Supplementary material:

Methods:

Supplementary tables:

Figures:

Supplementary methods 1: Description of the models

To estimate the effect of vitamin D deficiency on SRS, variations of linear models were fitted to the data. The choice of model depended on the distribution of close relatives in the sample and the method that was applied to control for population structure. Three different models were fitted to the data:

1. First, a multivariate mixed linear model was employed to predict SRS from vitamin D deficiency status. Vitamin D deficiency status and parental and offspring demographic variables were fitted as fixed effects. Family membership was fitted as a random effect to account for family clustering within the sample. Members of the same family (e.g., siblings, cousins), are likely to be more similar on traits that have a genetic component and/or are influenced by family factors such as upbringing. With appropriate adjustments, these factors could inflate the association estimate between exposure and outcome variable. The model was:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\gamma} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y}$ is an *n x 1* column vector phenotypic values (SRS) for *n* individuals; $\boldsymbol{X}$ is an *n x p* matrix of the *p* predictor variables (vitamin D deficiency status and other covariates included in the model); $\boldsymbol{\beta}$ is an *p x 1* column vector of the fixed-effects regression coefficients; $\boldsymbol{Z}$ is an n *x q* design matrix for the *q* random effects (families in the sample); $\boldsymbol{\gamma}$ is a *q x 1* vector of the random effects; and $\boldsymbol{\varepsilon}$ is an *n x 1* column vector of the residuals.

All p fixed effects were: vitamin D deficiency status, parental demographic variables and offspring demographic variables. Parental variables included maternal and paternal age, maternal Body Mass Index (BMI, measured mid-gestation), smoking status of the mother during pregnancy, and educational level of the mother. Offspring variables included ethnicity of the child, age at time of SRS assessment, gestational age at birth, birth weight, and sex of the child.

To test the effect of vitamin D on SRS we fitted a full model (i.e., a model that included the vitamin D predictor), and a nested model (i.e, a model that did not included the vitamin D predictor), and then compared the fit of the two models. Goodness-of-fit of the submodels was assessed by likelihood-ratio-tests. The difference in log-likelihoods between the two models was evaluated using a χ^2-difference test. A significant χ^2-difference test implies that the effect of the vitamin D predictor is significant.

Specifically, we fitted 4 full models to test the effect of deficient vs. sufficient and insufficient vs. sufficient at mid-gestation / cord (results of these analyses are presented in Table 3). Subsequently, we fitting a series of planned sensitivity analyses:

First, we fitted a similar model as in the main analyses but reduced the sample to those offspring with European ethnicity. In these models we did not include the ethnicity variable as a covariate as the sample consisted of Europeans only (results of these analyses are presented in Supplementary Table 2).

Second, we fitted a model utilizing genome-wide genotype data to capture sample structure in the data (explained below, results are presented in Supplementary 3).

Third, we fitted a model on the full sample replacing the categorical vitamin D predictor with a continuous measure of vitamin D (results are presented in Supplementary Table 4).

Fourth, a model in which we tested whether 25OHD deficiency at two time points (Mid-gestation *and* Cord was predictive of higher SRS scores compared with 25OHD sufficiency at two time points. Similarly, we tested whether 25OHD deficiency at only one time point (Mid-gestation *or* Cord) was predictive of higher SRS scores compared with 25OHD sufficiency at two time points. Results of these two-time point analyses are presented in Supplementary Table 5.

Fifth, we fitted a similar model as in the main analyses but included season of blood sampling as a covariate in the model to account for seasonality of 25OHD concentrations.

All mixed linear models were fitted using the 'lme4' package in R. [1]

2. In the next section, we will present a detailed technical description of the method we used to utilize genome-wide genotype data to capture structure in the data. In addition, we have provided a simplified 'worked example' to assist readers not familial with recent developments in statistical genetics.

Technical Description

Within the Generation R sample, there is substantial structure resulting from geographic population structure as well as from family- and cryptic-relatedness. Geographic population structure is a result of different ancestries being represented in the cohort while family- and cryptic-relatedness result from a proportion of the mothers participating more than once and from offspring related through common close ancestors (e.g., cousins), respectively. Because the exposure variable in this study (25OHD concentration) is highly associated with

skin colour and thus with ethnic background,[4] we aimed to fit a model in which potential confounding effects related to ethnic differences between individuals were accounted for.

To this end, we first created a Genetic Relationship Matrix (GRM) for all the children for whom genome-wide genotype data were available. This GRM is an *n* x *n* symmetrical matrix where each off-diagonal element represents the genome-wide genetic relatedness between two different individuals and each diagonal element represents a persons genetic relatedness with him- or herself. For individuals *j* and *k* from *m* genotyped SNPs, the GRM can be written as follows:

$$A_{jk} = \frac{1}{m}\sum_{i=1}^{m}\frac{(x_{ij}-2p_i)(x_{ik}-2p_i)}{2p_i(1-p_i)},$$

where *p* is the frequency of the reference allele and x_i is the genotype indicator of the i^{th} SNP (x_i = 0, 1, or 2).

The GRM is fitted in a mixed linear model together with fixed effects, including Vitamin D deficiency status as well as parental and offspring demographic variables. The mixed linear model can be written as follows:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Wu} + \boldsymbol{\varepsilon}, \text{with } \mathbf{var}(\mathbf{y}) = \mathbf{V} = \mathbf{WW}'\boldsymbol{\sigma}_{\mathbf{u}}^{\mathbf{2}} + \mathbf{I}\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^{\mathbf{2}},$$

where **y** is an *n x 1* matrix of phenotypic values (SRS or cASD) with n being the sample size, **β** is a vector of fixed effects including vitamin D deficiency status and parental and offspring variables, **u** is a vector of SNP effects with $u \sim N(0, \mathbf{I}\boldsymbol{\sigma}_{\boldsymbol{u}}^{\mathbf{2}})$ and **ε** is a vector of residual effects with $\varepsilon \sim N(0, \mathbf{I}\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^{\mathbf{2}})$**.** **W** is a standardized genotype matrix with the *ij*th element $w_{ij} = (x_{ij} - 2p_i)/\sqrt{2p_1}(1-p_1)$, where x_{ij} is the number of copies of the reference allele for the i^{th} SNP of the j^{th} individual and p_i is the frequency of the reference allele. **A** is defined as **WW'**/N and $\boldsymbol{\sigma}_{\mathbf{g}}^{\mathbf{2}}$ is defined as the variance explained by all the SNPs, i.e., $\boldsymbol{\sigma}_{\mathbf{g}}^{\mathbf{2}} = \boldsymbol{N}\boldsymbol{\sigma}_{\mathbf{u}}^{\mathbf{2}}$ with N being the number of SNPs, $\mathbf{y} = \mathbf{X\beta} + \mathbf{Wu} + \boldsymbol{\varepsilon}, \text{with } \mathbf{var}(\mathbf{y}) = \mathbf{V} = \mathbf{WW}'\boldsymbol{\sigma}_{\mathbf{u}}^{\mathbf{2}} + \mathbf{I}\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^{\mathbf{2}}$, will be equivalent to $\mathbf{y} = \mathbf{X\beta} + \mathbf{g} + \boldsymbol{\varepsilon}, \text{with } \mathbf{V} = \mathbf{A}\boldsymbol{\sigma}_{\mathbf{g}}^{\mathbf{2}} + \mathbf{I}\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^{\mathbf{2}}$, where **g** is an n x 1 vector of the total genetic effect of the individuals with $\boldsymbol{g} \sim N(0, \mathbf{A}\boldsymbol{\sigma}_{\boldsymbol{g}}^{\mathbf{2}})$ and **A** is interpreted as the genetic relationship matrix (GRM).

Application of a mixed linear model to human genetic data was proposed by Yang et al [5,6] in order to estimate the variance of a trait that can be explained by common SNPs (i.e., genetic variance). Rather than using this method to estimate the genetic variance, we use this method to test the effect of a fixed effect in the model thereby accounting for population

structure that is captured by the GRM, analogous to a related method proposed by Kang et al., [7] and discussed in Yang et al [8] in the context of genome-wide association studies. For simplicity we assume that there are no covariates, we estimate the effect of the predictor (i.e., vitamin D deficiency status) in a Generalized Least Squares (GLS) procedure as $\widehat{\boldsymbol{b}} = \mathbf{x}^{\boldsymbol{T}}\boldsymbol{G}^{-\mathbf{1}}\mathbf{y}^{*}/\left(\mathbf{x}^{\boldsymbol{T}}\boldsymbol{G}^{-\mathbf{1}}\mathbf{x}\right)$, with $\boldsymbol{var}(\widehat{\boldsymbol{b}}) = \mathbf{1}/\left(\mathbf{x}^{\boldsymbol{T}}\boldsymbol{G}^{-\mathbf{1}}\mathbf{x}\right)$ and **x** being a vector with the predictor values, and $\mathbf{y}^{*}$ being the adjusted phenotype $\mathbf{y}^{*} = \mathbf{y} - (\mathbf{1}^{\boldsymbol{T}}\boldsymbol{V}^{-\mathbf{1}}\mathbf{1})^{-\mathbf{1}}\mathbf{1}^{\boldsymbol{T}}\mathbf{y}.$ Significance of the vitamin D predictor was tested with a Chi-square test with $\chi^2 = \hat{b}^2/var(\hat{b})$. The models were fitted in GCTA, [6] the software package that was specifically designed to estimate genetic variance.

Results of the mixed model analyses using GCTA are presented in Supplementary Tables 4. The two time-point analyses (i.e., 25OHD deficient at two time points vs 25OHD sufficient at two time points and 25OHD deficient at only one time points vs 25OHD sufficient at two time points) was also done in a mixed model setting using the GRM to capture sample structure; these results are presented in Supplementary Table 5.

Simplified 'Worked Example'

Modern technology allows for many hundreds of thousands of common genetic variants (Single Nucleotide Polymorphisms; SNPs) to be determined quickly and cheaply. By looking at how similar these SNPs are between participants, we can determine if two individuals are closely related (e.g. siblings, cousins), or share common ancestry (e.g. are descendants from a particular ancestral populations). In research, it is important to understand these variables and adjust for them where needed. In the current example skin colour is linked to vitamin D status, thus variables related to ancestry (e.g. ethnicity/race) can influence this variable. Similarly, we need to be careful to adjust models if siblings are included. We can use variables like "ethnicity" and relatedness (e.g. two participants may have the same mother) to help make these adjustments, but by including information from SNPs, we can use *objective genetic information* to make these adjustments. In this setting, we are not wanting to adjust the analyses based on any particular SNP, but for an overall, genome-wide degree of relatedness. In this study, we used a recently developed technique that uses pairwise genome-wide comparisons for every participant with every other participant (e.g. subject 1 vs. subject 2, subject 1 vs. subject 3, etc. until all possible pairwise combinations are examined). By building a table (or matrix) of these results, we can summarize the degree of

relationship for the entire sample. It is relatively simple to use this matrix to adjust statistical models based on the pairwise genomic relatedness.

Here is a simplified example. We assume we have a sample of N=6 individuals, two of whom share a father (but not a mother, subjects 1 and 2). Three of the individuals are from European ancestry (subjects 1, 2, and 3), one individual is of Cape Verdian ancestry (subject 4), and two individuals are from Moroccan ancestry (subjects 5 and 6). In our example, subjects 1 and 2 will share approximately 25% of their genotypes while subjects 1 and 4 will share approximately 1% of their genotypes.

With regards to the different ancestries in our example, the calculated genomic relationships will be very close to zero for individuals who are related through very distant common ancestors and will be higher if individuals are related through more recent ancestors. That is, a pair of individuals with the same ethnic background (e.g., Moroccan ancestry, subjects 5 and 6) will be genetically more similar compared with a pair of individuals from different ethnic backgrounds.

For ease of computation, we store all pair-wise genomic relationships in an NxN matrix. The off-diagonal elements of matrix represent the genomic relationships for all pairs of subjects whereas the diagonal elements of matrix represent the genomic relationship of a person with him- or herself. We see that pair-wise genomic relationships for pairs of individuals with the *same ancestry* are larger than pair-wise genomics relationships for pairs of individuals with *different ancestries*.

	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6
Subject 1	1					
Subject 2	0.25	1				
Subject 3	0.08	0.06	1			
Subject 4	0.01	0.02	0.01	1		
Subject 5	0.02	0.03	0.02	0.03	1	
Subject 6	0.03	0.01	0.01	0.04	0.07	1

Using standard statistical models, we can then make adjustments that take into account the degree of relatedness between individuals (e.g. weighting by the inverse of the degree of relatedness).

Supplementary methods 2: Assessment of vitamin D status

Total 25OHD was reported as the sum of 25-hydroxyvitamin D2 (25OHD2) and 25-hydroxyvitamin D3 (25OHD3) species measured in plasma using a modification of a method previously described. [9] Briefly, 50µL milli-Q water and 500µL of acetonitrile (ACN) containing *6,19,19*-[$^2H_3$]-25OHD2 and *6,19,19*-[$^2H_3$]-25OHD3 at 10 nmol/L each were added to 3µL serum, sonicated, vortexed and centrifuged. The supernatant was filtered using a TiO_2/ZrO_2 filter plate (Glygen, USA) and evaporated to dryness. Samples were derivatised using 4-phenyl-1,2,4-triazoline-3,5-dione (PTAD) and reconstituted in ACN:H_2O (1:3) prior to analysis. Samples were quantified using isotope dilution liquid chromatography-tandem mass spectrometry. The analytical system was comprised of a Shimadzu Nexera UPLC coupled to an AbSciex 5500 QTRAP equipped with an APCI source. Chromatographic separation was achieved using a Kinetex XB-C18 column (50 ×2.1mm, 1.7µm; Phenomenex, USA), and 72% acetonitrile/32% aqueous 0.1% formic acid at a flow rate of 0.5 mL/min.

Linearity of 25OHD concentration was assessed using matrix-matched calibration standards, with R^2 values of >0.99 across the calibration range (10 – 125 nmol/L). Assay accuracy was assessed at four concentration levels for 25OHD3 (48.3, 49.4, 76.4, 139.2 nmol/L) and a single level for 25OHD2 (32.3 nmol/L) using certified reference materials purchased from the National Institute of Standards and Technology (NIST) (NIST SRM 972a Levels 1-4), and was excellent at all concentration levels tested (<10% and <17%, respectively). Assay repeatability was assessed via replicate analysis of an independent reference material (NIST SRM1950, 61.9 nmol/L 25OHD3). Inter-assay imprecision was <11% (n=343). The method limit of quantification was 1 and 5 nmol/L for 25OHD3 and 25OHD2, respectively. All samples were analysed at the Queensland Brain Institute in Brisbane, Australia, from July 2013 to August 2014.

Supplementary methods 3: Genotyping and Quality Control

Genotyping was performed using Illumina HumanHap 610 or 660 Quad chips, depending on collection time, following manufacturer protocols. Intensities were obtained from the BeadArray Reader and genotype calling was performed on normalized intensities using the Illumina Genome Studio software version 1.1.0.28426. Prior to Quality Control (QC), genotype data from the HumanHap 610 array and the 660 Quad array were merged retaining only SNPs in common to both arrays. QC was applied using PLINK[10] removing markers with a call rate lower than 80% (first round) and 95% (second round), minor allele frequency smaller than 0.001, differential missingness between the two genotyping arrays ($p< 10^{-7}$) and deviation from Hardy-Weinberg equilibrium ($p< 10^{-7}$). Samples were removed if duplicated, if sex was discordant between self-report and predicted by the genetic data, if genotype call-rate was smaller than 95% (first round) and 97.5% (second round), and if heterozygosity rate was high (>4 SD of the mean heterozygosity of all samples). Analyses involving genome-wide genotype data in the current study were based on 3,234 individuals on 518,245 single nucleotide polymorphisms (SNPs).

Supplementary methods 4: Social Responsiveness Scale (SRS)[11], Short Form Items

Probes were rated on a 4 point Likert scale:

- 0 (not true)
- 1 (sometimes true)
- 2 (often true)
- 3 (almost always true)

1. Is unable to pick up on any of the meaning of conversations of older children or adults.
2. Is slow or awkward in turn-taking interactions with peers.
3. Is able to understand the meaning of other people's tone of voice and facial expressions.
4. Avoids eye contact, or has unusual eye contact.
5. Does not attempt to interact with the other children when on the playground or in a group with other young children.
6. Has strange ways of playing with toys.
7. Has more difficulty than other children with changes in his/her routine.
8. Is regarded by other children as odd or weird.
9. Has trouble keeping up with the flow of a normal interaction with other children.
10. Has difficulty "relating" to peers.
11. Has a restricted (or unusually narrow) range of interests.
12. Is imaginative, good at pretending (without losing touch with reality).
13. Has repetitive odd behaviors such as hand flapping or rocking.
14. Responds to clear, direct questions in ways that don't seem to make any sense.
15. Talks to people with an unusual tone of voice (for example, talks like a robot or like he/she is giving a lecture).
16. Concentrates too much on parts of things rather than "seeing the whole picture" (for example, spins the wheels of a toy car, but doesn't play with it as a car, or plays with doll's hair but not with the whole doll).
17. Is inflexible, has a hard time changing his/her mind.
18. Gives unusual or illogical reasons for doing things.

Supplementary Table 1. Raw counts and proportions of parental and offspring variables with and without imputed data

	Without imputed data	With imputed data		Without imputed data	With imputed data
Parental variables	**N (Prop.)**	**N (Prop.)**	**Offspring variables**	**N (Prop.)**	**N (Prop.)**
Age mother at intake	4229	4229	**Sex child**	4229	4229
< 25 years	443 (0.10)	443 (0.10)	Boy	2 114 (0.50)	2 114 (0.50)
>=25 & < 30	1 090 (0.26)	1 090 (0.26)	Girl	2 115 (0.50)	2 115 (0.50)
>=30 & < 35	1 905 (0.45)	1 905 (0.45)	*Missing*	*0*	
> 35 years	791 (0.19)	791 (0.19)	**Gestational age at time of birth**	4229	4229
Missing	0		>= 41.5 weeks	627 (0.15)	627 (0.15)
Age biological father at intake	3899	4229	>= 40.5 weeks & < 41.5 weeks	1 050 (0.25)	1 050 (0.25)
< 30 years	869 (0.22)	964 (0.23)	>= 39.5 weeks & < 40.5 weeks	1 173 (0.28)	1 173 (0.28)
>=30 & < 35	1 613 (0.41)	1 717 (0.41)	>=38.5 weeks & < 39.5 weeks	748 (0.18)	748 (0.18)
>=35 & < 40	970 (0.25)	1 071 (0.25)	< 38.5 weeks	631 (0.15)	631 (0.15)
>=40	447 (0.11)	477 (0.11)	*Missing*	*0*	
*Missing**	330 (0.08)		**Birth weight infant**	4229	4229
BMI mother	4095	4229	>= 4 000 grams	645 (0.15)	645 (0.15)
< 20	146 (0.04)	155 (0.04)	< 4 000 grams & >= 3500	1 434 (0.34)	1 434 (0.34)
>= 20 - < 25	2 110 (0.52)	2 148 (0.51)	< 3 500 grams & >= 3000	1 413 (0.33)	1 413 (0.33)
>= 25 - < 30	1 355 (0.33)	1 424 (0.34)	< 3 000 grams	737 (0.17)	737 (0.17)
> 30	484 (0.12)	502 (0.12)	*Missing*	*0*	
*Missing**	134 0.03)				
Smoking mother during pregnancy	3804	4229			
No	3 369 (0.89)	3 744 (0.89)			
Yes	435 (0.11)	485 (0.11)			
Missing	425 (0.10)				
Educational level mother	4110	4229			
Primary school or less	253 (0.06)	267 (0.06)			
Secondary school phase 1	423 (0.10)	436 (0.10)			

Secondary school phase 2	1 174 (0.29)	1 216 (0.29)
Higher education phase 1	994 (0.24)	1 015 (0.24)
Higher education phase 2	1 266 (0.31)	1 295 (0.31)
*Missing**	119 (0.03)	

Notes: Counts and proportions are based on all individuals with data on vitamin D measured from Mid-gestation and/or Cord blood and data on the SRS; Prop. = proportion; Parental variables and Offspring variables except for 'Ethnicity of child' were imputed for missing data (see Methods section for details on imputation protocol); * The count of missing values and related proportions (which were based on adjusted totals combining observed plus missing counts). All other proportions based on observed data.

Supplementary Table 2. Distribution of midgestation and cord 25OHD across season of testing and ethnic background of the child

		autumn	%	spring	%	summer	%	winter	%	European	%	Non-European	%
midgestation	Sufficient	541	61	519	50	759	76	423	44	1969	69	273	27
	Insufficient	203	23	317	31	166	17	311	33	706	25	289	29
	Deficient	137	16	202	19	70	7	219	23	185	6	439	44
	Total	881	100	1038	100	995	100	953	100	2860	100	1001	100
cord	Sufficient	204	29	96	13	319	43	65	10	647	30	37	5
	Insufficient	310	44	298	40	267	36	261	39	966	45	169	24
	Deficient	184	26	350	47	150	20	347	52	520	24	508	71
	Total	698	100	744	100	736	100	673	100	2133	100	714	100

Supplementary Table 3. Means and standard deviations for weighted SRS scores for different midgestation and cord 25OHD strata

Midgestation	N	Mean (sd) SRS
Sufficient	2242	0.20 (0.20)
Insufficient	997	0.24 (0.24)
Deficient	628	0.32 (0.28)
Cord		
Sufficient	684	0.19 (0.20)
Insufficient	1136	0.21 (0.22)
Deficient	1031	0.27 (0.25)

Notes: Scores of the 18 item SRS are gender-weighted, such that the recommended cut-offs for screening in population-based settings are consistent with weighted scores of 1.078 for boys and 1.000 for girls; N = number of observations; sd = standard deviation

Supplementary Table 4. Association between Mid-gestation and cord 25OHD deficiency and high-SRS (above the cut-off) versus low-SRS (below cut-off)

	N	OR	lower 95% CI	upper 95% CI	χ2(1)	p
midgestation 25OHD deficient vs sufficient	2711	3.80	1.73	8.09	10.85	0.001
midgestation 25OHD insufficient vs sufficient	3042	1.74	0.92	3.26	2.88	0.09
cord 25OHD deficient vs sufficient	1634	1.18	0.44	3.31	0.11	0.74
cord 25OHD insufficient vs sufficient	1720	1.15	0.48	2.96	0.09	0.76

Notes: Estimates are based on general linear model; a Firth correction was applied to the model; N = sample size; OR = odds ratio, 95% CI = 95% confidence interval; $X^2(1)$ = chi-squared test statistic with 1 degree of freedom; p = p-value; * = significant at alpha of 0.05; covariates included in the model are: ethnicity of child; sex child, birth weight child, gestational age at time of birth, age mother at intake, age father at intake, smoking mother during pregnancy, educational level mother, and BMI mother at mid-gestation; Deficient is 25OHD concentrations <25 nmol/L, Insufficient is 25OHD concentrations 25 to <50 nmol/L; Sufficient is 25OHD concentrations ≥ 50 nmol/L. Note that only one sibling per family was included in this analyses; individuals with SRS scores above the cut-off were scored as 1, individuals with SRS scores below the cut-off were scored as 0.

Supplementary Table 5. Association between Mid-gestation and cord 25OHD deficiency and Social Responsiveness Scale in children with European ethnic background only

	N	**β**(s.e.)	$\chi^2(1)$	p
Mid-gestation - 25OHD deficient vs sufficient	**2 154**	**0.06 (0.02)**	**14.34**	**0.0002*** **
Mid-gestation - 25OHD insufficient vs sufficient	**2 675**	**0.02 (0.01)**	**4.66**	**0.03***
Cord - 25OHD deficient vs sufficient	1 167	0.02 (0.01)	3.83	0.05
Cord - 25OHD insufficient vs sufficient	1 613	0.01 (0.01)	2.13	0.14

Notes: Estimates are based on multivariate mixed linear model; family membership was fitted as a random effect in the model; N = sample size, β = effect size; s.e. = standard error; $X^2(1)$ = chi-squared test statistic with 1 degree of freedom; p = p-value; * = significant at alpha of 0.05; *** = significant at alpha of 0.001; covariates included in the model are: sex of the child, age child at time of SRS assessment, birth weight of the child, gestational age at time of birth, age mother at intake, age father at intake, smoking mother during pregnancy, educational level mother, and BMI mother at mid-gestation.

Supplementary Table 6. Association between Mid-gestation and cord 25OHD deficiency and Social Responsiveness Scale including a genetic component to adjust for sample structure

	N	**β**(s.e.)	$\chi^2(1)$	p
Mid-gestation - 25OHD deficient vs sufficient	**2 083**	**0.05 (0.01)**	**3.59**	**0.0003*** **
Mid-gestation - 25OHD insufficient vs sufficient	**2 410**	**0.02 (0.01)**	**2.21**	**0.03***
Cord - 25OHD deficient vs sufficient	**1 590**	**0.03 (0.01)**	**2.04**	**0.04***
Cord - 25OHD insufficient vs sufficient	1 684	0.01 (0.01)	1.13	0.26

Notes: Estimates are based on a multivariate mixed linear model; the genetic relationship matrix (GRM) of genome-wide genotype data was fitted as a random effect in the model to account for sample structure; N = sample size, β = effect size; s.e. = standard error; $X^2(1)$ = chi-squared test statistic with 1 degree of freedom; p = p-value; * = significant at alpha of 0.05; *** = significant at alpha of 0.001; covariates included in the model are:, sex child, age child at time of SRS assessment, birth weight child, gestational age at time of birth, age mother at intake, age father at intake, smoking mother during pregnancy, educational level mother, and BMI mother at mid-gestation.

Supplementary Table 7. Association between Mid-gestation and cord 25OHD deficiency and Social Responsiveness Scale correcting for season of 25OHD measurement (midgestation or birth)

	N	B (s.e.)	χ2(1)	p
midgestation 25OHD deficient vs sufficient	**2 866**	**0.06 (0.01)**	**20.70**	**<0.001*** **
midgestation 25OHD insufficient vs sufficient	**3 237**	**0.02 (0.01)**	**5.29**	**0.02***
cord 25OHD deficient vs sufficient	**1 712**	**0.04 (0.01)**	**9.68**	**0.002****
cord 25OHD insufficient vs sufficient	**1819**	**0.03 (0.01)**	**5.92**	**0.02***

Notes: Estimates are based on multivariate mixed linear model; family membership was fitted as a random effect in the model; N = sample size, β = effect size; s.e. = standard error; $X^2(1)$ = chi-squared test statistic with 1 degree of freedom; p = p-value; * = significant at alpha of 0.05; ** = significant at alpha of 0.01; *** = significant at alpha of 0.001; covariates included in the model are: ethnicity of child; sex child, age child at time of SRS assessment, birth weight of the child, gestational age at time of birth, age mother at intake, age father at intake, smoking mother during pregnancy, educational level mother, BMI mother at mid-gestation and season of blood draw.

Supplementary Table 8a. Responder analyses for categorical variables

	Total sample		SRS sample		Non-SRS sample	
	N	**Proportion**	**N**	**Proportion**	**N**	**Proportion**
Ethnicity offspring						
Cape Verdean	255	0.03	81	0.02	174	0.03
European	5765	0.62	3120	0.74	2645	0.53
Moroccan	630	0.07	153	0.04	477	0.1
Other	1122	0.12	372	0.09	750	0.15
Surinamese	720	0.08	233	0.06	487	0.1
Turkish	732	0.08	263	0.06	469	0.09
χ2-test distribution SRS sample vs non-SRS *sample*	*χ2(5) 451.882, p < 0.001*					
Educational level mother						
Primary school or less	254	0.04	105	0.03	149	0.07
Secondary school phase 1	549	0.09	309	0.08	240	0.11
Secondary school phase 2	1913	0.31	1141	0.28	772	0.35
Higher education phase 1	1710	0.27	1158	0.29	552	0.25
Higher education pahse 2	1830	0.29	1336	0.33	494	0.22

χ2-test distribution SRS sample vs non-SRS *sample*	*χ2(4) 161.268, p < 0.001*					
Smoking during pregnancy mother						
yes	1084	0.15	435	0.11	649	0.2
no	5913	0.85	3369	0.89	2544	0.8
χ2-test distribution SRS sample vs non-SRS *sample*	*χ2(1) 104.121, p< 0.001*					
Gender child						
boy	4938	0.51	2114	0.5	2824	0.51
girl	4809	0.49	2115	0.5	2694	0.49
χ2-test distribution SRS sample vs non-SRS *sample*	*χ2(1) 1.309, p = 0.25*					

Supplementary Table 8b. Responder analyses for continuous variables

	Total sample			**SRS sample**			**non-SRS sample**			*t-test for mean differences SRS sample vs non-SRS sample*	
	mean	sd	n	mean	sd	n	mean	sd	n	t	p
Age biological mother	29.92	5.37	9900	31.13	4.61	4229	29.02	5.72	5671	*20.33,*	*<0.001*
Age biological father	33.03	6.04	7564	33.77	5.41	3899	32.24	6.56	3665	*11.02,*	*<0.001*
Gestational Age in weeks	39.7	2.01	9662	39.96	1.7	4229	39.5	2.2	5433	*11.56,*	*<0.001*

Notes: Total sample = is the total number of individuals with data available; SRS-sample = the number of individuals with SRS data available; non-SRS sample = the number of individuals for whom no SRS data are available; N = sample size; sd = standard deviation; vs = versus; *χ2(1) = chi-square test statistic with degrees of freedom in between brackets;* t = t-test statistic, p = p-value.

References

1. *lme4: Linear Mixed-Effects Models using 'Eigeh' and S4* [computer program]. Version 1.1-102015.
2. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. *Stat Med.* Aug 30 2002;21(16):2409-2419.
3. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol.* Nov 1985;122(5):904-914.
4. Vinkhuyzen AA, Eyles DW, Burne TH, et al. Prevalence and predictors of vitamin D deficiency based on maternal mid-gestation and neonatal cord bloods: The Generation R Study. *J Steroid Biochem Mol Biol.* Sep 15 2015.
5. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nature genetics.* Jul 2010;42(7):565-569.
6. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* Jan 7 2011;88(1):76-82.
7. Kang HM, Sul JH, Service SK, et al. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet.* Apr 2010;42(4):348-354.
8. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics.* Feb 2014;46(2):100-106.
9. Eyles D, Anderson C, Ko P, et al. A sensitive LC/MS/MS assay of 25OH vitamin D3 and 25OH vitamin D2 in dried blood spots. *Clin Chim Acta.* May 2009;403(1-2):145-151.
10. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* Sep 2007;81(3):559-575.
11. Constantino JN. *Social Responsiveness Scale (SRS), Manual*. Los Angeles: Western Psychological Services; 2002.
12. Falconer DF, Mackay TFC. *Introduction to quantitative genetics*. Essex: Longman; 1996.