MAX PLANCK INSTITUTE FOR PSYCHOLINGUISTICS

PGHC Population Genetics of Human Communication

Genetic influences underlying early language development and their links with later-life traits

Dr. Ellen Verhoef

Population Genetics of Human Communication research group

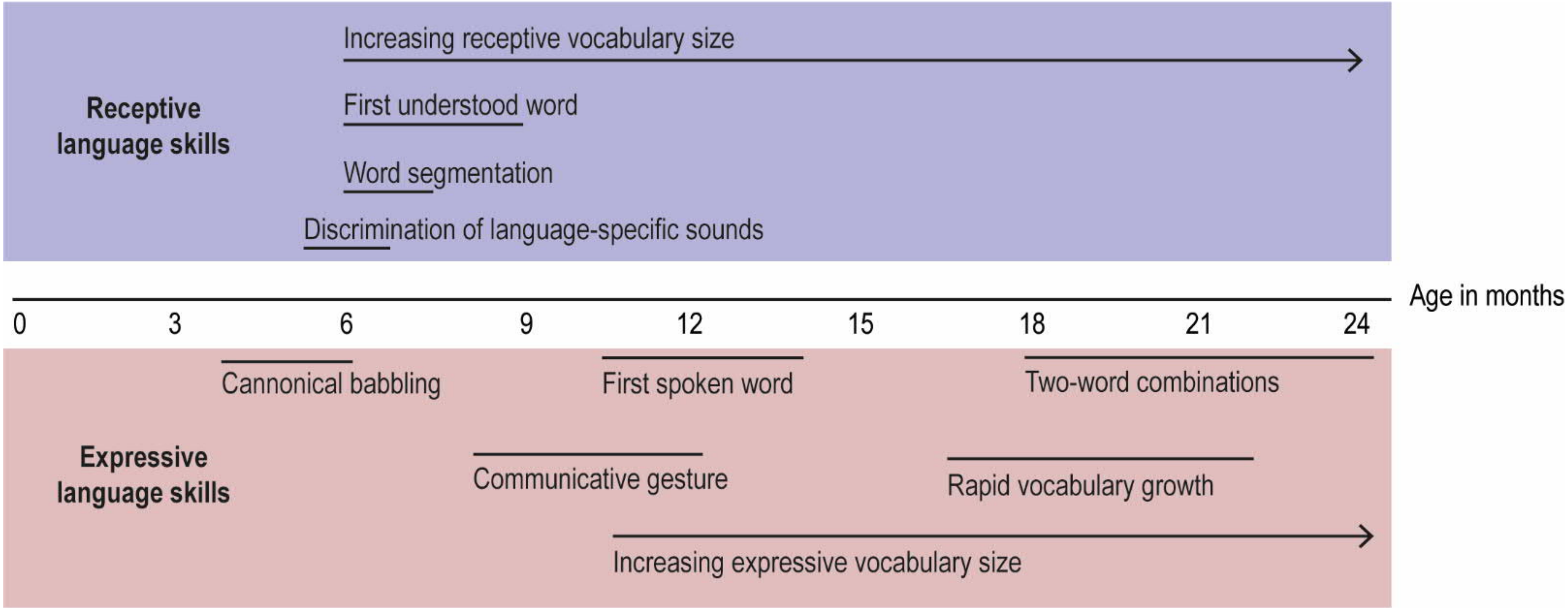
Language and Genetics department

Max Planck Institute for Psycholinguistics

MPI staff meeting

19 September 2023

Language development in infants and toddlers



Vocabulary size in young children

- Parental questionnaires
  - MacArthur Communicative Development Inventories
  - Language Development Survey

- Age-specific word lists (~100 - 700 words)

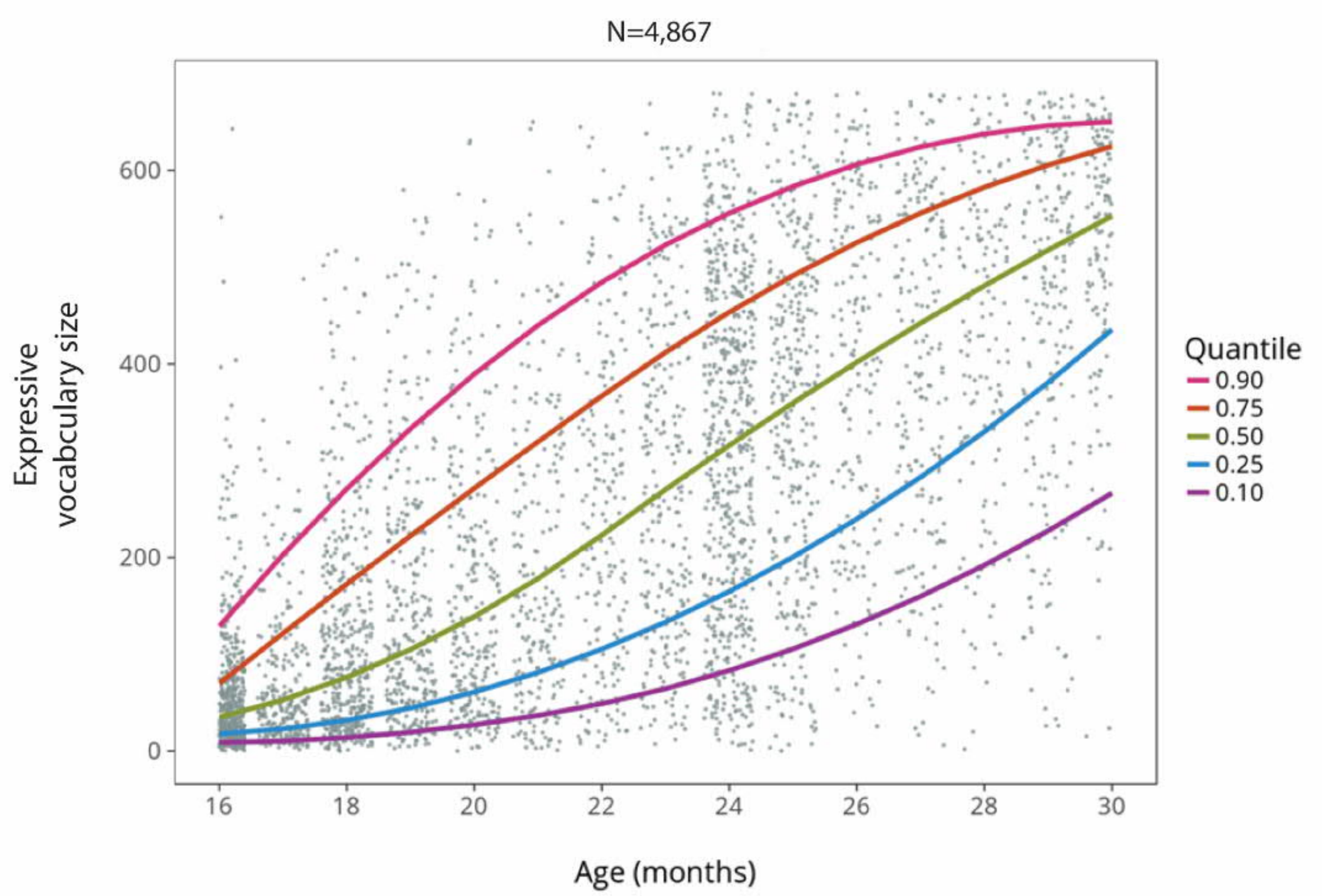
Item	Say	Understand
Ball	x	
Tree		x
Dog	x	x

- Relatively easy to assess → large sample sizes

Fenson et al. Singular Publishing. 1993
Rescorla et al. Journal of Speech and Hearing Disorders. 1989

MAX PLANCK

Individual differences in vocabulary size



Cross-sectional MacArthur-Bates Communicative Development Inventory expressive vocabulary data (N=4,687) based on English-speaking children between 16-30 months of age. Data were downloaded from the Wordbank.·

Wordbank: an open repository for developmental vocabulary data. J Child Lang 44, 677–694 (2017)

MAX PLANCK

PGHC
Population Genetics of Human Communication

Individual differences in language development and our DNA

Are individual differences in early-life vocabulary size (partially) attributable to genetic variation?

Are the same genetic influences related to vocabulary size throughout early development?

Do early-life vocabulary and later cognition-related abilities share genetic influences?

Are genetic influences related to early-life vocabulary associated with childhood-onset neurodevelopmental conditions (NDCs)?

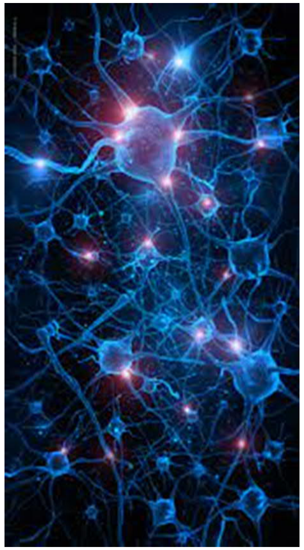
MAX PLANCK

DNA

Proteins

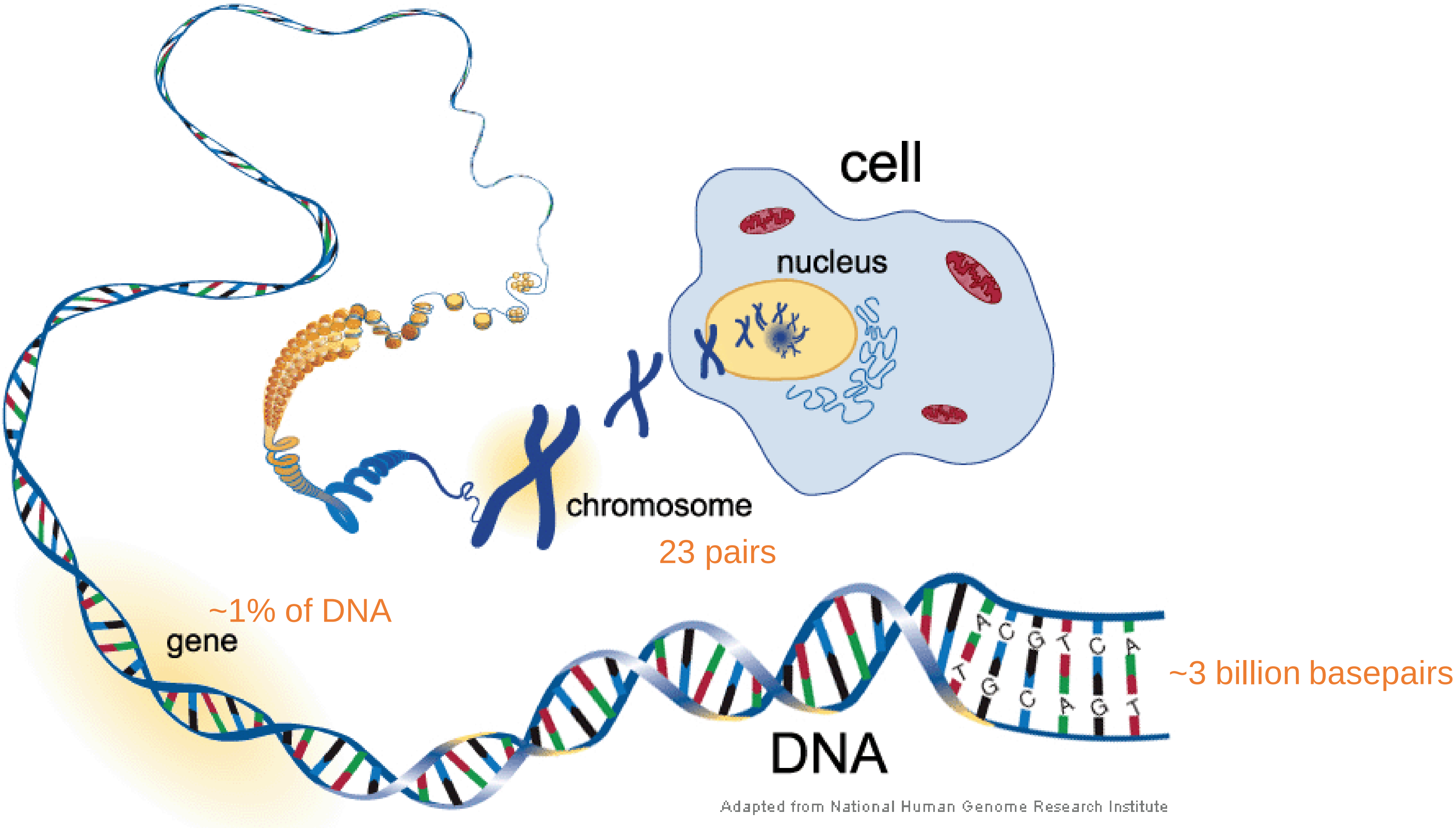
Cells

Cellular
networks



Brain

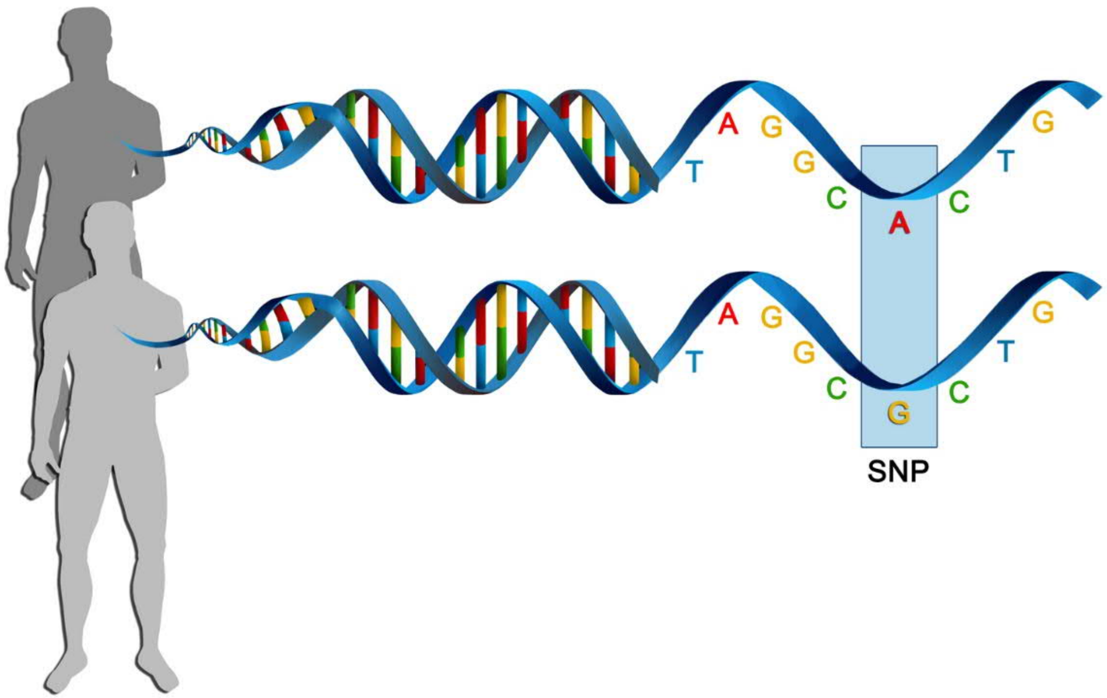
Language skills



Adapted from National Human Genome Research Institute

Single-Nucleotide Polymorphism (SNP)

- Same genetic code at majority of bases in our genome
- ~ 85 million sites were differences occur



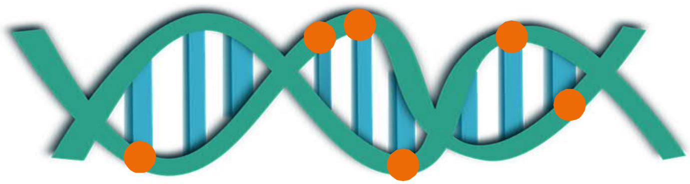
Size: one basepair
Frequency: ≥1%
Effect: very small

The 1000 Genomes Consortium, Nature, 2015

MAX PLANCK

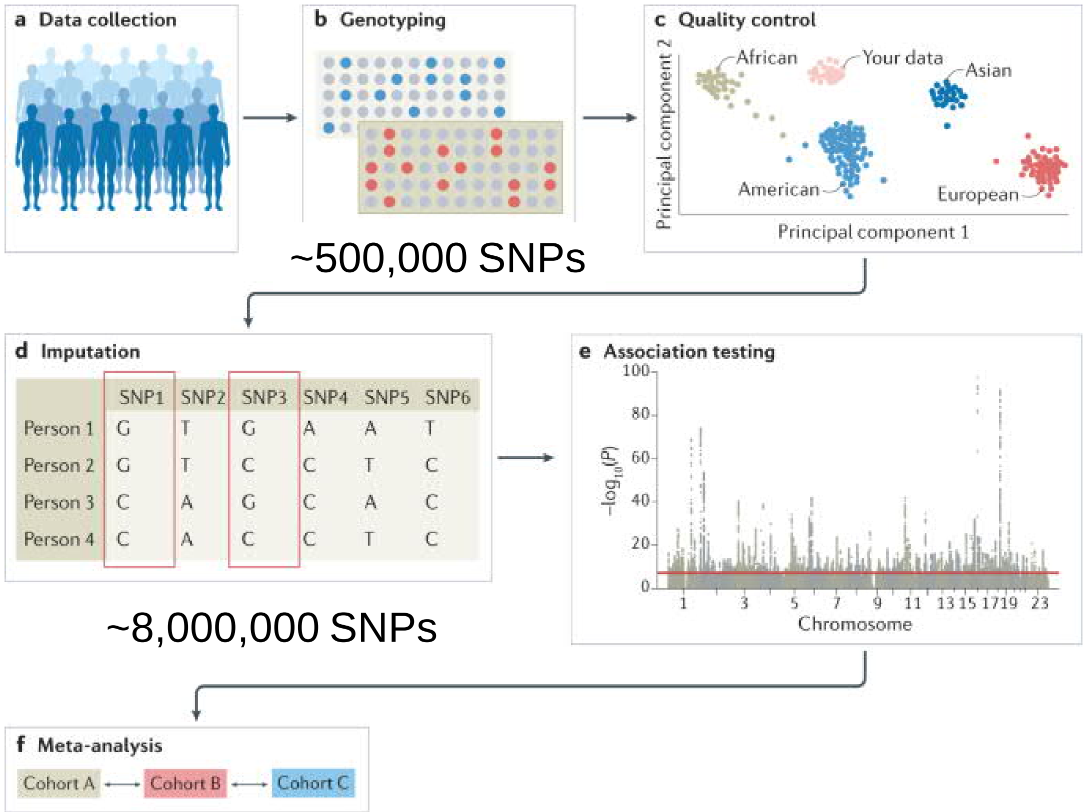
Are individual differences in early-life vocabulary size (partially) attributable to genetic variation?

Per SNP



Across all SNPs

Genome-wide association study (GWAS)



Linear regression
Phenotype ~ genotype

Output: summary statistics

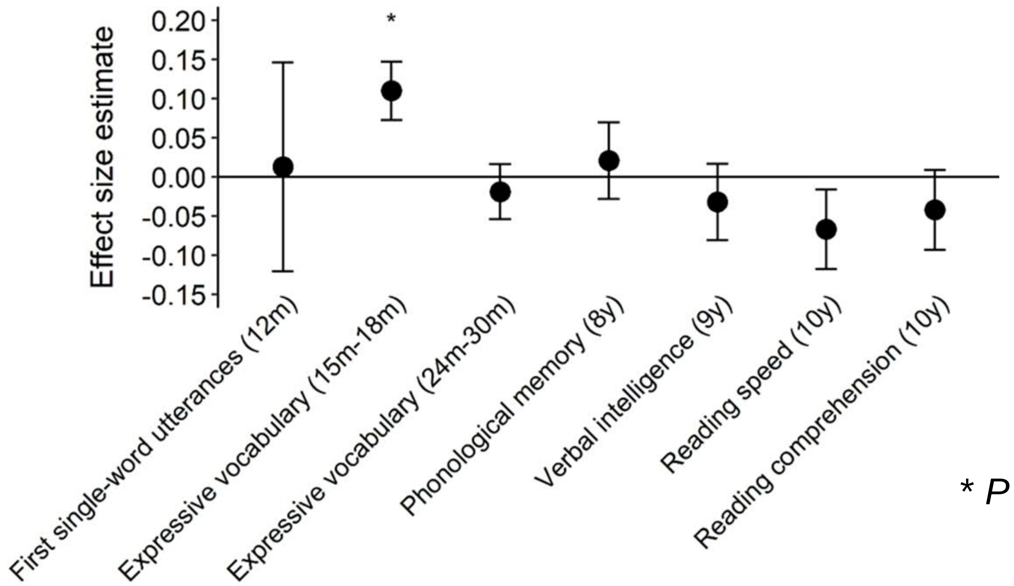
SNP	Beta(SE)	*P*
rs1	-0.11(0.022)	9.5x10^{-7}
rs2	0.081(0.020)	6.2x10^{-5}

Uffelmann et al. Nat. Rev. Methods. 2021

MAX PLANCK

SNP-vocabulary GWAS

rs7642482



* $P < 5x10^{-8}$

nature COMMUNICATIONS

ARTICLE

Received 15 Jan 2014 | Accepted 28 Jul 2014 | Published 16 Sep 2014 DOI: 10.1038/ncomms5831 OPEN

Common variation near *ROBO2* is associated with expressive vocabulary in infancy

Beate St Pourcain[1,2,3,*], Rolieke A.M. Cents[4,5,*], Andrew J.O. Whitehouse[6,*], Claire M.A. Haworth[7,8,*], Oliver S.P. Davis[8,9,*], Paul F. O'Reilly[8,10], Susan Roulstone[11], Yvonne Wren[11], Qi W. Ang[12], Fleur P. Velders[4,5], David M. Evans[1,13,14], John P. Kemp[1,13,14], Nicole M. Warrington[12,14], Laura Miller[13], Nicholas J. Timpson[1,13], Susan M. Ring[1,13], Frank C. Verhulst[5], Albert Hofman[15], Fernando Rivadeneira[15,16], Emma L. Meaburn[17], Thomas S. Price[18], Philip S. Dale[19], Demetris Pillas[10], Anneli Yliherva[20], Alina Rodriguez[10,21], Jean Golding[13], Vincent W.V. Jaddoe[4,15,22], Marjo-Riitta Jarvelin[10,23,24,25,26], Robert Plomin[8], Craig E. Pennell[12], Henning Tiemeier[5,15,*] & George Davey Smith[1,13]

St Pourcain et al. Nature Communications. 2014

SNP-vocabulary GWAS

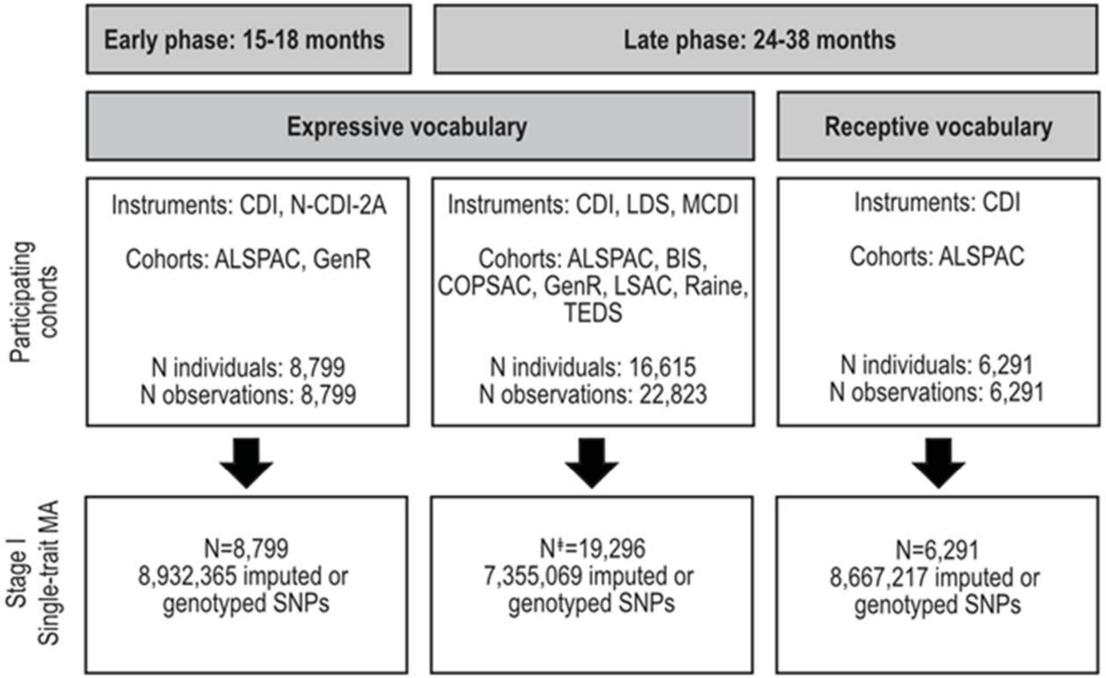
- ~50% sample size increase
- Multivariate approach to maximize statistical power
- More detailed study of lower frequency variants (0.5-1%)
- Inclusion of receptive vocabulary

EAGLE
EArly Genetics & Lifecourse Epidemiology Consortium

Verhoef et al. BioRxiv. 2022

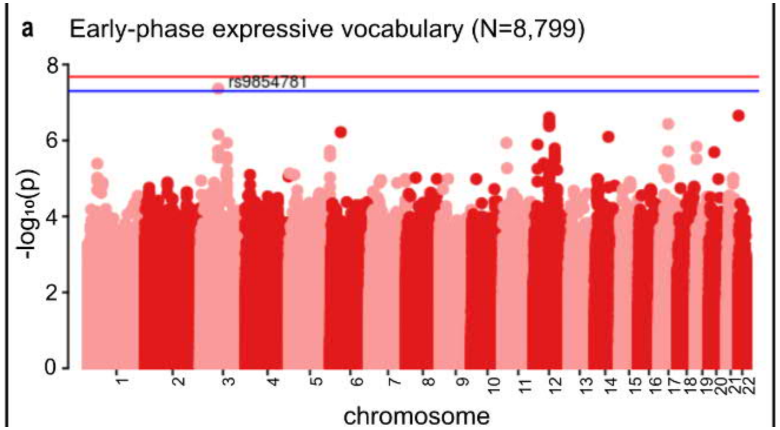
MAX PLANCK

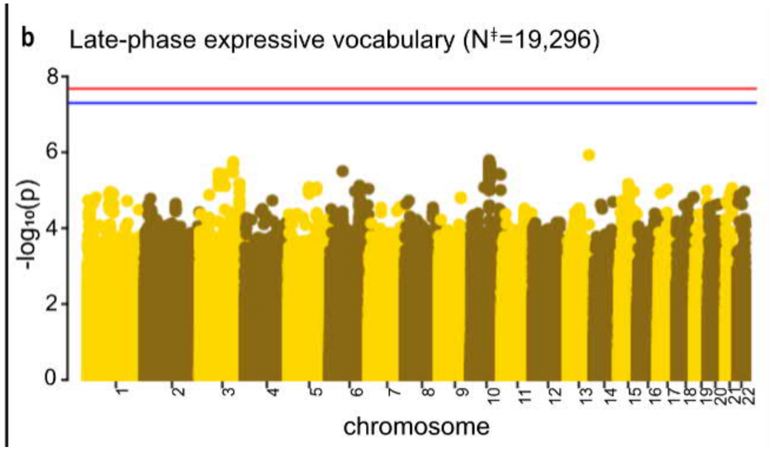
SNP-vocabulary GWAS

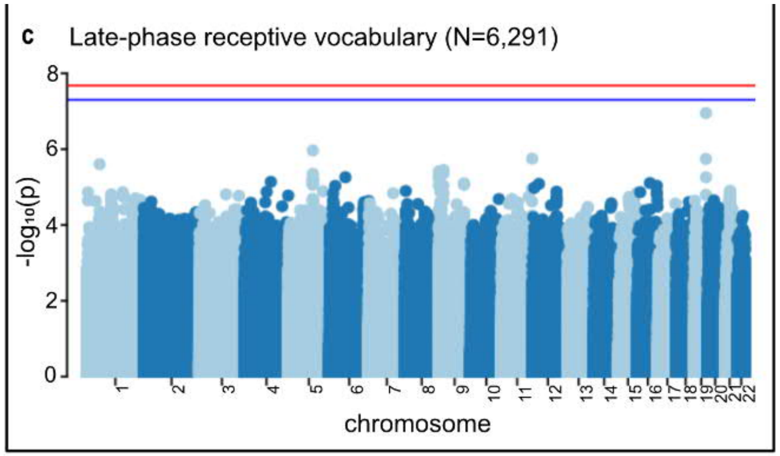


Verhoef et al. BioRxiv. 2022

SNP-vocabulary GWAS







$P < 2.1 \times 10^{-8}$

$P < 5 \times 10^{-8}$

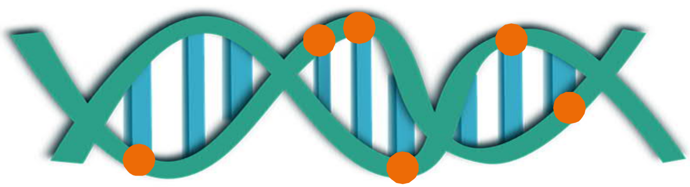
Verhoef et al. BioRxiv. 2022

MAX PLANCK

Are individual differences in early-life vocabulary size (partially) attributable to genetic variation?

Per SNP

Yes, but effects are very small



Across all SNPs

Heritability

The phenotypic variance (σ_P^2) explained by genetic variance (σ_G^2)

$$h^2 = \sigma_G^2 / \sigma_P^2$$

Heritability

The phenotypic variance (σ_P^2) explained by genetic variance (σ_G^2)

$$h^2 = \sigma_G^2 / \sigma_P^2$$

Twin heritability

The proportion of phenotypic variance explained by all genetic variance

Heritability

The phenotypic variance (σ_P^2) explained by genetic variance (σ_G^2)

$$h^2 = \sigma_G^2 / \sigma_P^2$$

Twin heritability
The proportion of phenotypic variance explained by <u>all</u> genetic variance

Single-Nucleotide Polymorphism (SNP) heritability
The proportion of phenotypic variance explained by <u>common</u> genetic influences

Variation in DNA present in >1% population
500,000 SNPs across genome

Heritability

The phenotypic variance (σ_P^2) explained by genetic variance (σ_G^2)

$$h^2 = \sigma_G^2 / \sigma_P^2$$

Expressive vocabulary (15-38m)

Receptive vocabulary (14-38m)

St Pourcain et al. Nat Comm. 2014; Hayiou-Thomas et al. Dev Sci. 2012; Dionne et al. Child Development. 2003; Dale et al. Journal of Child Language. 2000; Reznick et al. Monogr Soc Res Child Dev. 1997; Verhoef et al. PLOS Genet. 2021

Heritability

The phenotypic variance (σ_P^2) explained by genetic variance (σ_G^2)

$$h^2 = \sigma_G^2 / \sigma_P^2$$

Expressive vocabulary (15-38m)

Receptive vocabulary (14-38m)

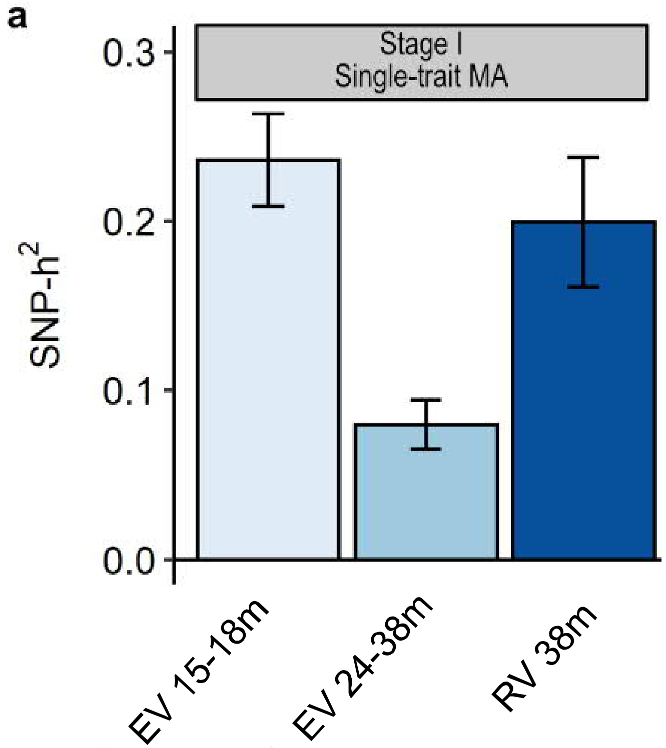
St Pourcain et al. Nat Comm. 2014; Hayiou-Thomas et al. Dev Sci. 2012; Dionne et al. Child Development. 2003; Dale et al. Journal of Child Language. 2000; Reznick et al. Monogr Soc Res Child Dev. 1997; Verhoef et al. PLOS Genet. 2021

Heritability

The phenotypic variance (σ_P^2) explained by genetic variance (σ_G^2)

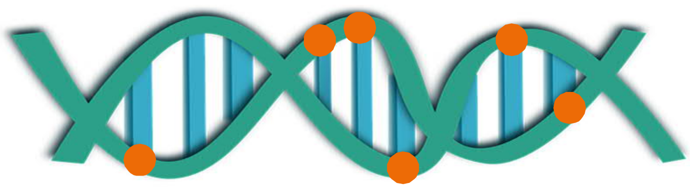
$$h^2 = \sigma_G^2 / \sigma_P^2$$

7 cohorts
$N_{individuals}$ = 6,291 – 16,615
$N_{observations}$ = 6,291 - 22,823



Verhoef et al. BioRxiv. 2022

Are individual differences in early-life vocabulary size (partially) attributable to genetic variation?



Per SNP

Yes, but effects are very small

Across all SNPs

Yes, ~10-20% of the variance

MAX PLANCK

Are the same genetic influences related to vocabulary size throughout early development?

Genetic influences across vocabulary development

- DNA code is (relatively) stable throughout life
- Genetic associations with a certain phenotype do not have to be stable.
  - Gene expression patterns change over time *(Li et al. 2018)*
  - Different underlying biological processes *(Plomin & Deary, 2015)*
  - Environmental variance changes *(Haworth et al. 2010)*

Genetic influences across vocabulary development

Genetic correlation (between 2 phenotypes)

The extent to which two phenotypes are influenced by the same genetic variation

Values range from

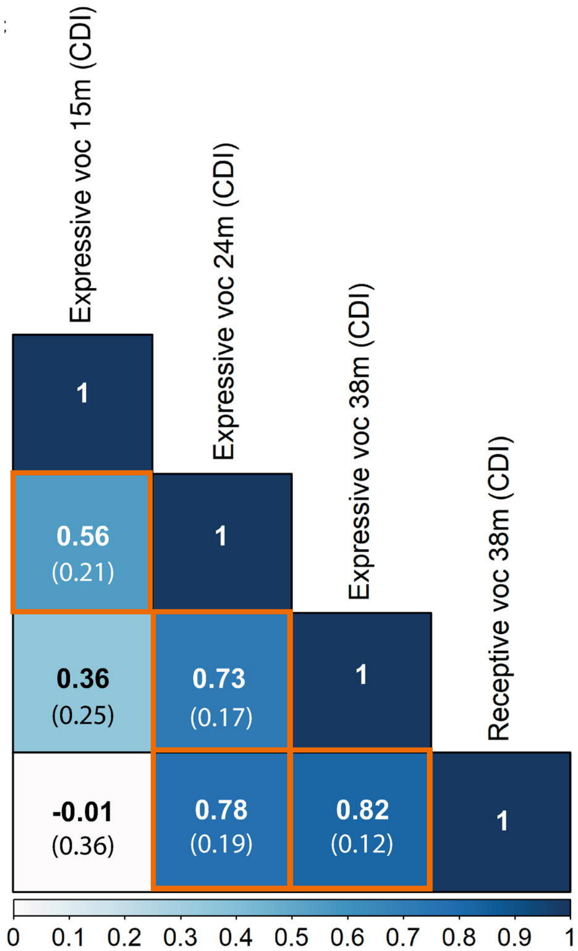
1: all genetic variation is shared, with the same direction of effect

0: no genetic variation is shared

-1: all genetic variation is shared, with opposite direction of effect

MAX PLANCK

Genetic influences across vocabulary development



Genetic correlations

Avon Longitudinal Study of Parents and Children
6,014 ≤ N ≤ 6,524

CDI Words & Gestures (abbreviated) - 15 months
CDI Words & Sentences (abbreviated) - 24 & 38 months

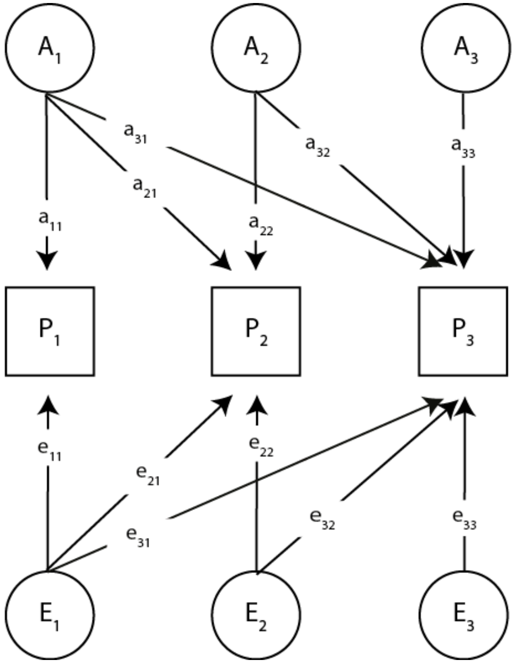
SNP data

Verhoef et al. PLOS Genet. 2021

Genetic influences across vocabulary development

Genetic-relationship-matrix structural equation modelling (grm-sem)

Multivariate methodology that allows to simultaneously examine genetic relationships for ≥2 traits



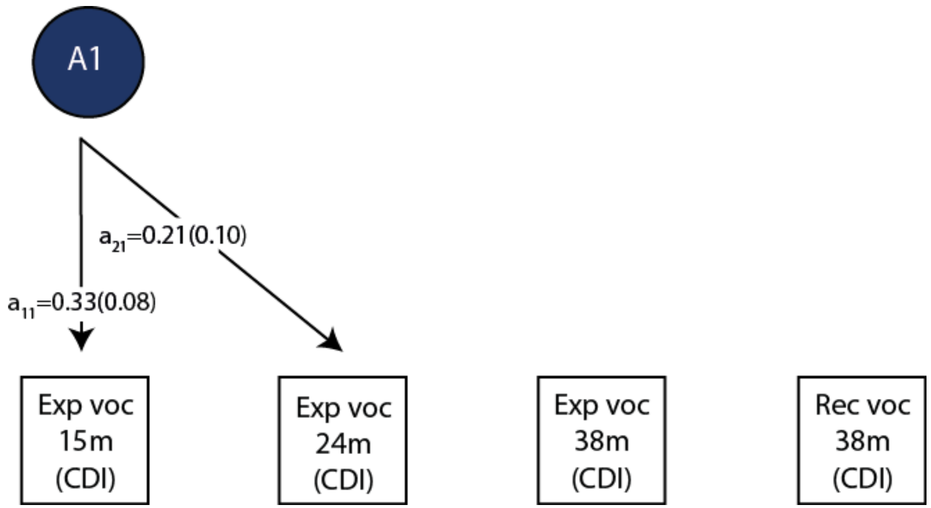
Cholesky decomposition

Genetic factors

Phenotype

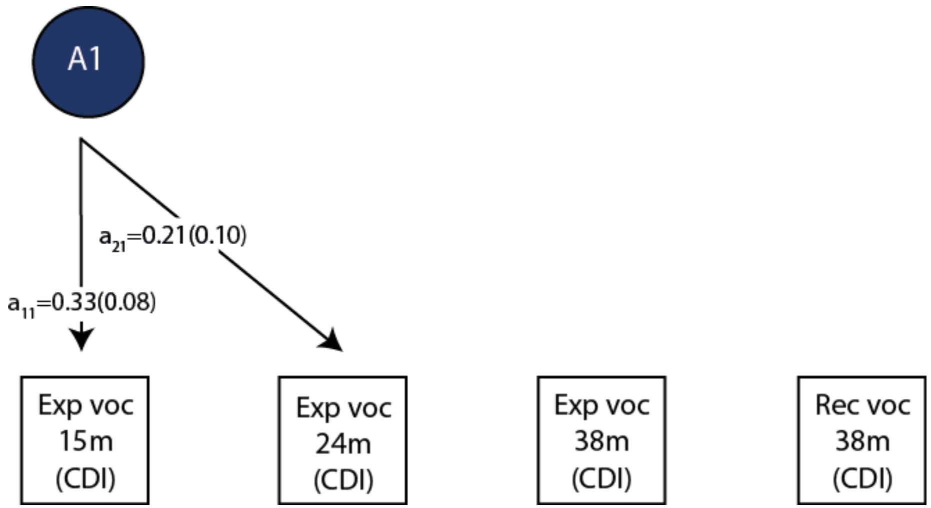
Residual factors

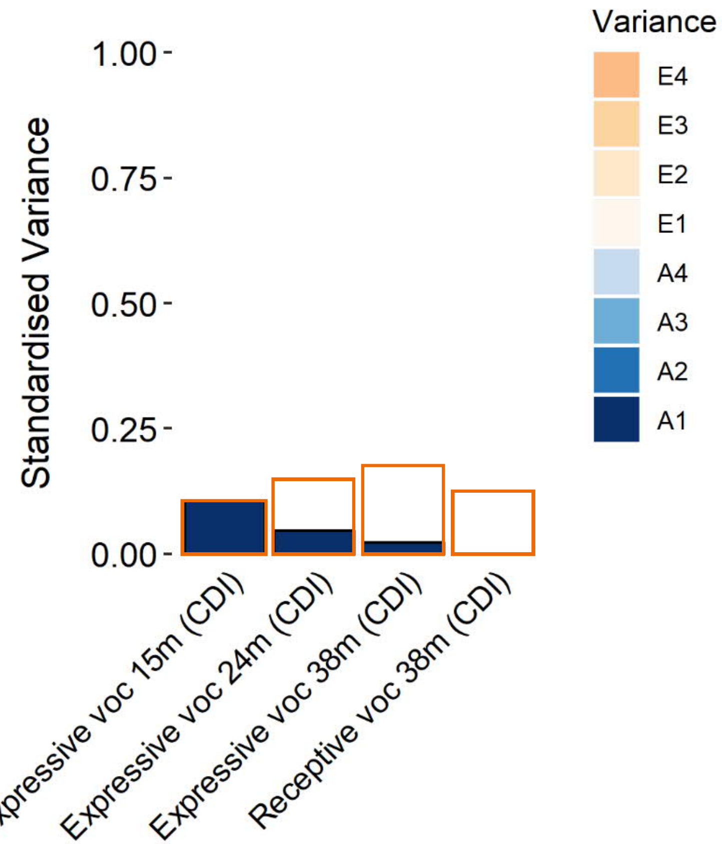
St Pourcain et al. Biological Psychiatry. 2018
https://gitlab.gwdg.de/beate.stpourcain/grmsem



Paths are shown for path coefficients P<0.05 only

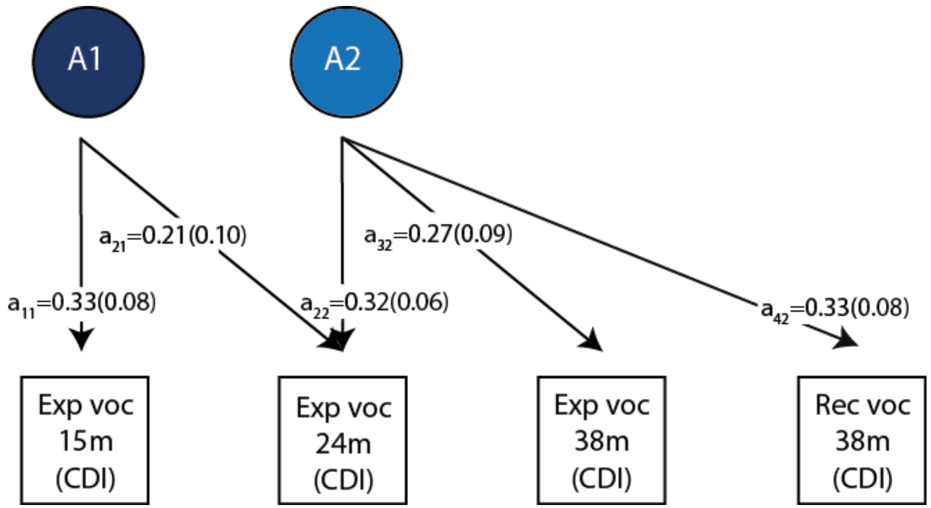
Verhoef et al. *PLOS Genet.* 2021

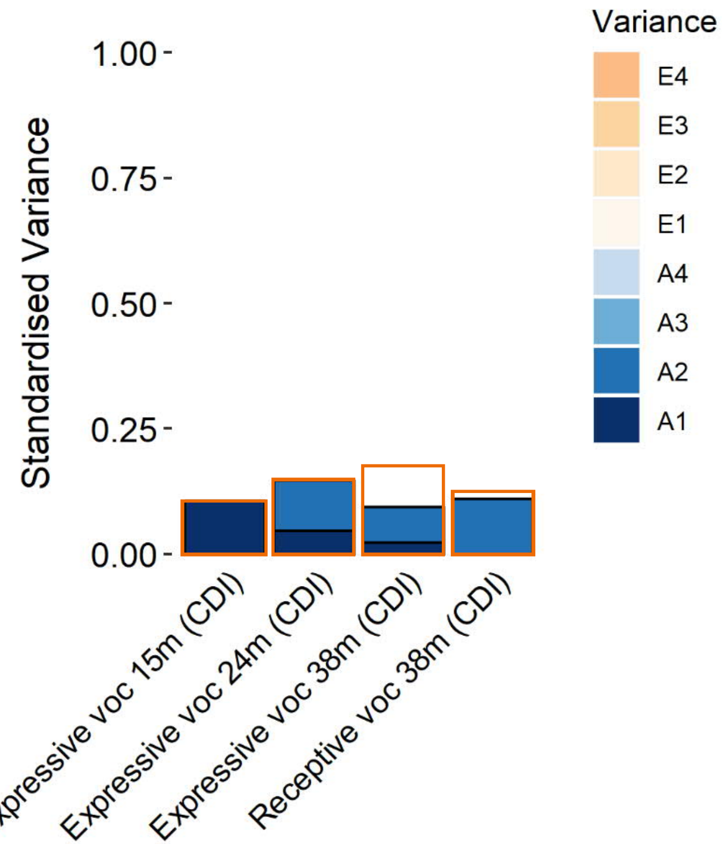




Paths are shown for path coefficients *P*<0.05 only

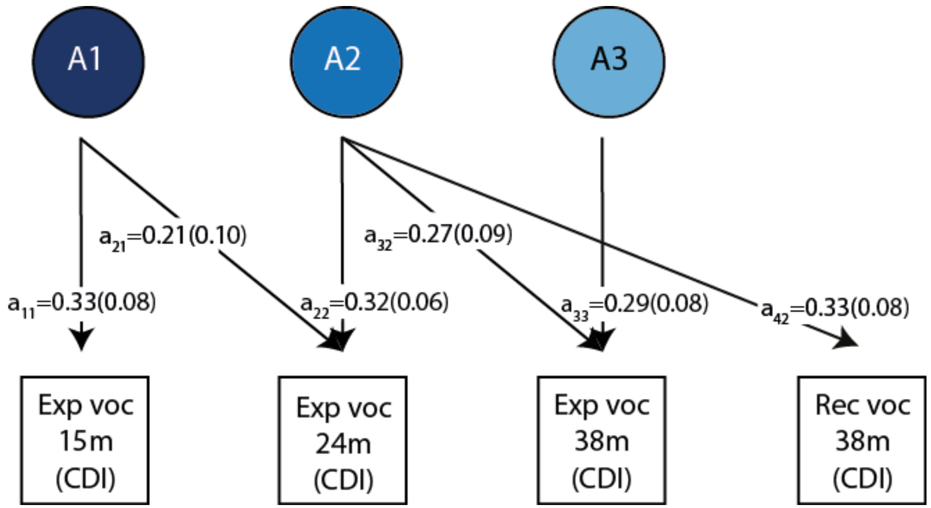
Verhoef et al. PLOS Genet. 2021

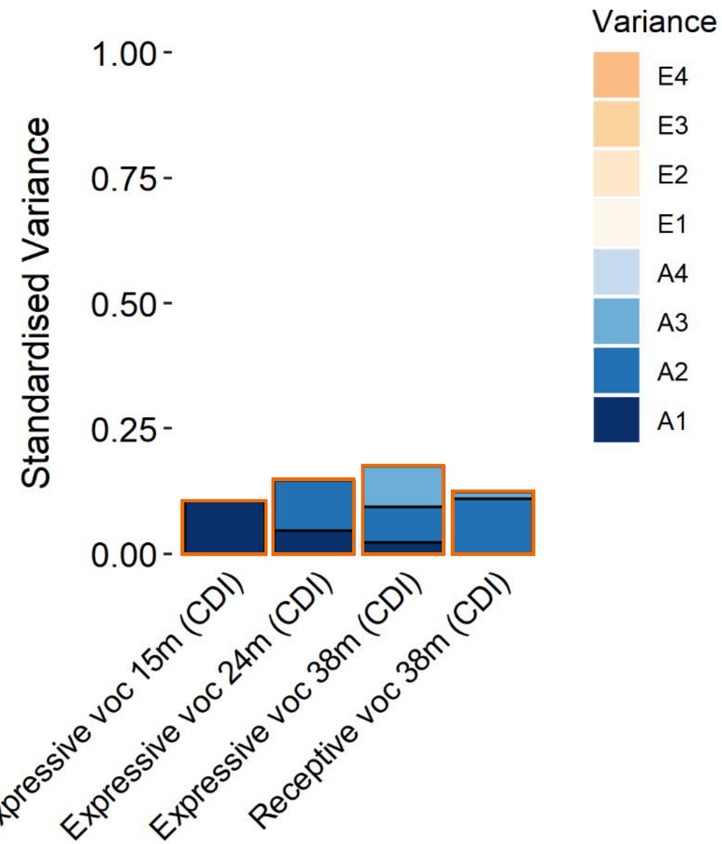




Paths are shown for path coefficients *P*<0.05 only

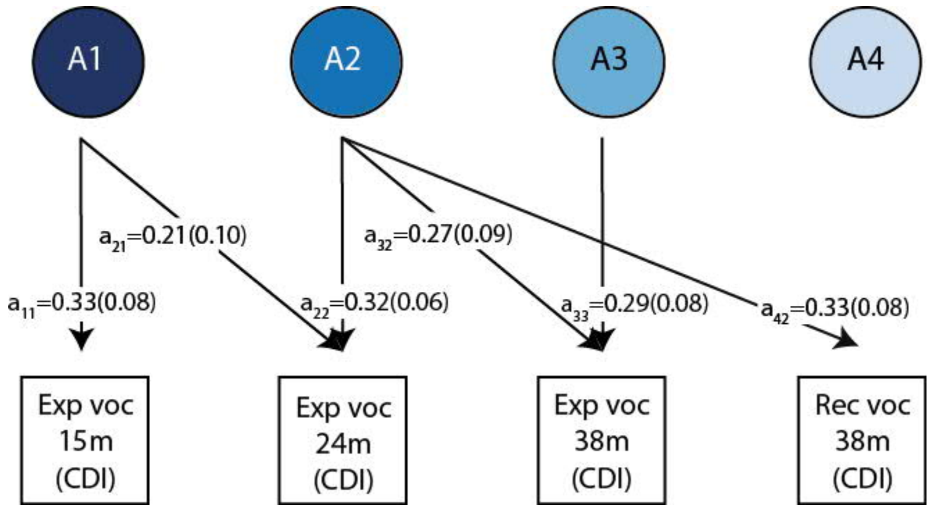
Verhoef et al. *PLOS Genet.* 2021

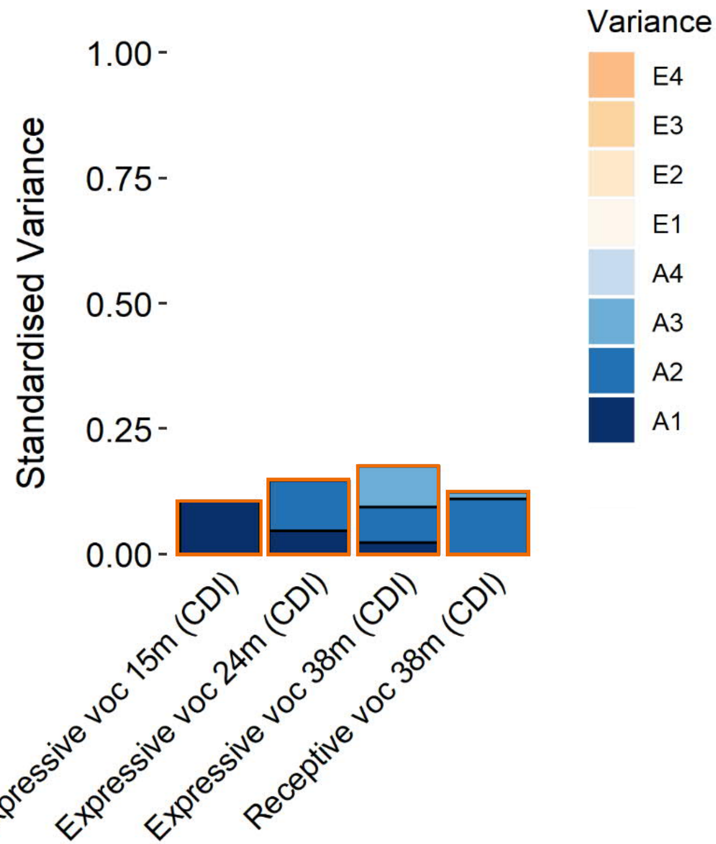




Paths are shown for path coefficients *P*<0.05 only

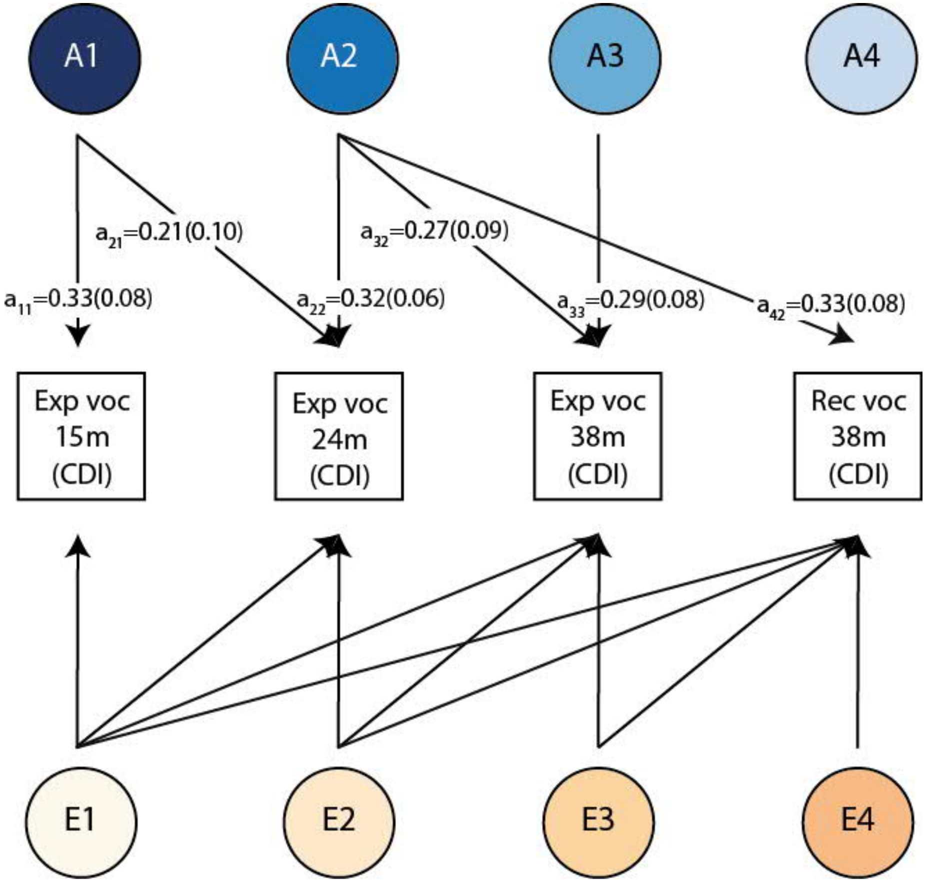
Verhoef et al. *PLOS Genet.* 2021

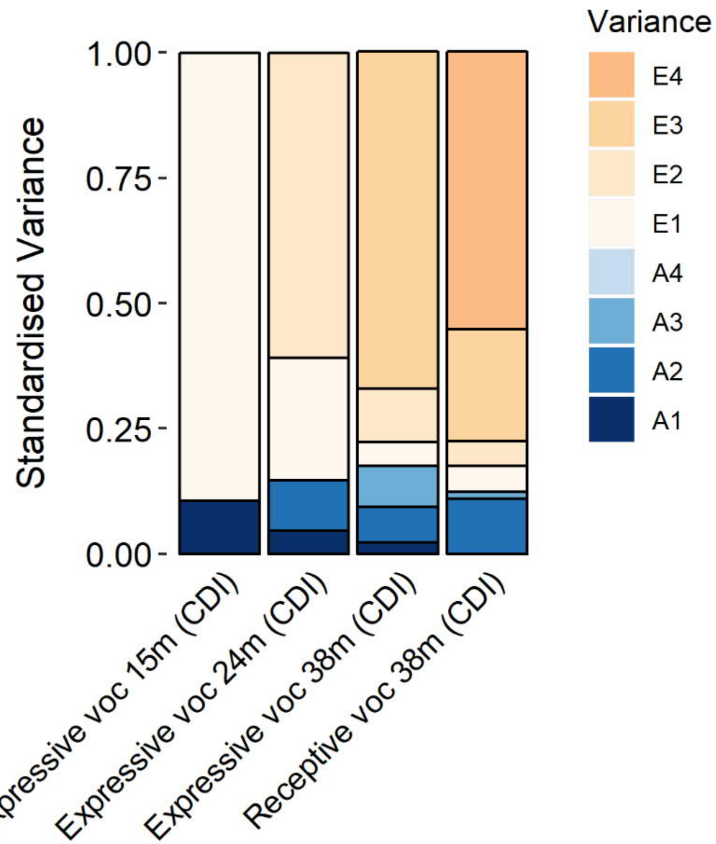




Paths are shown for path coefficients *P*<0.05 only

Verhoef et al. *PLOS Genet.* 2021





Paths are shown for path coefficients *P*<0.05 only

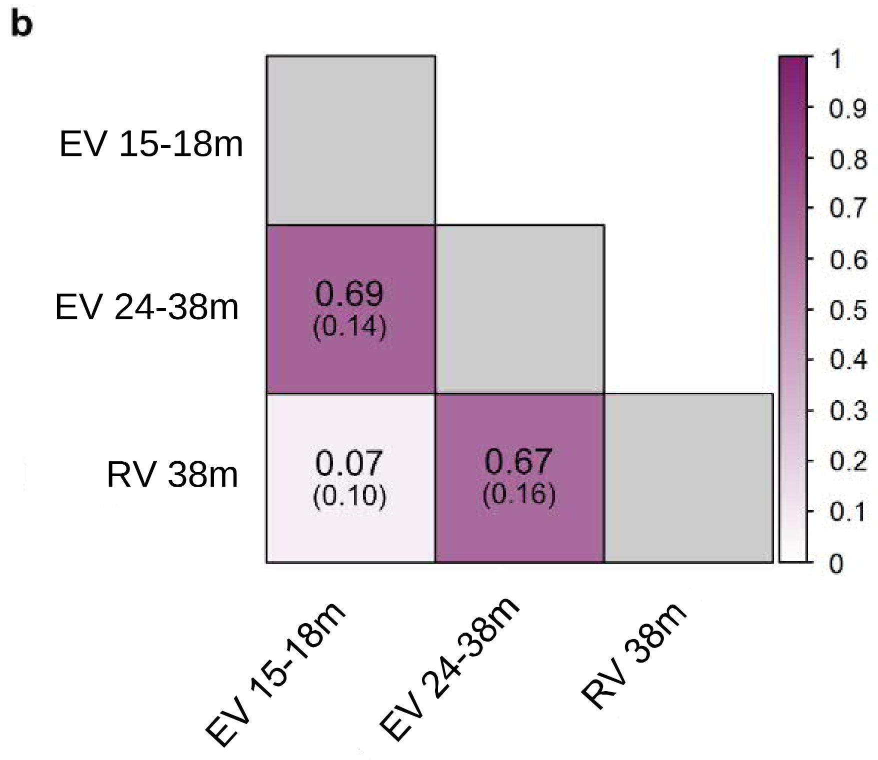
Verhoef et al. *PLOS Genet.* 2021

Genetic influences across vocabulary development

EAGLE
EArly Genetics & Lifecourse Epidemiology Consortium

7 cohorts
$N_{individuals}$ = 6,291 – 16,615
$N_{observations}$ = 6,291 - 22,823

Genetic correlations



Verhoef et al. BioRxiv. 2022

Are the same genetic influences related to vocabulary size throughout early development?

No,

multiple genetic factors contribute to individual differences in vocabulary size during the first few years of life.

Early language and later cognitive development

- Early vocabulary (2-4 years) predicts vocabulary, but also aspects of syntax, figurative language and pragmatics at 12 years of age *(Hayiou-Thomas et al. 2012)*
- Vocabulary at 10 months predicts non-verbal skills at 10 years and school achievement at the end of primary school *(Hohm et al. 2007)*
- Delayed language abilities are associated with lower intelligence quotient *(Webster et al. 2006; Liao et al. 2015)*

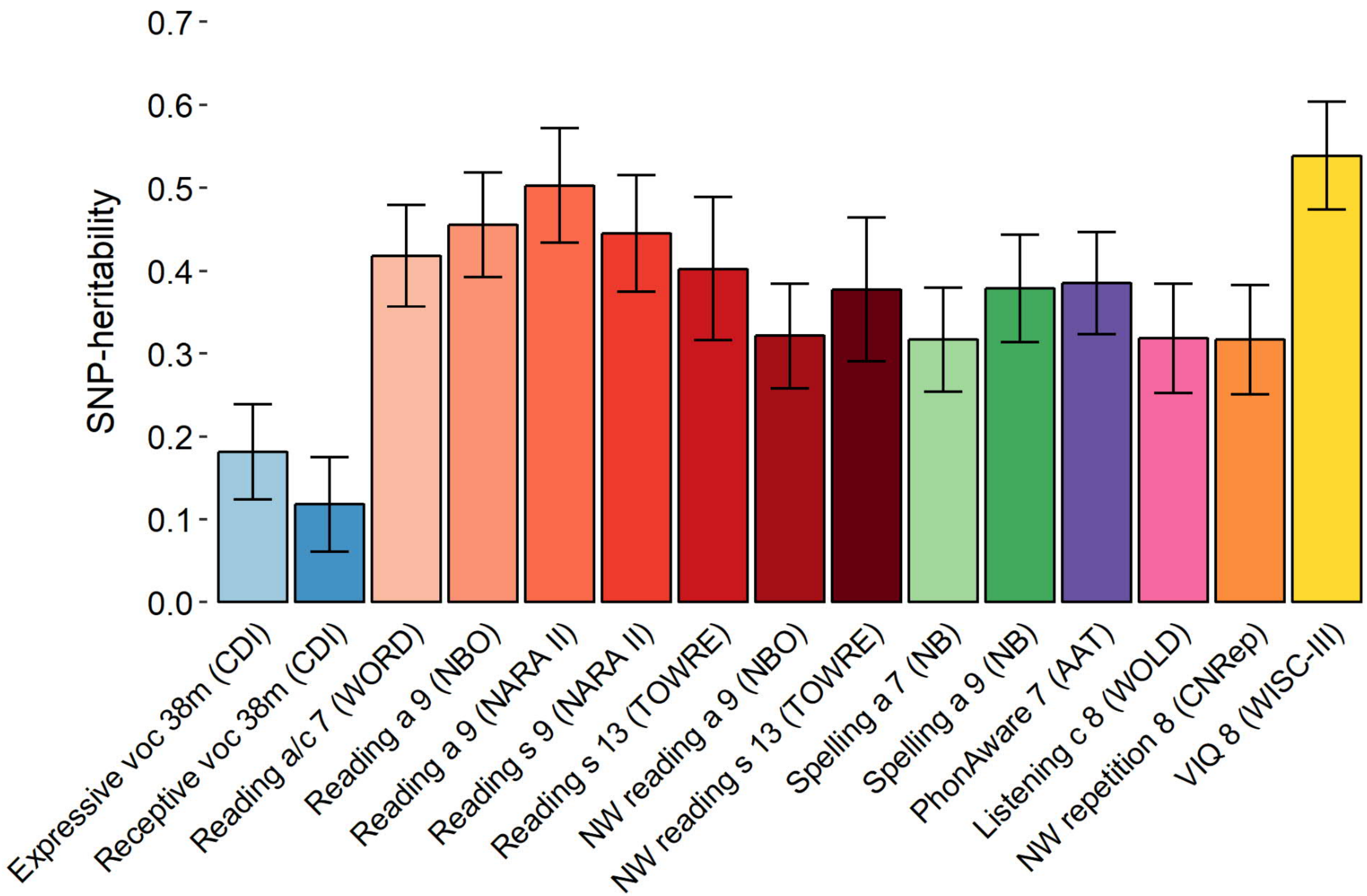
Do early-life vocabulary and
later cognition-related abilities share genetic influences?

Early-life vocabulary and later cognition-related abilities

- Several psychological instruments
- Sample size > 4,000

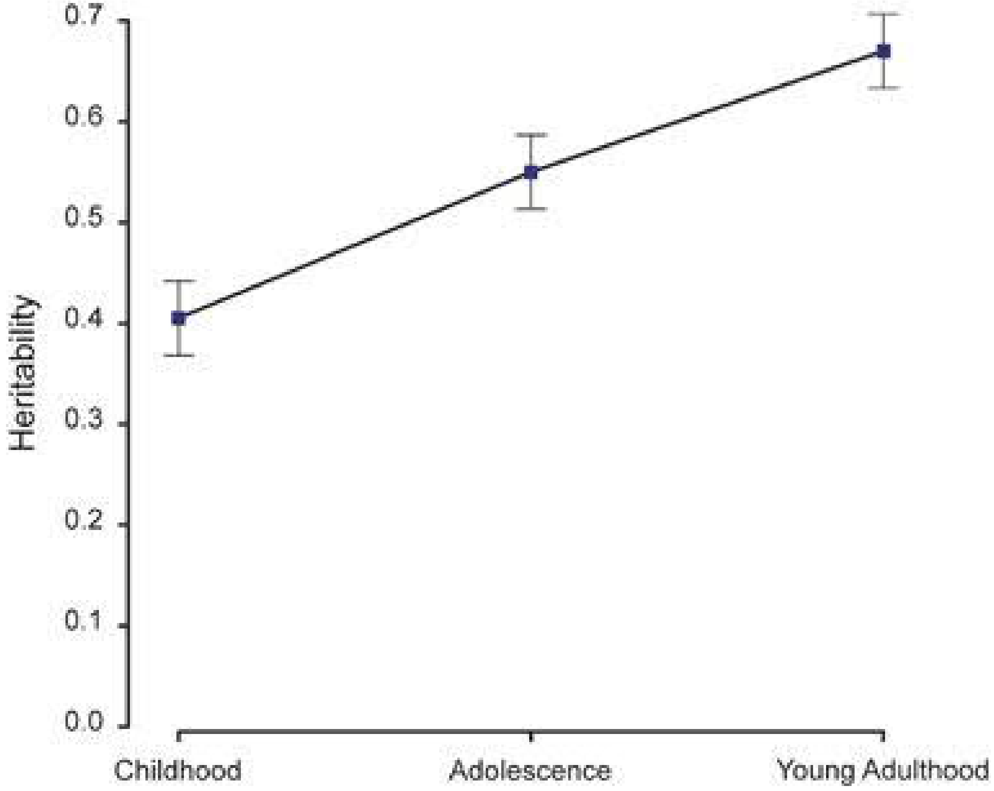
Ability	Age (year)
Reading (comprehension, speed, accuracy)	7,9,13
Spelling (accuracy)	7,9
Non-word reading (speed, accuracy)	9,13
Phonemic awareness	7
Verbal intelligence	8
Listening comprehension	8
Non-word repetition	8

Verhoef et al. JCPP 2020



Bars represent standard errors

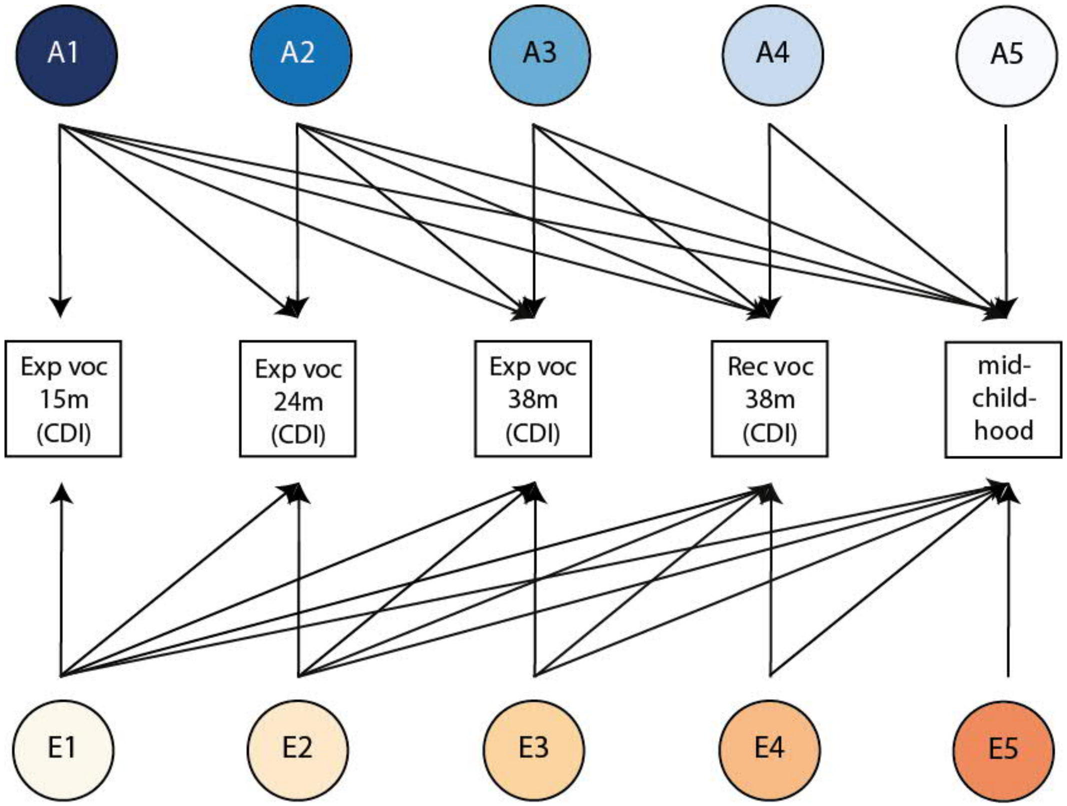
Verhoef et al. JCPP 2020



Innovation
Novel genetic influences arise during development

Amplification
Early genetic influences become increasingly important with age

Plomin & Deary, 2015; Plomin & DeFries 1958, Briley & Tucker-Drob 2013



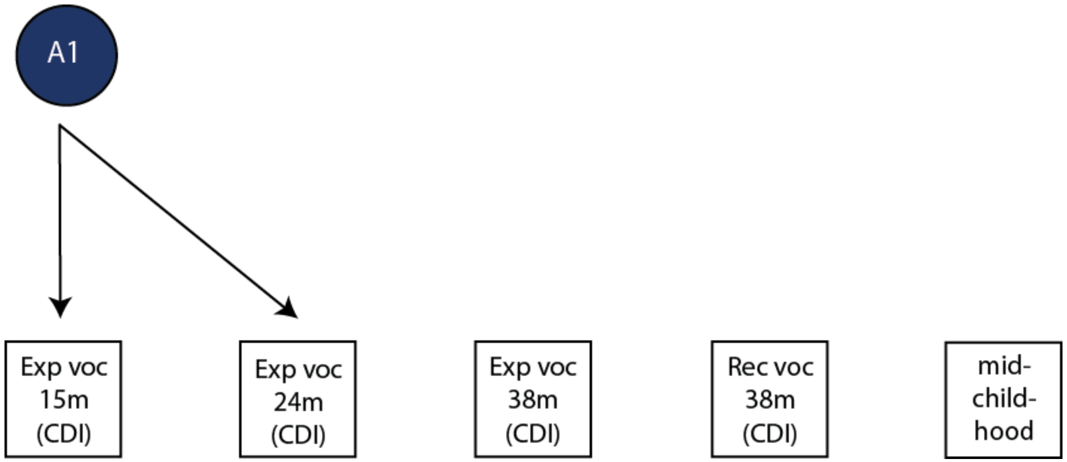
Genetic-relationship-matrix structural equation modelling (grm-sem)

3 Cholesky decompositions

- Reading a/c 7 (WORD)
- VIQ 8 (WISC-III)
- PIQ 8 (WISC-III)

Verhoef et al. PLOS Genet. 2021

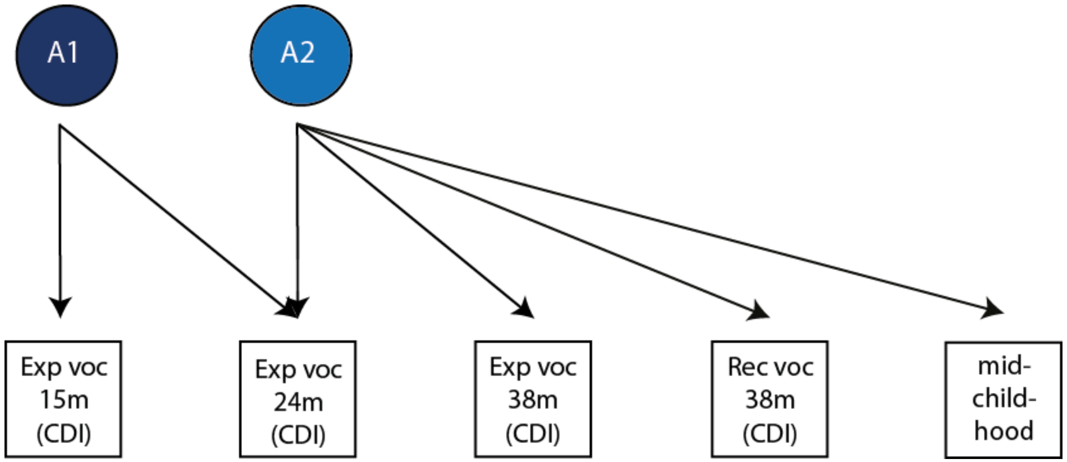
MAX PLANCK



- A1 unrelated to all mid-childhood traits

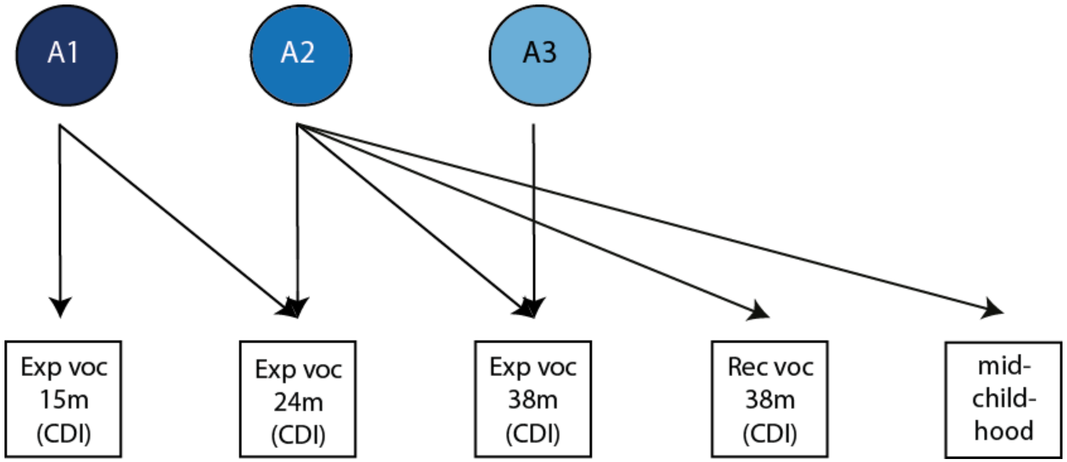
Verhoef et al. PLOS Genet. 2021

MAX PLANCK



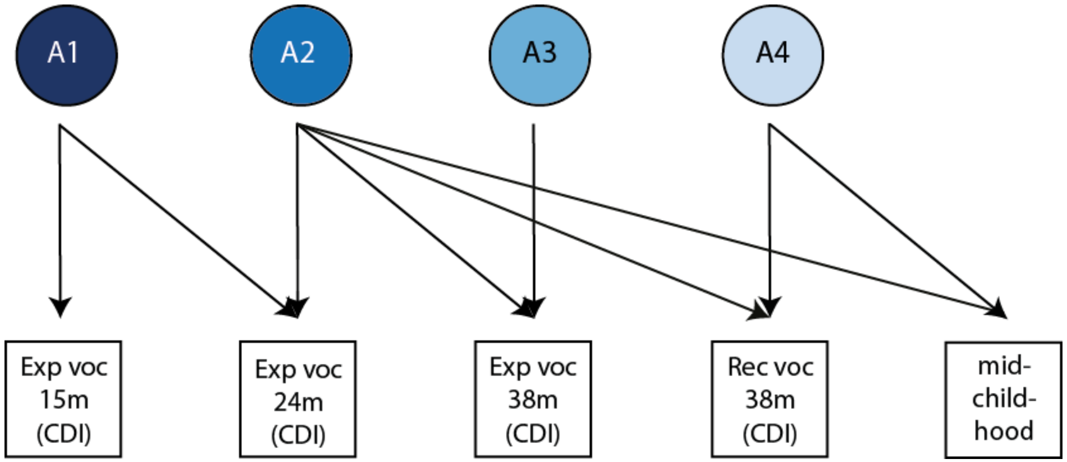
- A1 unrelated to all mid-childhood traits
- A2 only related to reading and VIQ

Verhoef et al. *PLOS Genet.* 2021



- A1 unrelated to all mid-childhood traits
- A2 only related to reading and VIQ
- A3 unrelated to all mid-childhood traits

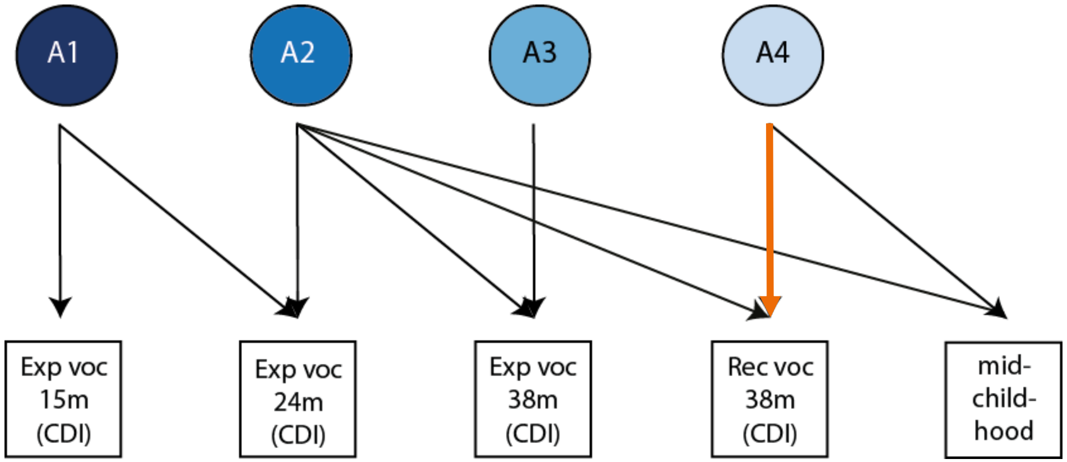
Verhoef et al. PLOS Genet. 2021



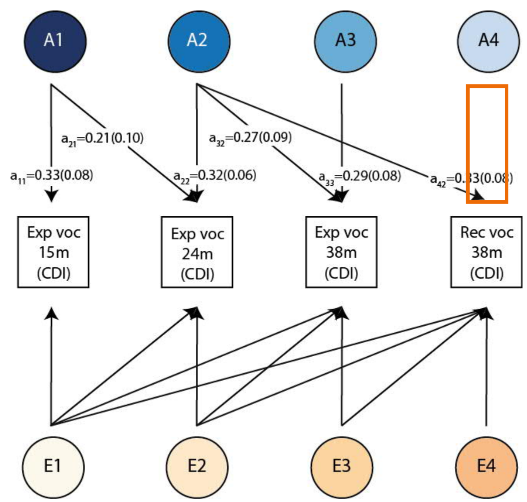
- A1 unrelated to all mid-childhood traits
- A2 only related to reading and VIQ
- A3 unrelated to all mid-childhood traits
- A4 related to all mid-childhood traits

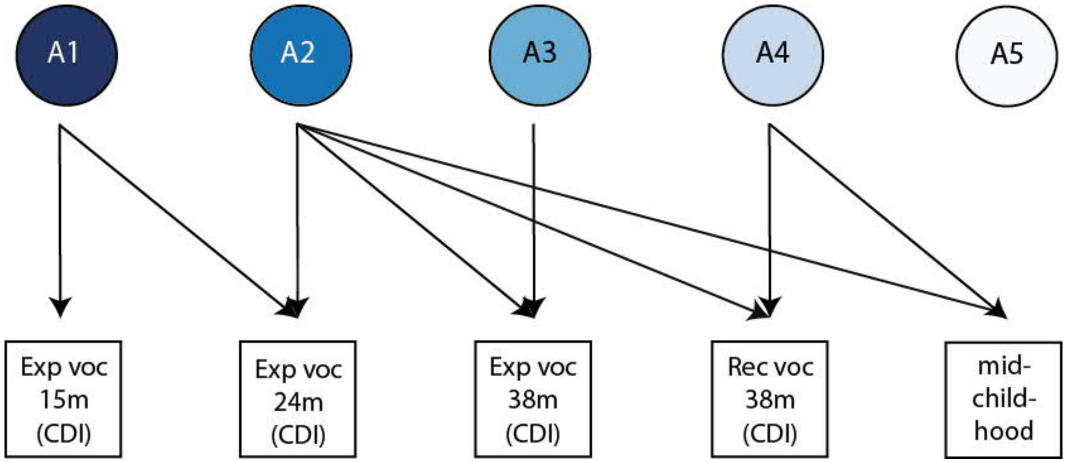
Verhoef et al. PLOS Genet. 2021

MAX PLANCK



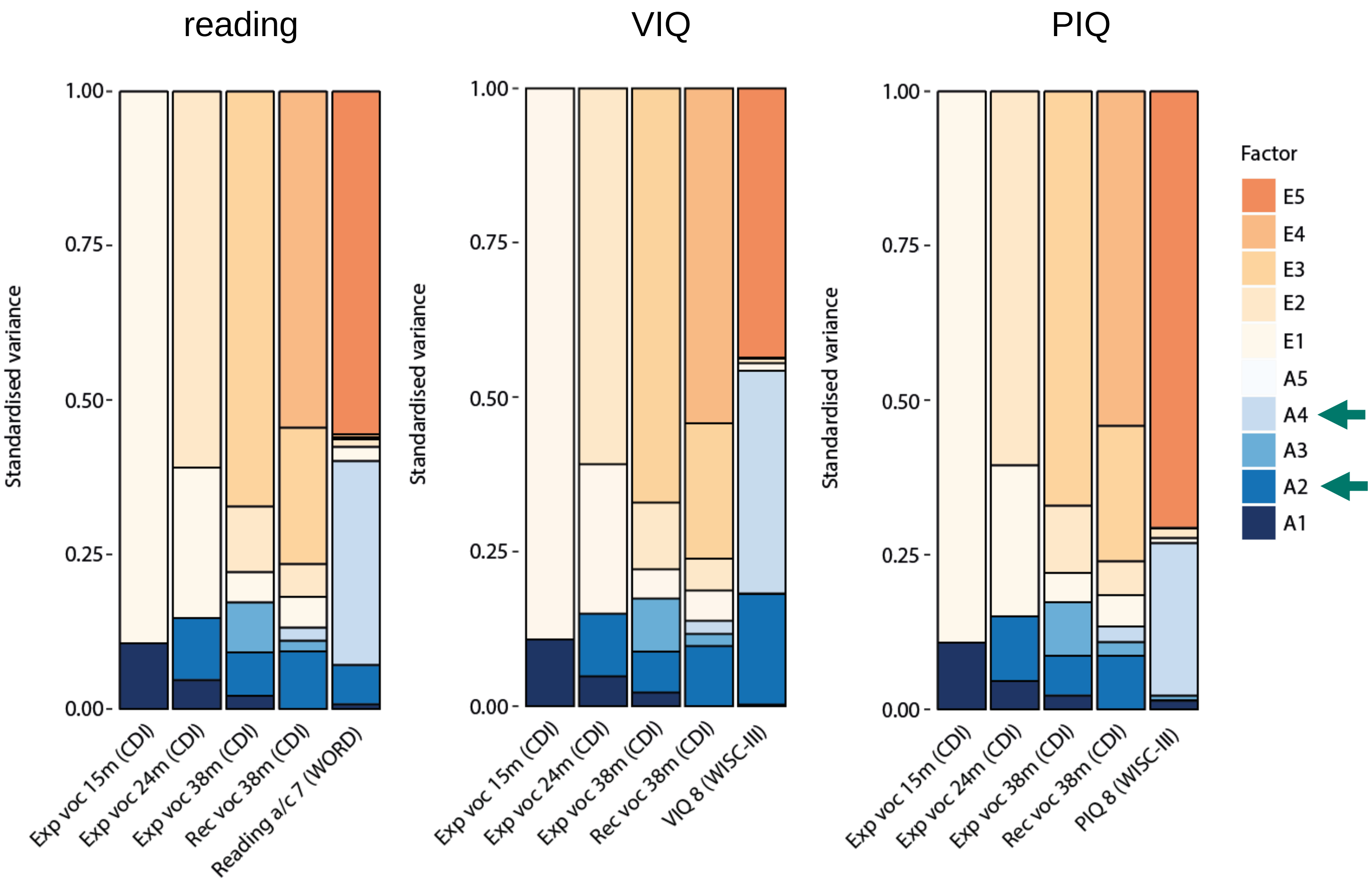
- A1 unrelated to all mid-childhood traits
- A2 only related to reading and VIQ
- A3 unrelated to all mid-childhood traits
- A4 related to all mid-childhood traits





- A1 unrelated to all mid-childhood traits
- A2 only related to reading and VIQ
- A3 unrelated to all mid-childhood traits
- A4 related to all mid-childhood traits
- No evidence for A5

Verhoef et al. PLOS Genet. 2021



Innovation
Novel genetic influences arise during development

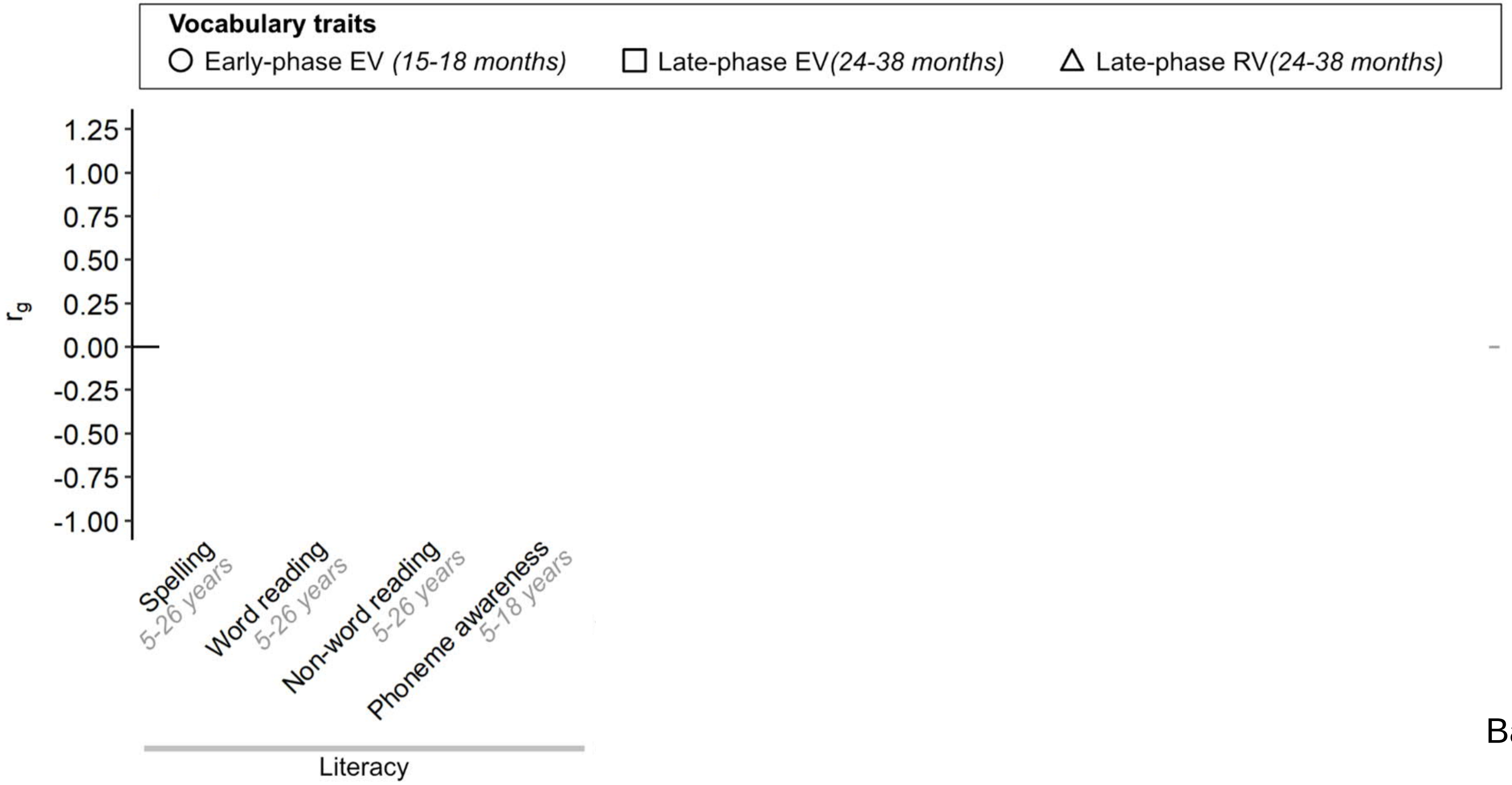
Amplification
Early genetic influences become increasingly important with age

Verhoef et al. PLOS Genet. 2021

Do early-life vocabulary and later cognition-related abilities share genetic influences?

Yes, verbal cognitive processes are genetically linked to word production at 2 years of age, while more general cognitive processes are related to genetic influences emerging for receptive vocabulary 3 years of age.

Genetic correlations



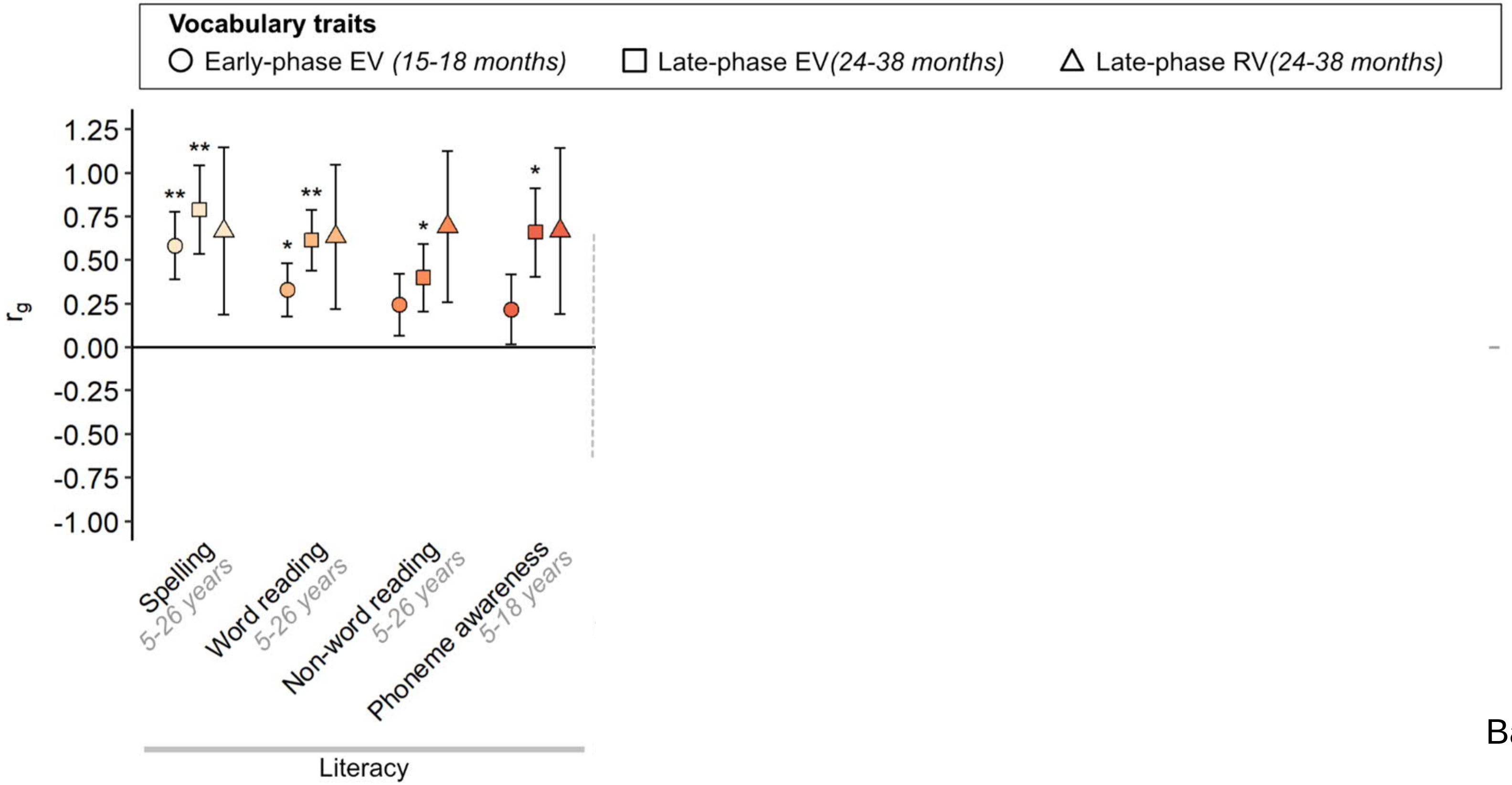
* P <0.05

** $P < 5.57 \times 10^{-3}$

Bars represent standard errors

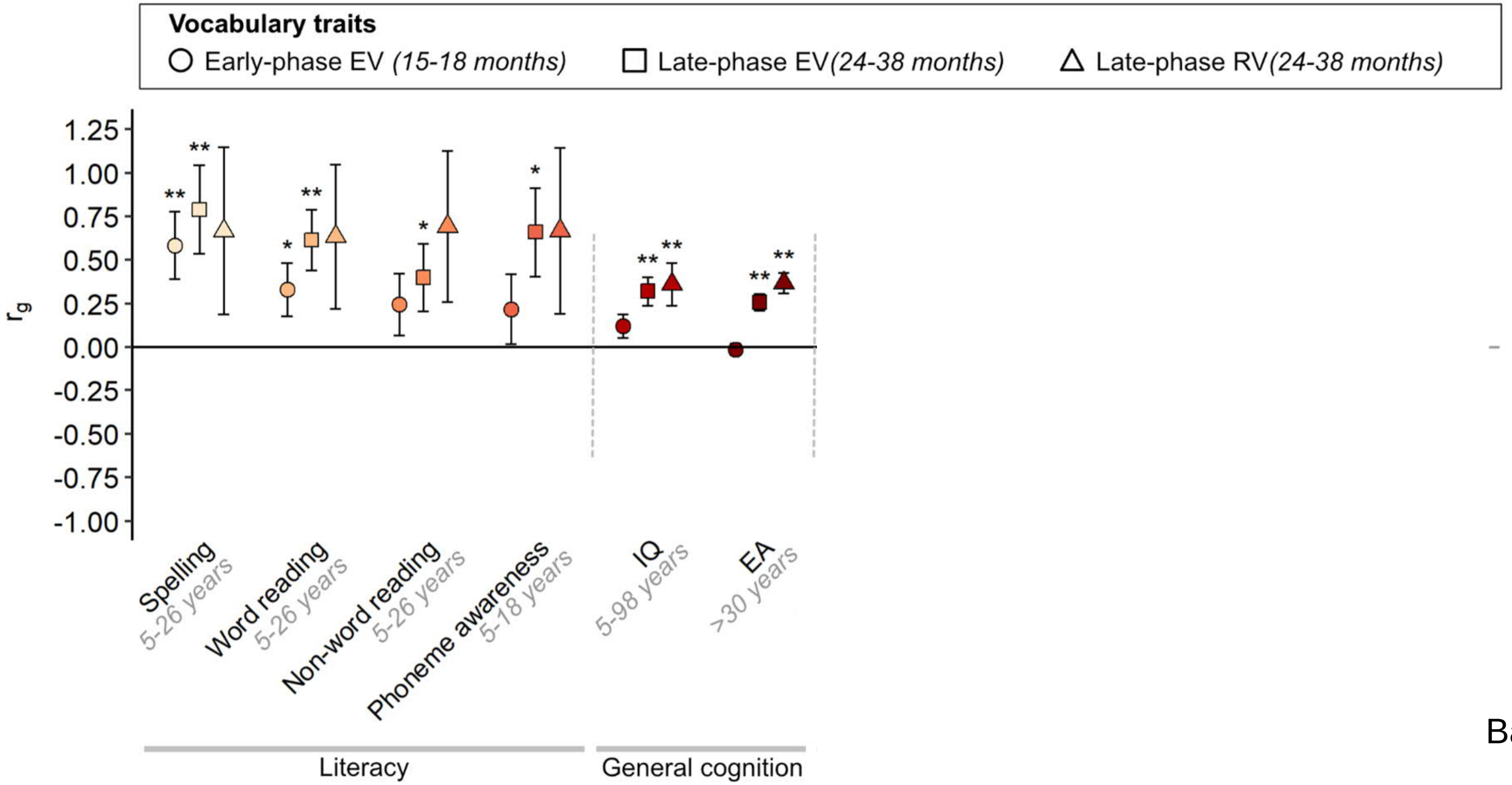
Verhoef et al. BioRxiv. 2022

Genetic correlations



* P <0.05
** $P < 5.57 \times 10^{-3}$
Bars represent standard errors

Genetic correlations



* P <0.05

** $P < 5.57 \times 10^{-3}$

Bars represent standard errors

Do early-life vocabulary and
later cognition-related abilities share genetic influences?

CHILDREN OF THE 90s

Yes, verbal cognitive processes are genetically linked to word production at 2 years of age, while more general cognitive processes are related to genetic influences emerging for receptive vocabulary 3 years of age.

EAGLE
EArly Genetics & Lifecourse Epidemiology Consortium

Yes, literacy abilities are genetically linked to word production, also in infancy, while more general cognitive processes are related to genetic influences emerging in toddlerhood.

MAX PLANCK

Early language and childhood-onset NDCs

ADHD

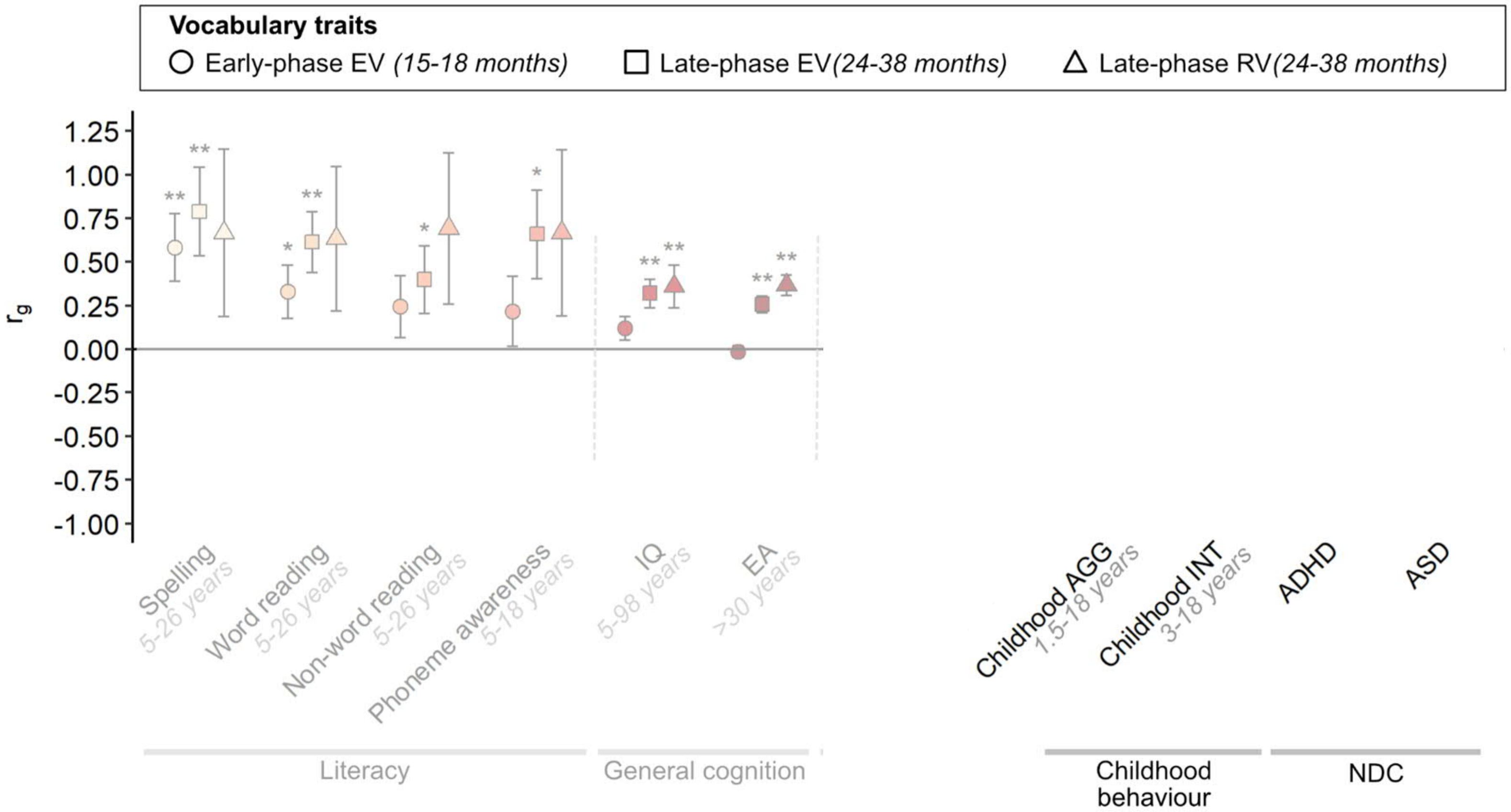
- Poor language skills at 3 years of age were predictive of inattention and hyperactivity at 5 years of age *(Peyre et al. 2016)*
- Genetic overlap of mid-childhood language & literacy abilities with ADHD risk (*Verhoef et al. 2019*)

ASD

- Wide phenotypic spectrum *(Tager-Flusberg et al. 2005)*
  - Little or no spontaneous speech by school age *(Ozonoff et al. 2000)*
  - Few language problems *(Ozonoff et al. 2000)*

Are genetic influences related to early-life vocabulary associated with childhood-onset neurodevelopmental conditions (NDCs)?

Early-life vocabulary, childhood behaviour and NDCs



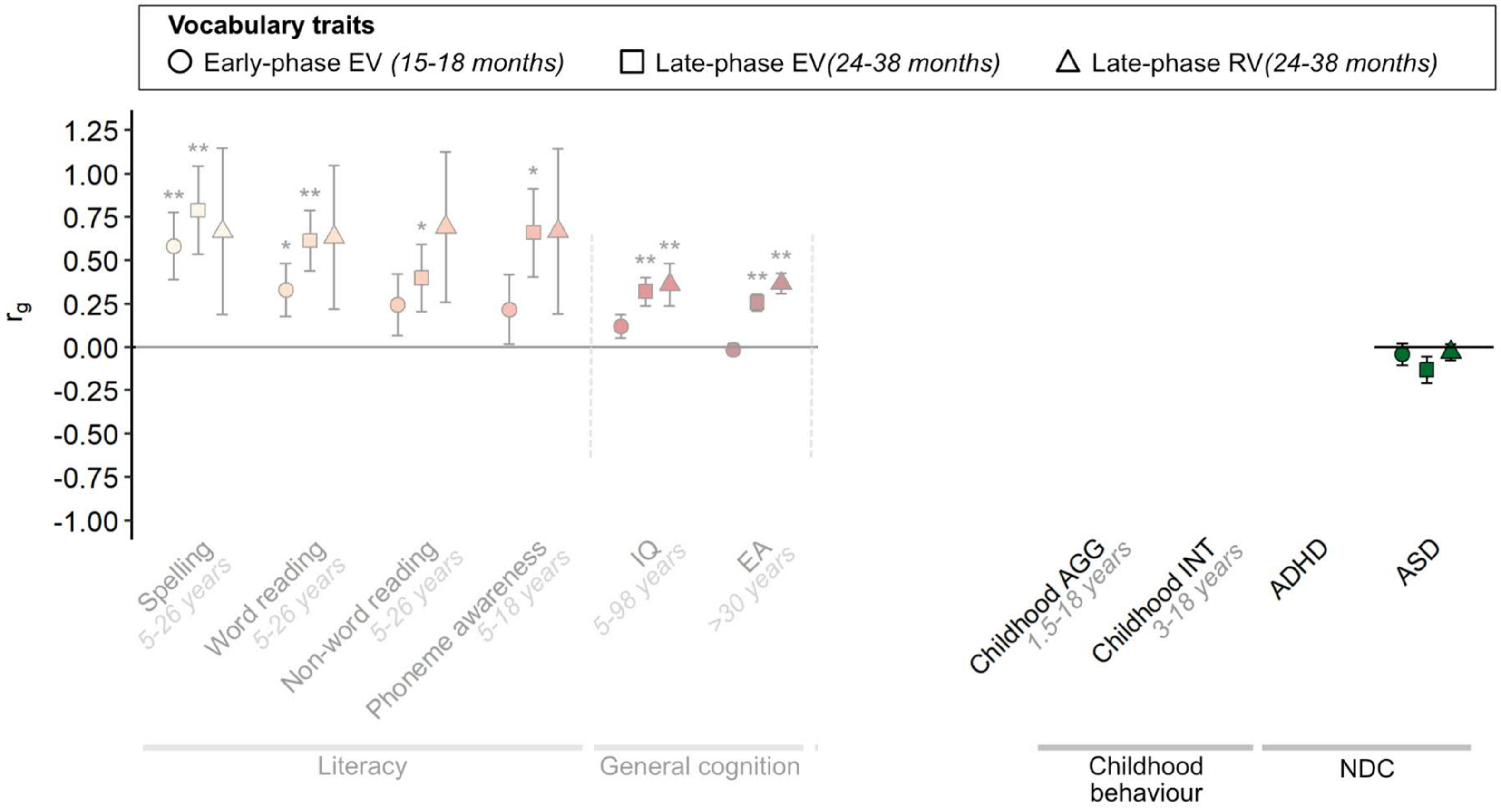
* P <0.05

** $P < 5.57\text{x}10^{-3}$

Bars represent standard errors

Verhoef et al. BioRxiv. 2022

Early-life vocabulary, childhood behaviour and NDCs



EAGLE
EArly Genetics & Lifecourse Epidemiology Consortium

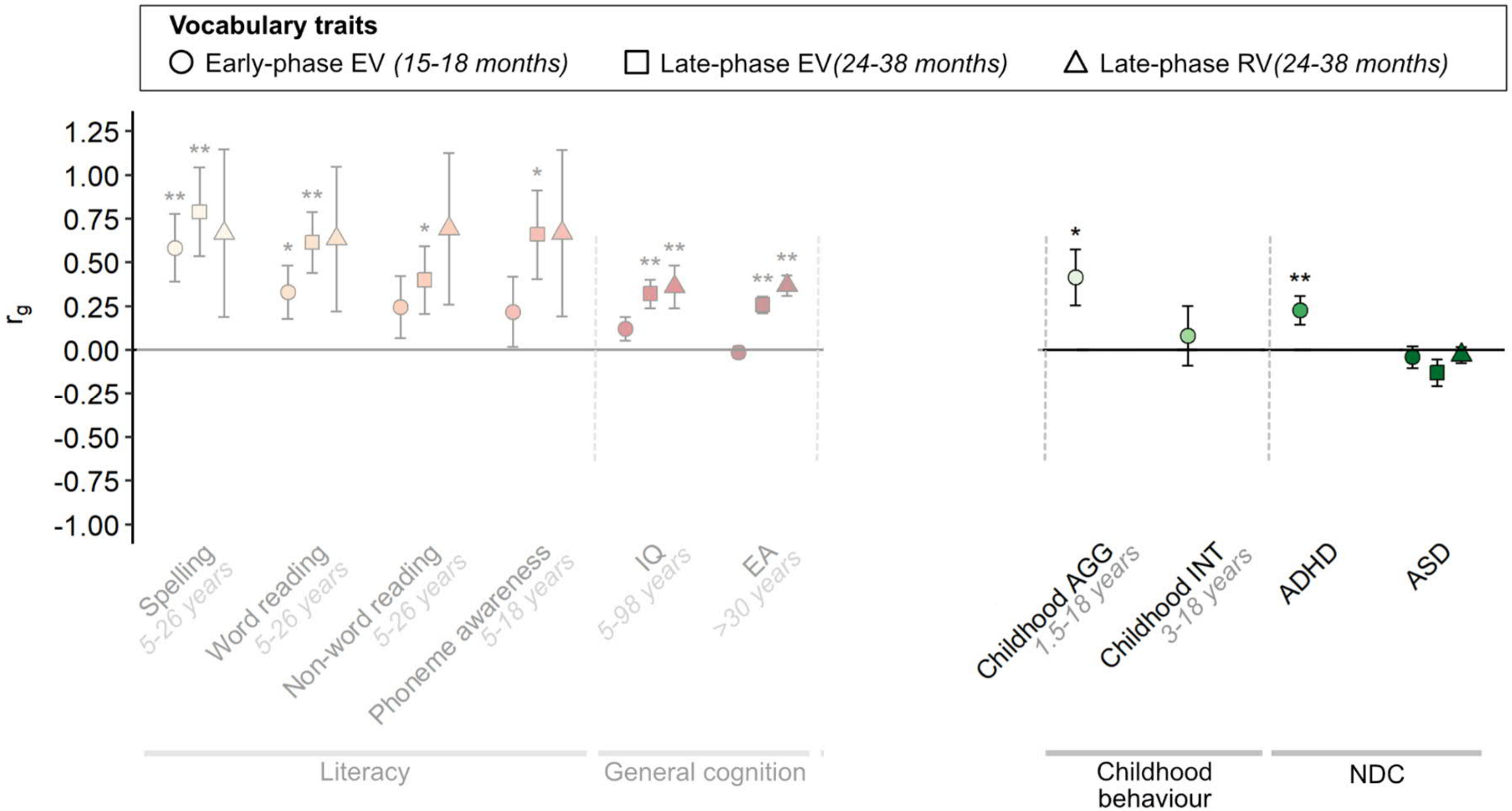
* P <0.05
** $P < 5.57 \times 10^{-3}$
Bars represent standard errors

EAGLE PGC iPSYCH

Verhoef et al. BioRxiv. 2022

MAX PLANCK

Early-life vocabulary, childhood behaviour and NDCs



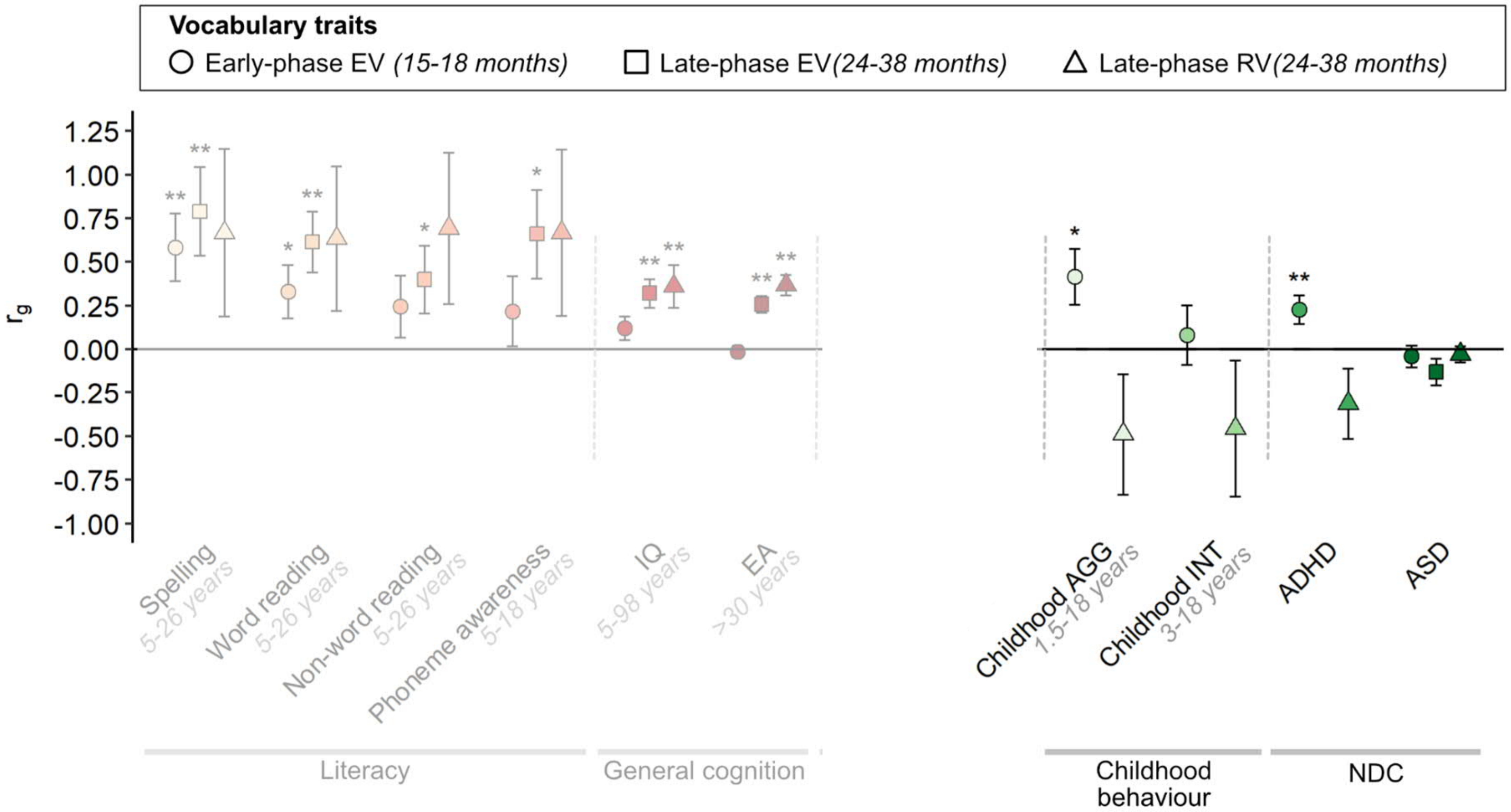
* P <0.05

** $P < 5.57 \times 10^{-3}$

Bars represent standard errors

Verhoef et al. BioRxiv. 2022

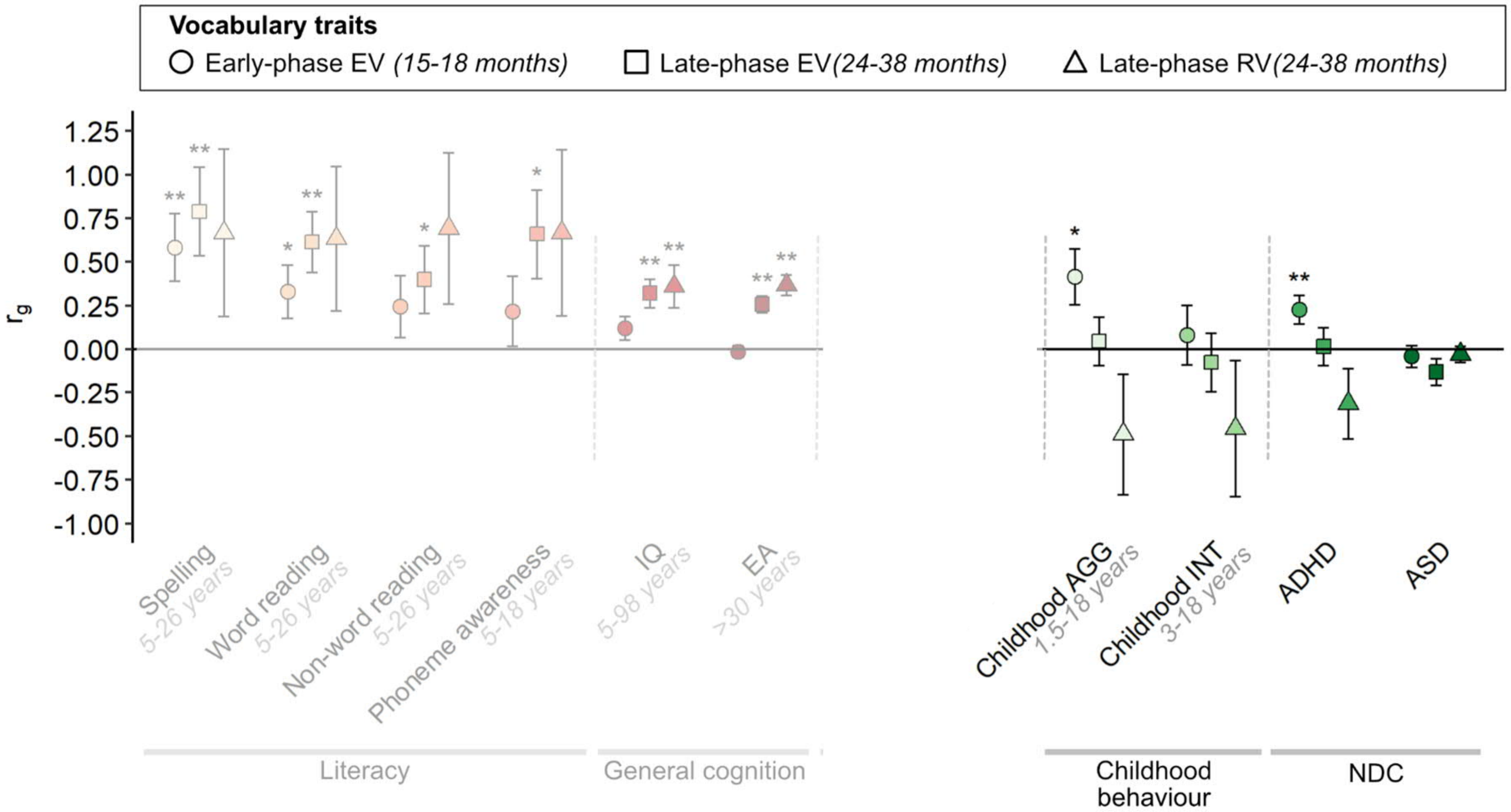
Early-life vocabulary, childhood behaviour and NDCs



* P <0.05
** $P < 5.57 \times 10^{-3}$
Bars represent standard errors

Verhoef et al. BioRxiv. 2022

Early-life vocabulary, childhood behaviour and NDCs



* $P < 0.05$

** $P < 5.57 \times 10^{-3}$

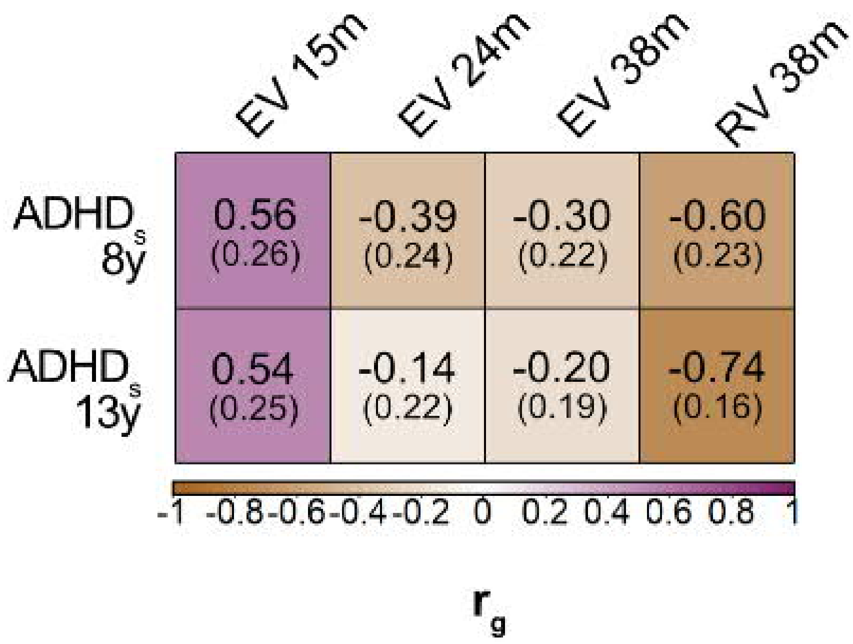
Bars represent standard errors

Verhoef et al. BioRxiv. 2022

Genetic links of early-life vocabulary with ADHD

Genetic correlations

N≤6,524



Verhoef et al. BioRxiv. 2022

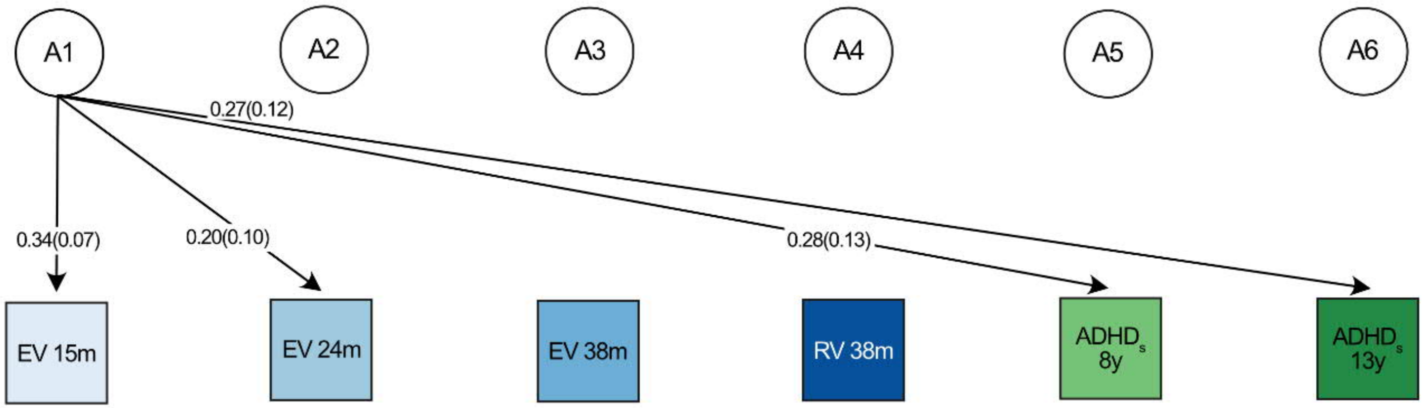
Genetic links of early-life vocabulary with ADHD

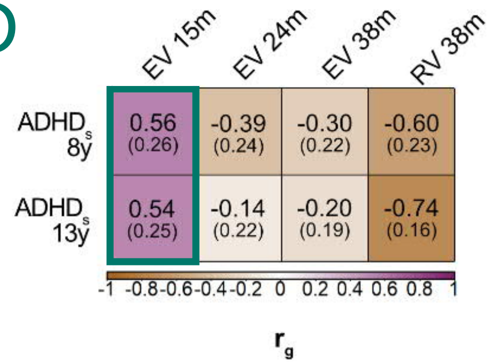
Genetic-relationship-matrix structural equation modelling (grm-sem)

Cholesky decomposition

Verhoef et al. BioRxiv. 2022

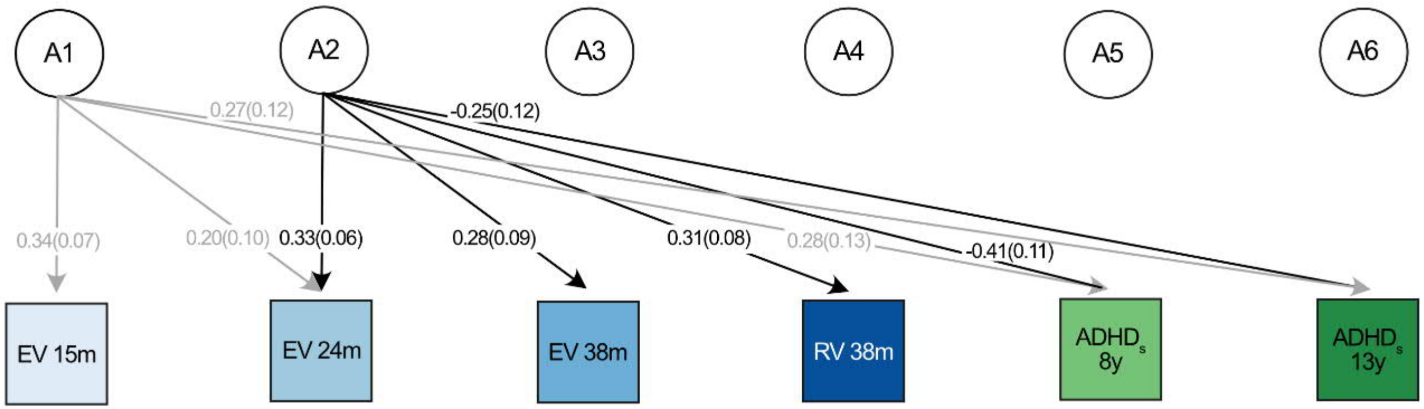
Genetic links of early-life vocabulary with ADHD

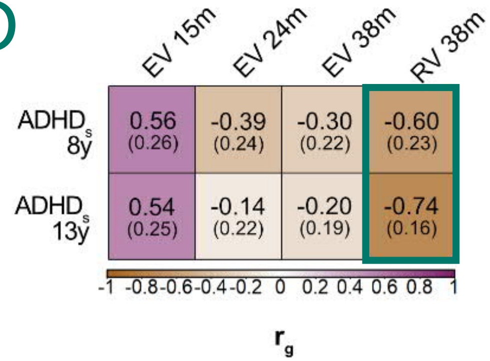




Verhoef et al. BioRxiv. 2022

Genetic links of early-life vocabulary with ADHD

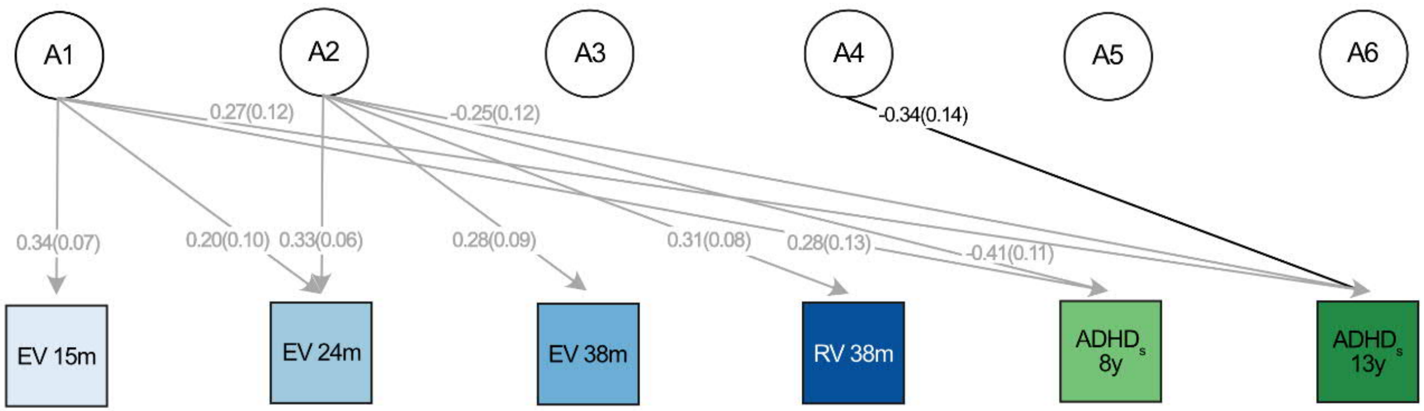


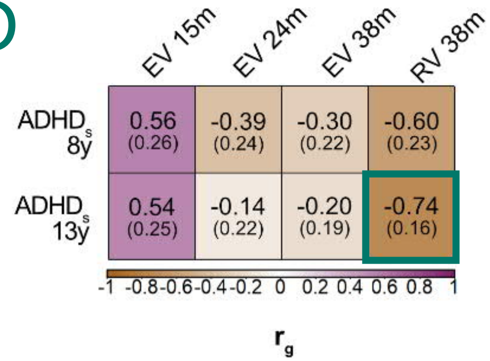


Verhoef et al. *BioRxiv.* 2022

MAX PLANCK

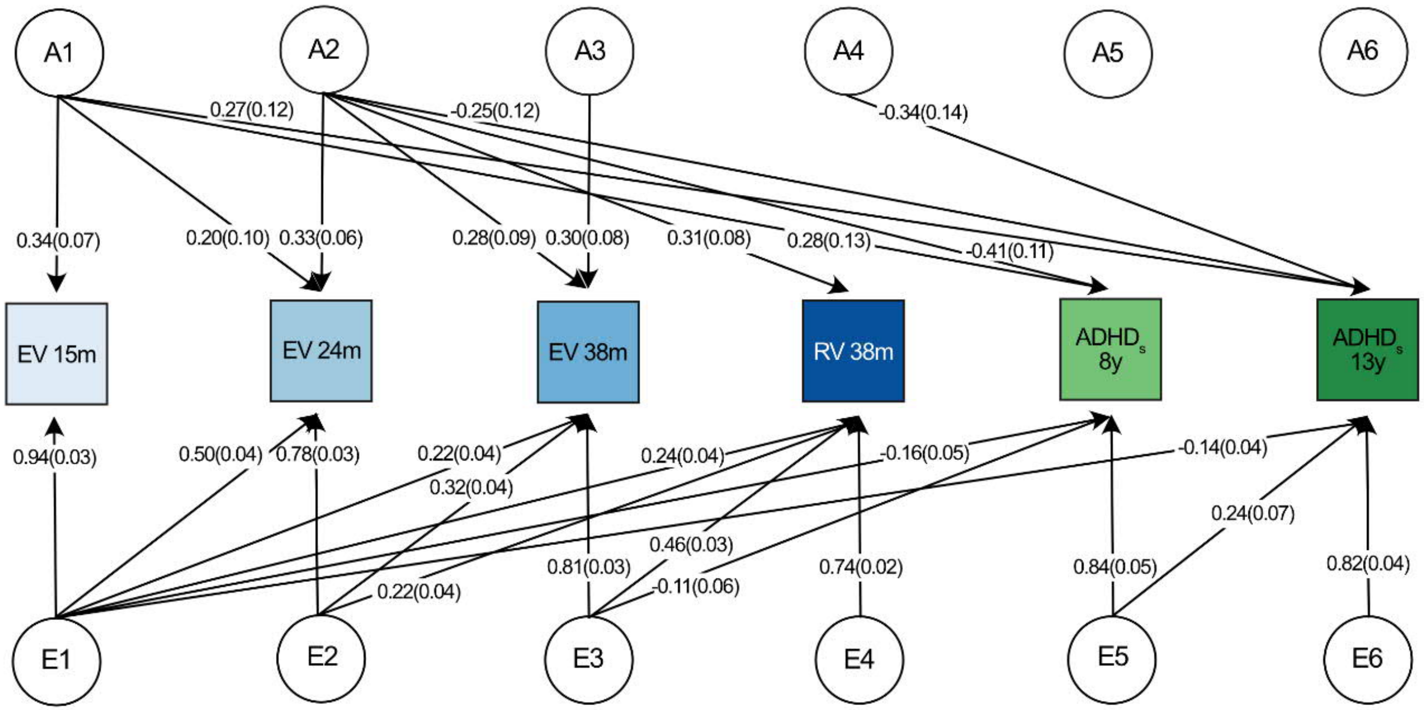
Genetic links of early-life vocabulary with ADHD





Verhoef et al. BioRxiv. 2022

Genetic links of early-life vocabulary with ADHD



	EV 15m	EV 24m	EV 38m	RV 38m
ADHDs 8y	0.56 (0.26)	-0.39 (0.24)	-0.30 (0.22)	-0.60 (0.23)
ADHDs 13y	0.54 (0.25)	-0.14 (0.22)	-0.20 (0.19)	-0.74 (0.16)

-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1

r_g

Verhoef et al. *BioRxiv. 2022*

MAX PLANCK

Are genetic influences related to early-life vocabulary associated with childhood-onset neurodevelopmental conditions (NDCs)?

Yes, especially ADHD.
Genetic relationships change, however, during development.

Take home messages

- Genetic differences play a role in individual differences in vocabulary size.
- Multiple, independent genetic factors may play a role in vocabulary size during the first few years of life, suggesting developmental change.
- Genetic association patterns of early-life vocabulary measures with later-life traits, such as cognition and ADHD, match with developmental heterogeneity.

Hypothetical interpretation

Learning to speak

Emerging speech

Single-words

- No genetic association with general cognition
- Genetic association with higher ADHD risk

Speaking to learn

Some language fluency

Combining words

- Genetic association with higher general cognition
- Genetic association with lower ADHD risk

PGHC
Population Genetics of Human Communication

Beate St Pourcain

LANGUAGE & GENETICS

Simon E Fisher

Chin Yang Shapland, Philip S. Dale, Andrea G. Allegrini, Philip R. Jansen, Katherine Lange, Carol A. Wang, Angela T. Morgan, Tarunveer S. Ahluwalia, Christos Symeonides, EAGLE working group, Else Eising, Marie-Christine Franken, Elina Hypponen, Toby Mansell, Mitchell Olislagers, Emina Omerovic, Kaili Rimfeld, Fenja Schlag, Saskia Selzam, Chin Yang Shapland, Henning Tiemeier, Andrew J.O. Whitehouse, Richard Saffery, Klaus Bønnelykke, Sheena Reilly, Craig E. Pennell, Melissa Wake, Charlotte A.M. Cecil, Robert Plomin

MAX
PLA
NCK

MAX PLANCK INSTITUTE
FOR **PSYCHOLINGUISTICS**

WWW.MPI.NL