fields RD. White matter in learning, cognition and psychiatric disorders. *Trends Neurosci*. 2008;31(7):361–370.
28 Friederici AD. The brain basis of language processing: from structure to function. *Physiol Rev*. 2011;91(4):1357–1392.
29 Price CJ. The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann N Y Acad Sci*. 2010;1191:62–88.
30 Friederici AD. Towards a neural basis of auditory sentence processing. *Trends Cogn Sci*. 2002;6(2):78–84.
31 Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291–295.
32 Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for Colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*. 2014;10(5):e1004383.
33 VanderWeele TJ. Principles of confounder selection. *Eur J Epidemiol*. 2019;34(3):211–219.
34 Choi SW, O'Reilly PF. PRSice-2: polygenic risk score software for biobank-scale data. *Gigascience*. 2019;8(7):giz082.
35 Ballinger GA. Using generalized estimating equations for longitudinal data analysis. *Organ Res Methods*. 2004;7(2):127–150.
36 Sanderson E, Glymour MM, Holmes MV, et al. Mendelian randomization. *Nat Rev Methods Primers*. 2022;2(1):6.
37 Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet*. 2017;13(11):e1007081.
38 Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol*. 2016;40(4):304–314.
39 Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. 2015;44(2):512–525.
40 Gkatzionis A, Burgess S, Conti DV, Newcombe PJ. Bayesian variable selection with a pleiotropic loss function in Mendelian randomization. *Stat Med*. 2021;40(23):5025–5045.
41 Bowden J, Spiller W, Del Greco MF, et al. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the radial plot and radial regression. *Int J Epidemiol*. 2018;47(4):1264–1278.
42 Skrivankova VW, Richmond RC, Woolf BAR, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomisation (STROBE-MR): explanation and elaboration. *BMJ*. 2021;375:n2233.
43 Szklarczyk D, Kirsch R, Koutrouli M, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*. 2023;51(D1):D638–D646.
44 Dunn R, Dudbridge F, Sanderson CM. The use of edge-betweenness clustering to investigate biological function in protein interaction networks. *BMC Bioinformatics*. 2005;6(1):39.
45 Bobb JF, Claus Henn B, Valeri L, Coull BA. Statistical software for analyzing the health effects of multiple concurrent exposures via Bayesian kernel machine regression. *Environ Health*. 2018;17(1):67.
46 Huang J, Tan AP, Law E, et al. Maternal preconception circulating blood biomarker mixtures, child behavioural symptom scores and the potential mediating role of neonatal brain microstructure: the S-PRESTO cohort. *Transl Psychiatry*. 2023;13(1):38.
47 Hamra G, MacLehose R, Richardson D. Markov chain Monte Carlo: an introduction for epidemiologists. *Int J Epidemiol*. 2013;42(2):627–634.
48 VanderWeele TJ. A unification of mediation and interaction: a 4-way decomposition. *Epidemiology*. 2014;25(5):749–761.
49 Corrigan NM, Yarnykh VL, Huber E, Zhao TC, Kuhl PK. Brain myelination at 7 months of age predicts later language development. *Neuroimage*. 2022;263:119641.
50 Miguel PM, Pereira LO, Barth B, et al. Prefrontal cortex dopamine transporter gene network moderates the effect of perinatal hypoxic-ischemic conditions on cognitive flexibility and brain gray matter density in children. *Biol Psychiatry*. 2019;86(8):621–630.
51 GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509):1318–1330.
52 Chen LM, Yao N, Garg E, et al. PRS-on-Spark (PRSoS): a novel, efficient and flexible approach for generating polygenic risk scores. *BMC Bioinformatics*. 2018;19(1):295.
53 Chen R, Yang X, Zhang B, et al. EphA4 negatively regulates myelination by inhibiting Schwann cell differentiation in the peripheral nervous system. *Front Neurosci*. 2019;13:1191.
54 Klein R. Bidirectional modulation of synaptic functions by Eph/ephrin signaling. *Nat Neurosci*. 2009;12(1):15–20.
55 Shu Y, Xiao B, Wu Q, et al. The Ephrin-A5/EphA4 interaction modulates neurogenesis and angiogenesis by the p-Akt and p-ERK pathways in a mouse model of TLE. *Mol Neurobiol*. 2016;53(1):561–576.
56 Fibla L, Forbes SH, McCarthy J, et al. Language exposure and brain myelination in early development. *J Neurosci*. 2023;43(23):4279–4290.
57 Del Tufo SN, Earle FS, Cutting LE. The impact of expressive language development and the left inferior longitudinal fasciculus on listening and reading comprehension. *J Neurodevelop Disord*. 2019;11(1):37.
58 Altieri R, Melcarne A, Junemann C, et al. Inferior Fronto-Occipital fascicle anatomy in brain tumor surgeries: from anatomy lab to surgical theater. *J Clin Neurosci*. 2019;68:290–294.
59 Tranchevent LC, Halder R, Glaab E. Systems level analysis of sex-dependent gene expression changes in Parkinson's disease. *NPJ Parkinsons Dis*. 2023;9(1):8.
60 Chatterjee M, Schild D, Teunissen CE. Contactins in the central nervous system: role in health and disease. *Neural Regen Res*. 2019;14(2):206–216.
61 Hart CG, Karimi-Abdolrezaee S. Bone morphogenetic proteins: new insights into their roles and mechanisms in CNS development, pathology and repair. *Exp Neurol*. 2020;334:113455.
62 Oh HSH, Rutledge J, Nachun D, et al. Organ aging signatures in the plasma proteome track health and disease. *Nature*. 2023;624(7990):164–172.
63 You J, Guo Y, Zhang Y, et al. Plasma proteomic profiles predict individual future health risk. *Nat Commun*. 2023;14(1):7817.
64 Harris SE, Cox SR, Bell S, et al. Neurology-related protein biomarkers are associated with cognitive ability and brain volume in older age. *Nat Commun*. 2020;11(1):800.
65 Gudgeon J, Marín-Rubio JL, Trost M. The role of macrophage scavenger receptor 1 (MSR1) in inflammatory disorders and cancer. *Front Immunol*. 2022;13:1012002.
66 Chu EM, Tai DC, Beer JL, Hill JS. Macrophage heterogeneity and cholesterol homeostasis: classically-activated macrophages are associated with reduced cholesterol accumulation following treatment with oxidized LDL. *Biochim Biophys Acta*. 2013;1831(2):378–386.
67 Gamba P, Testa G, Gargiulo S, Staurenghi E, Poli G, Leonarduzzi G. Oxidized cholesterol as the driving force behind the development of Alzheimer's disease. *Front Aging Neurosci*. 2015;7:119.
68 Tin A, Fohner AE, Yang Q, et al. Identification of circulating proteins associated with general cognitive function among middle-aged and older adults. *Commun Biol*. 2023;6(1):1117.
69 Malik-Moraleda S, Ayyash D, Gallée J, et al. An investigation across 45 languages and 12 language families reveals a universal language network. *Nat Neurosci*. 2022;25(8):1014–1019.
70 Adams C, Lockton E, Freed J, et al. The social communication intervention project: a randomized controlled trial of the effectiveness of speech and language therapy for school-age children who have pragmatic and social communication problems with or without autism spectrum disorder. *Int J Lang Commun Disord*. 2012;47(3):233–244.