

# Genome-wide association meta-analysis of expressive and receptive vocabulary from infancy to early childhood

E. Verhoef<sup>1</sup>, T.S. Ahluwalia<sup>2,3</sup>, P. Jansen<sup>4,5</sup>, K. Lange<sup>6,7</sup>, A. Allegrini<sup>8</sup>, C. Wang<sup>9</sup>, C. Symeonides<sup>6,7,10</sup>, A. Morgan<sup>6,7,10</sup>, EAGLE working group, S.E. Fisher<sup>1,11</sup>, B. St Pourcain<sup>1,11,12</sup>

1. Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands; 2. COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark; 3. Steno Diabetes Center Copenhagen, Gentofte, Denmark; 4. Department of Complex Trait Genetics, Center for Neurogenetics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam, Amsterdam, the Netherlands; 5. Department of Child and Adolescent Psychiatry, Erasmus University Medical Center, Rotterdam, the Netherlands; 6. Murdoch Children's Research Institute, Parkville, VIC, Australia; 7. Department of Paediatrics & Department of Audiology and Speech Pathology, University of Melbourne, Parkville, VIC, Australia; 8. Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; 9. School of Medicine and Public Health, The University of Newcastle, Newcastle, NSW, 2308, Australia; 10. Royal Children's Hospital, Melbourne, Victoria, Australia; 11. Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands; 12. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

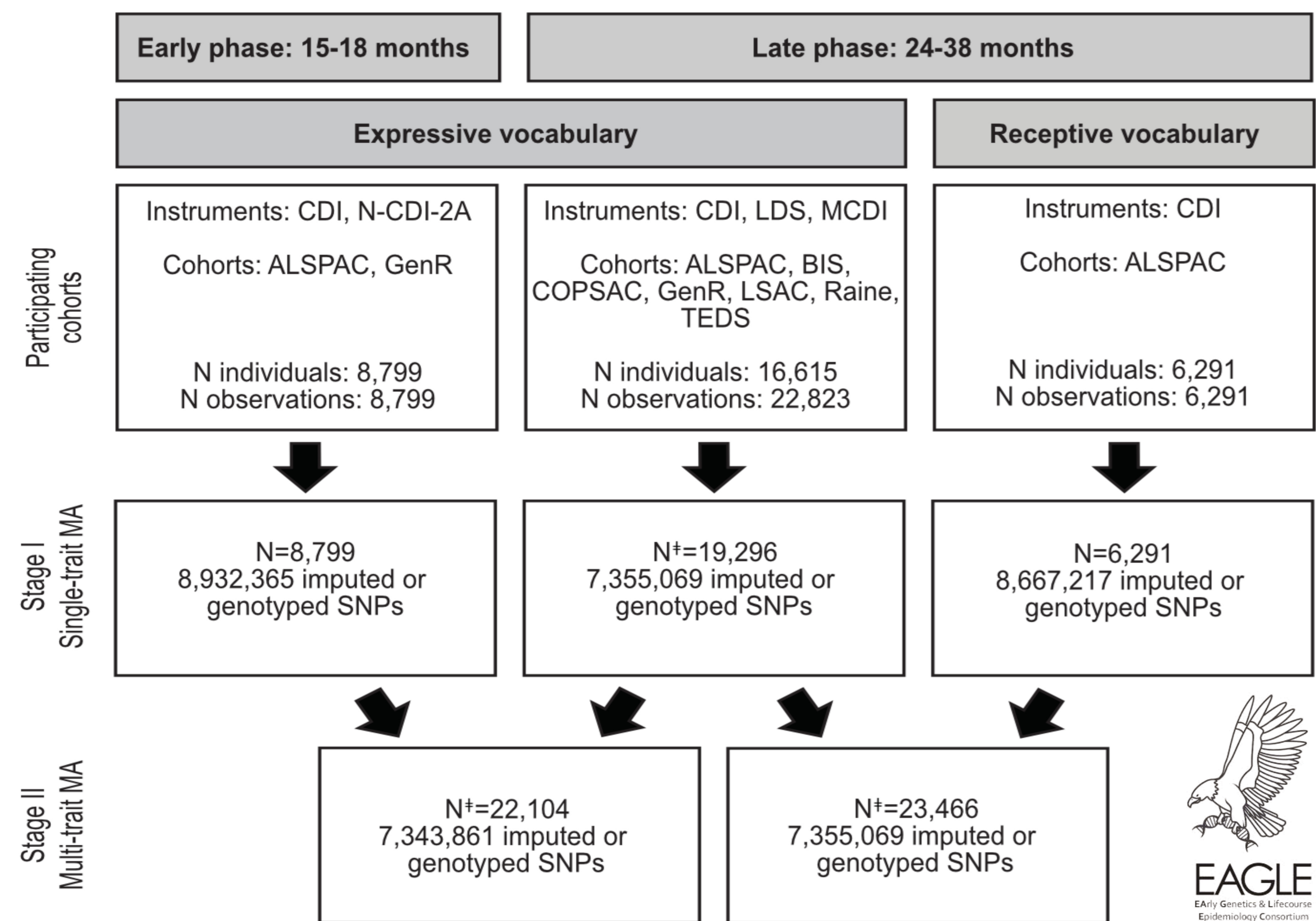
## In a nutshell - we aimed to ...

(i) identify genome-wide association signals for early vocabulary, (ii) elucidate the underlying genetic architecture and (iii) characterise polygenic links with later-life cognitive, behavioral and health outcomes.

## Background

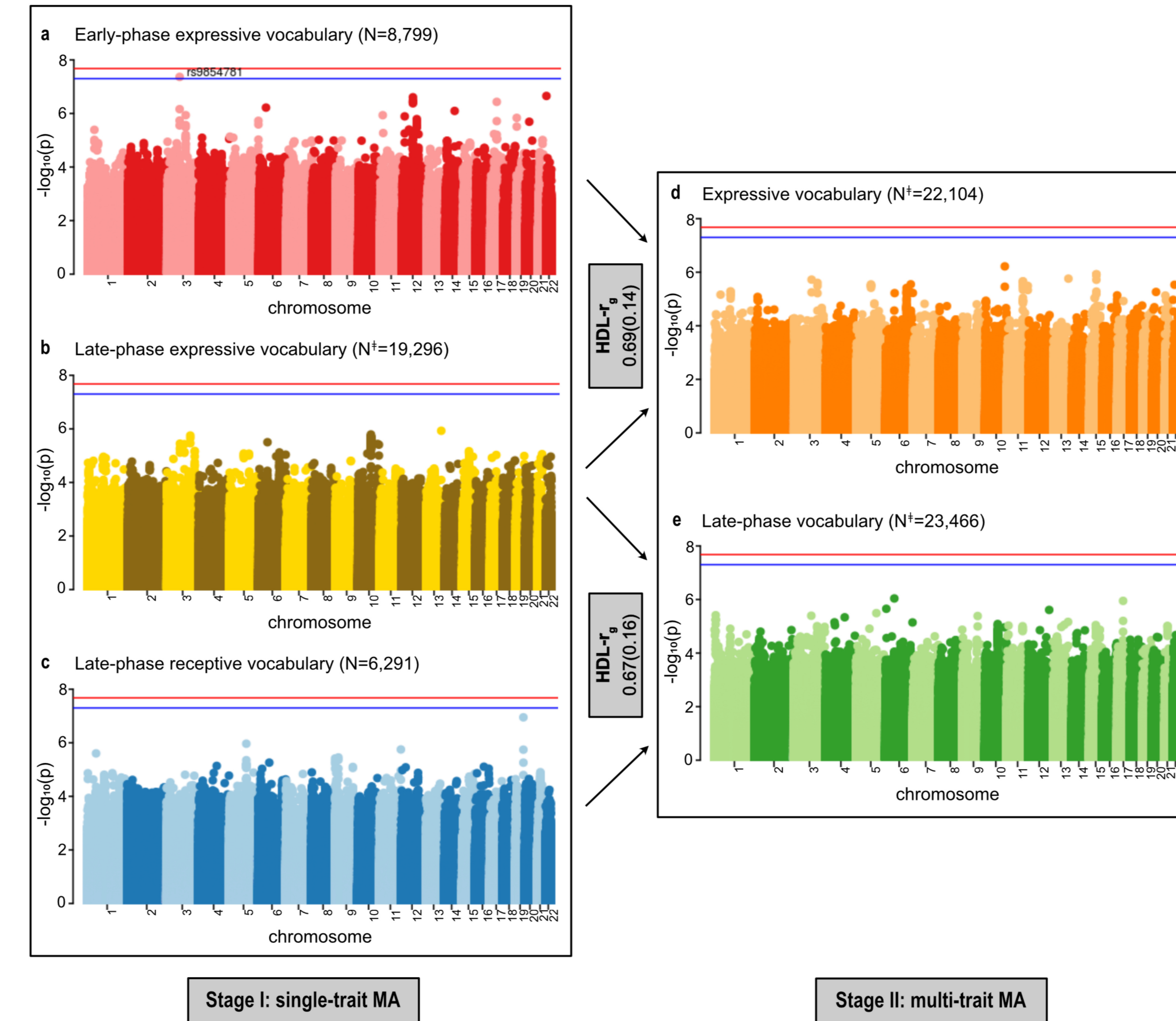
The genetic architectures underlying expressive and receptive vocabulary in early life are complex and little characterised. Here, we present the largest meta-genome-wide association study (meta-GWAS) of early vocabulary (15-38 months) till date, based on 37,913 observations and 17,298 individuals, within the Early Genetics and Life Course Epidemiology (EAGLE) Consortium.

## Study design



Expressive and/or receptive vocabulary was assessed using parent questionnaires in children of European descent from seven independent population-based or community-based cohorts. Meta-GWASs were performed for early-phase expressive vocabulary, late-phase expressive vocabulary, and late-phase receptive vocabulary using METAL as part of stage I. To increase statistical power, multi-trait SNP-vocabulary associations were examined by combining genetically related vocabulary traits using MTAG as part of stage II. † Estimated sample size based on the increase in mean  $\chi^2$  statistic using MTAG. Abbreviations: CDI, Communicative Development Inventory; LDS, Language Development Survey; MA, meta-analysis; SNP, Single-Nucleotide Polymorphism

## SNP-vocabulary associations



Single-trait (stage I) and multi-trait (stage II) SNP-vocabulary associations were assessed using METAL and MTAG, respectively. No SNP-vocabulary associations passed the adjusted threshold (red line,  $P < 2.1 \times 10^{-8}$ ). rs9854781 was associated with early-phase expressive vocabulary at the unadjusted genome-wide significance threshold (blue line,  $P < 5 \times 10^{-8}$ ). Genetic correlations among single-trait vocabulary summary statistics were derived with HDL. † Estimated sample size based on the increase in mean  $\chi^2$  statistic using MTAG. Abbreviations: HDL, High-Density Likelihood; MA, meta-analysis.

## SNP-heritability of early vocabulary

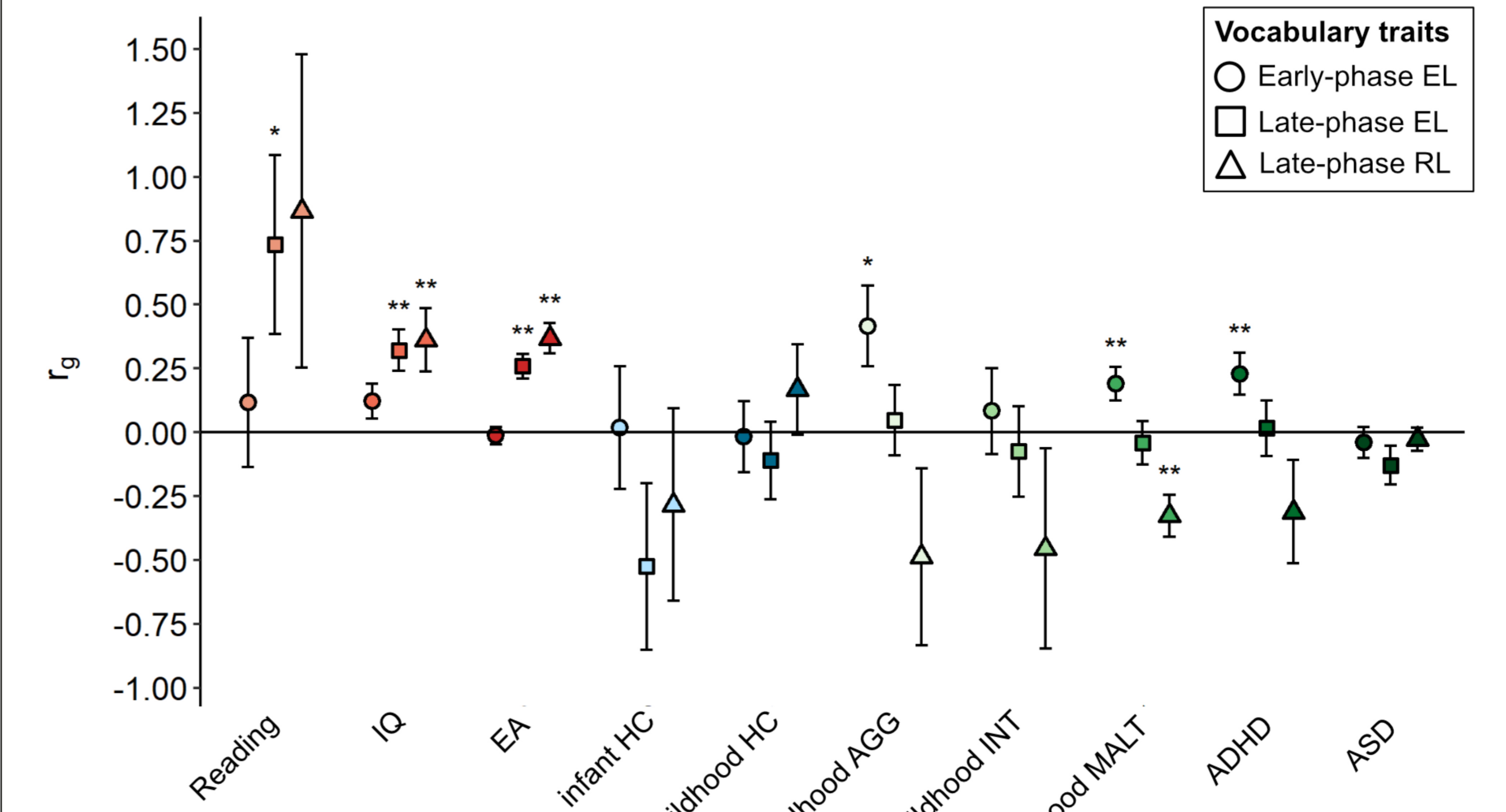
Analysis	Trait	SNP-h <sup>2</sup> (SE)	Z-score	P	N
Stage I: single-trait MA	Early-phase EV	0.24(0.02)	8.64	$< 1 \times 10^{-10}$	8,799
	Late-phase EV	0.08(0.01)	5.53	$3 \times 10^{-8}$	19,296†
	Late-phase RV	0.20(0.04)	5.21	$2 \times 10^{-7}$	6,291
Stage II: multi-trait MA	EV	0.10(0.01)	6.91	$< 1 \times 10^{-10}$	22,104†
	Late-phase vocabulary	0.07(0.01)	5.00	$5 \times 10^{-7}$	23,466†

SNP-heritability (SNP-h<sup>2</sup>), corresponding standard error (SE) and P-value were estimated with High-Density Likelihood (HDL) software using a HapMap3 reference panel. SNP-h<sup>2</sup> Z-scores were calculated by dividing SNP-h<sup>2</sup> by its standard error. † Estimated sample size based on the increase in mean  $\chi^2$  statistic using MTAG.

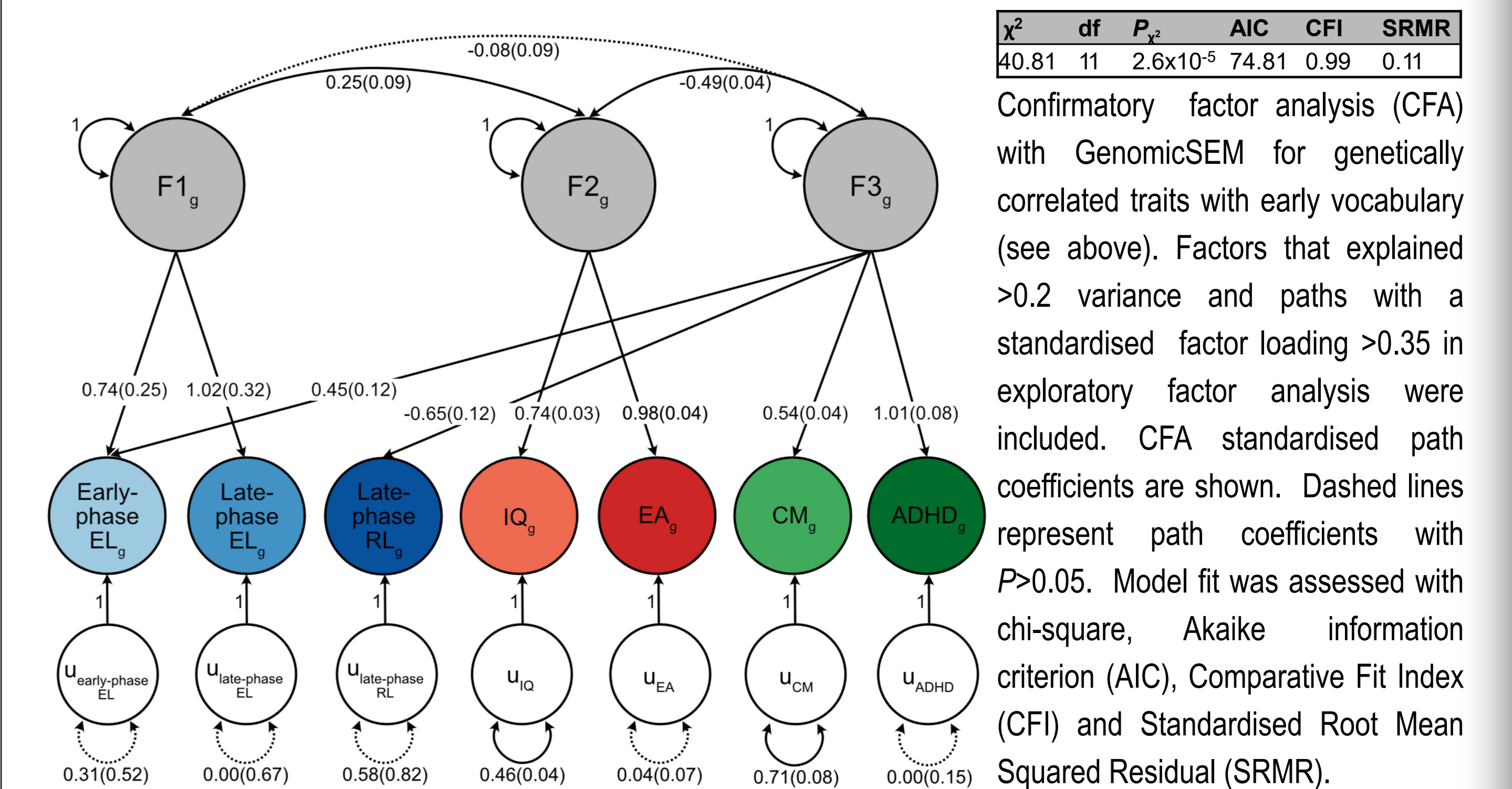
## Gene-, gene-set and gene-property analyses

All analyses were carried out within FUMA software. No results passed multiple-testing-adjusted thresholds.

## Genetic links with later-life cognitive, behavioral and health outcomes



Genetic correlations of vocabulary traits with later-life cognitive, behavioral and health outcomes estimated using HDL software. Bars represent standard errors. \*  $P < 0.05$ ; \*\*  $P < 1.1 \times 10^{-6}$ . Abbreviations: ADHD, Attention-Deficit/Hyperactivity Disorder; AGG, aggression; ASD, Autism Spectrum Disorder; EA, educational attainment; EL, expressive language; HC, head circumference; INT, internalising symptoms; IQ, general intelligence; MALT, maltreatment; RL, receptive language



## In a nutshell - we found that ...

early vocabulary traits show low to moderate SNP-h<sup>2</sup>. Their genetic architectures may change during early development: genetic links with cognition trace back to toddlerhood and not infancy, and childhood maltreatment has an opposite association with early-phase expressive compared to late-phase receptive vocabulary.

