

### *Early life socioeconomic conditions*

Parental occupational position defined as the highest household occupation was used as an indicator of socioeconomic conditions in early life (main analysis). In YFS, the occupational position (manual, lower-grade non-manual and higher-grade non-manual, farmer) of the head of the household was assessed in 1980 (study's baseline). In SKIPOGH, participants reported the profession of their father when they were children at the follow-up survey. Paternal occupational position is a commonly used indicator of socioeconomic conditions in early life<sup>1</sup> and is closely related to parental occupational position in the Swiss context until the early 1980s. Indeed, according to census data<sup>2</sup>, in 1970 the percentage of non-working women living in a household with children (0-13 years) was 73%, and the percentage of those working full time was 8%. In 1980, the percentage of non-working women living in a household with children (0-13 years) was 67%, and the percentage of those working full time was 12%. SKIPOGH participants were born on average in 1963, with 75% of participants being born by 1978.

Parental education defined as tertiles of a score computed from paternal and maternal education was also used as an exposure (sensitivity analysis). In YFS, years of education of both the mother and the father were assessed in 1980, and they were summed up to estimate a score of parental education. In SKIPOGH, participants reported the highest attained levels of education of their mother/father into 9 levels (from no diploma to university degree). The score was computed by summing up father/mother levels so ranging between 2 (lowest) and 18 (highest).

### *Systemic inflammatory levels*

In YFS, serum CRP was assessed using an automated analyzer (Olympus AU400; Tokyo, Japan) and a highly sensitive turbidimetric immunoassay kit ('CRP-UL'-assay; Wako Chemicals, Neuss, Germany). Detection limit of the assay was 0.06 mg/L. The inter-assay coefficient of variation was 3.3% at 1.5 mg/L and 2.7% at 2.5 mg/L.

In SKIPOGH, CRP levels in plasma were measured at follow-up through standard immunoturbidimetry with different detection limits (1 mg/L or 2 mg/L or 3 mg/L) depending on the center or diagnostic machine used (Cobas 8000 Roche Diagnostics, Modular Roche Diagnostics, Beckmann). Inter-assay coefficients of variation were 7.5% at 10 mg/L for N=317 (Beckmann), 3.3% at 1.5 mg/L for N=372 (Modular Roche), and 4.0% at 0.9 mg/L for N=344 (Cobas 8000 Roche). Since ~60% of SKIPOGH participants had CRP values below the detection limits, we imputed left censored values via censored maximum likelihood multiple imputation<sup>3</sup>. This method allows to accommodate different values of detection limits in multiple batches and in simulation studies performed well even in heavy censoring scenarios as in our study<sup>3</sup>. We run 30 imputed data sets with a model of CRP including paternal occupational position, BMI, age, sex, and center as predictors (*lodi* R package)<sup>3</sup>.

#### *DNA methylation and transcriptome*

Laboratory procedures. In YFS, 2.5 mL of whole blood was collected into PaXgene tubes (PreAnalytix, Hombrechtikon, Switzerland). Each tube was inverted 8–10 times and then stored at room temperature for at least 2 h. PaXgene tubes were frozen and stored for <1 year at –80 °C. After thawing, tubes were stored at RT 2–12 h as instructed by the manufacturer. RNA was then isolated with the PAXgene Blood RNA Kit (Qiagen) with the DNase Set according to the manufacturer's instructions. A QiaCube isolation robot was used. The transcription levels were analyzed with an Illumina HumanHT-12 version 4 Expression BeadChip (Illumina Inc.) containing 47,231 expression and 770 control probes. In brief, 200 ng of RNA was reverse-transcribed into cDNA and biotin-UTP-labeled using the Illumina TotalPrep RNA Amplification Kit (Ambion); 1500 ng of cDNA was then hybridized to the Illumina HumanHT-12 v4 Expression BeadChip. The BeadChips were scanned with the Illumina HiScan system. Raw Illumina probe data were exported from Beadstudio and analyzed in R 3.1.2 using the Bioconductor 3.0 packages. The transcriptomic data were processed using a nonparametric background correction, followed by quantile normalization with control and expression probes, and a log<sub>2</sub> transformation. The transcriptomic analysis was successful for 1,650 individuals.

In YFS, leukocytes' DNA was obtained from EDTA-blood samples using a Wizard® Genomic DNA Purification Kit (Promega Corporation, Madison, WI, USA) according to the manufacturer's instructions. In 179 YFS participants genome-wide DNA methylation was interrogated through the Illumina Infinium HumanMethylation450 BeadChip, while in 1,369 individuals DNA methylation was measured through the Illumina Infinium MethylationEPIC BeadChip (1,548 individuals in total). Raw data were processed according to the CPACOR pipeline<sup>4</sup>.

In SKIPOGH, whole blood was collected in PAXgene tubes (PreAnalytiX, Hombrechtikon, Switzerland), handled according to the manufacturer's instructions and stored at -80 °C. RNA was extracted using the Qiagen PAXgene Blood RNA kit (Qiagen, Venlo, Netherlands) with quality assessment and quantification by Qbit (ThermoFisher, Waltham, Massachusetts). RNA was sequenced via Illumina HiSeq 2000. Mapping of reads was performed with GEM algorithm<sup>5</sup> and gene counts quantification via QTLtools<sup>6</sup>. Counts were transformed through the *voom*<sup>7</sup> procedure and normalized them via trimmed mean of *M*-values. The transcriptomic analysis was successful for 723 individuals.

DNA was extracted using standard methods on a bead-based KingFisher Duo robot extraction system (ThermoFisher, Waltham, Massachusetts). DNA quality assessment and quantification was performed using a Nanodrop system (ThermoFisher, Waltham, Massachusetts). For bisulfite conversion, the protocol started with ~1.2ug of DNA extracted. For the PCR step: alternative incubation conditions was performed when using the Illumina Infinium® Methylation Assay (Appendix page 6 of bisulfite conversion protocol pdf). The final elution was done with 8uL of M-Elution Buffer. In 250 SKIPOGH participants genome-wide DNA methylation was interrogated through the Illumina Infinium HumanMethylation450 BeadChip, while in 451 individuals DNA methylation was measured through the Illumina Infinium MethylationEPIC BeadChip (701 individuals in total). Raw data were processed according to the CPACOR pipeline<sup>4</sup>.

Statistical analyses. For mediation analyses, in SKIPOGH gene transcription's residuals were obtained by removing the effects of RNA integrity number and mean GC. For Bayesian network scoring analyses, we estimated gene transcription's residuals by removing the effects of age, sex, centre of blood collection and family membership (random intercept, only in SKIPOGH). In analyses involving DNA methylation data, we transformed the proportion of methylated cells into *M*-values and then estimated residuals after adjustment for chip type (Illumina Infinium HumanMethylation450 or MethylationEPIC BeadChips), first thirty principal components of control probes, age, sex, centre of blood collection and family membership (random intercept, only in SKIPOGH).

#### *Summary statistics databases*

The eQTLGen database includes summary statistics for 153,437,930 associations of 10,823,015 SNPs nearby (<1Mbp) 16,989 genes. Associations between SNPs and gene transcription levels were computed and pooled across 37 cohorts for a total of 31,684 adult individuals. Gene transcription levels were profiled by arrays or RNA sequencing (20.3% of individuals). SNPs with a gene-level *P* value <  $1.9 \times 10^{-5}$  (equivalent to permutation false discovery rate < 0.05) are considered as *cis* expression quantitative trait loci (*cis* eQTLs)<sup>8</sup>. The CHARGE database includes summary statistics for 10,019,203 SNPs from pooled associations of 49 1000G imputed-genotype GWAS (N=148,164 adults). Serum CRP was measured in mg/L and natural-log transformed<sup>9</sup>.

#### *2-sample Mendelian Randomization methods*

First, the SNP-to-gene transcription and SNP-to-CRP associations were combined in a random effect meta-analysis using inverse variance weighting. This method takes into account the potential heterogeneity of multiple variants, but provides biased estimates when the net pleiotropic effect is different from zero or variant's association with the exposure is dependent in magnitude from its pleiotropic effects (violation of so-called InSIDE - Instrument Strength Independent of Direct Effect - assumption)<sup>10</sup>. Second, we conducted MR Egger regression, which provides a consistent estimate even if the average pleiotropic effect is different from zero, but requires the InSIDE assumption. Third,

we conducted penalized weighted median regression analyses, which can provide a consistent estimate for the true causal effect when up to half of the weight in the MR analysis comes from genetic variants violating all assumptions. Given the orthogonality in assumptions, if the results from these three MR methods are broadly consistent, then our causal inference is strengthened.

For each gene, available *cis* eQTLs were clumped by imposing a linkage disequilibrium  $r^2 < 0.25$ . The correlation between instruments was taken into account by running inverse-variance weighted methods<sup>11</sup>. All 2sMR methods were run with a minimum of 10 instruments. We estimated *F*-statistics to assess instruments strengths (*F*-stat > 10 is suggestive of strong instruments).

We computed the *q* value<sup>11</sup> from the *P* values estimated via the three 2sMR methods. We declared a gene transcription to be driving CRP levels (at any period before adulthood) when  $q \leq 0.05$  in all three 2sMR methods. Mendelian randomization was run via the *TwoSampleMR*<sup>13</sup> and *Mendelian Randomization*<sup>12</sup> R packages. *Q* values were estimated via the *qvalue* R package.

#### *Counterfactual mediation approach*

We adopted a counterfactual mediation framework to disentangle direct and indirect effects of the exposure(s) on the outcome, via multiple mediators. Contrary to traditional path analysis based on parametric structural equation modelling, mediation analysis based on counterfactuals relies on formal causal arguments and it allows path tracing even with binary outcomes, whereas the traditional path analysis rely on stringent parametric constraints<sup>14</sup>. In this study, we fitted marginal natural effect models, a class of marginal structural models for parametrizing and estimating so-called natural direct and indirect effects<sup>15</sup>.

Let nested counterfactual  $Y(a, M(a^*))$  denote the CRP value that would have been observed if exposure were set to low parental occupational position (*a*) and mediators *M* to the value they would have taken if exposure were set to high parental occupational position (*a*\*). Denoting as *C* the set of confounders and hypothesizing no product terms, we posited the following natural effect model

$$\log[E\{Y(a, M(a^*))|C\}] = \beta_0 + \beta_1 a + \beta_2 a^* + \beta_3 C$$

From it, we can simultaneously estimate the natural direct effect (% of change in geometric mean of Y) as

$$(\exp\{\beta_1 (a - a^*)\} - 1) * 100$$

and the natural joint indirect effect (% of change in geometric mean of Y) as

$$(\exp\{\beta_2 (a - a^*)\} - 1) * 100.$$

The sum  $(\exp\{\beta_1 (a - a^*) + \beta_2 (a - a^*)\} - 1) * 100$  measures the total effect and the ratio  $\beta_2 / (\beta_1 + \beta_2)$  measures the proportion mediated.

Under certain identifying conditions (no unmeasured confounders of exposure-outcome, exposure-mediators, mediators-outcome, and mediator-mediator induced by exposure relations), the effect of exposure on outcome mediated simultaneously by all mediators (so-called joint indirect effect) can be estimated<sup>16</sup>. Importantly, identification of joint indirect effects is not dependent on the true causal order of the mediators<sup>17</sup>. In other words, joint indirect effects are robust to potential misspecifications of the causal order among the examined mediators. This property is useful in our case since the interrelations between the transcription levels of different genes cannot be determined unequivocally.

In counterfactual terms, the effect mediated by all genes' transcription taken together or *en-bloc* (see Figure 2A) represents the effect on CRP values when altering the levels of the mediators from levels observed if exposed to high parental occupational position to levels that would have been observed at low parental occupational position while otherwise remaining exposed to low parental occupational position.

The aforementioned joint effects correspond to parameters of natural effects models (see above) and were estimated via an imputation model of the nested counterfactuals<sup>18</sup>. This was performed in three

steps by: i) fitting linear regression models for log-transformed CRP with parental occupational position, the mediators and confounders as explanatory variables; ii) imputing the counterfactual log-CRP for each combination of unobserved (counterfactual) exposures and observed mediators' levels; iii) fitting a linear natural effect model for imputed log-CRP. In all cases, we specified models for the outcome imputation to reflect the structure of the natural effects models<sup>18</sup>. Under the assumption of an outcome imputation model correctly specified and congenial with the natural effect model<sup>18</sup>, this estimation approach is appealing because it is parsimonious in that it requires specifying only one model even in the presence of multiple mediators. Furthermore, in empirical simulations, it provided adequate estimates compared to inverse probability and doubly robust estimators<sup>18</sup>.

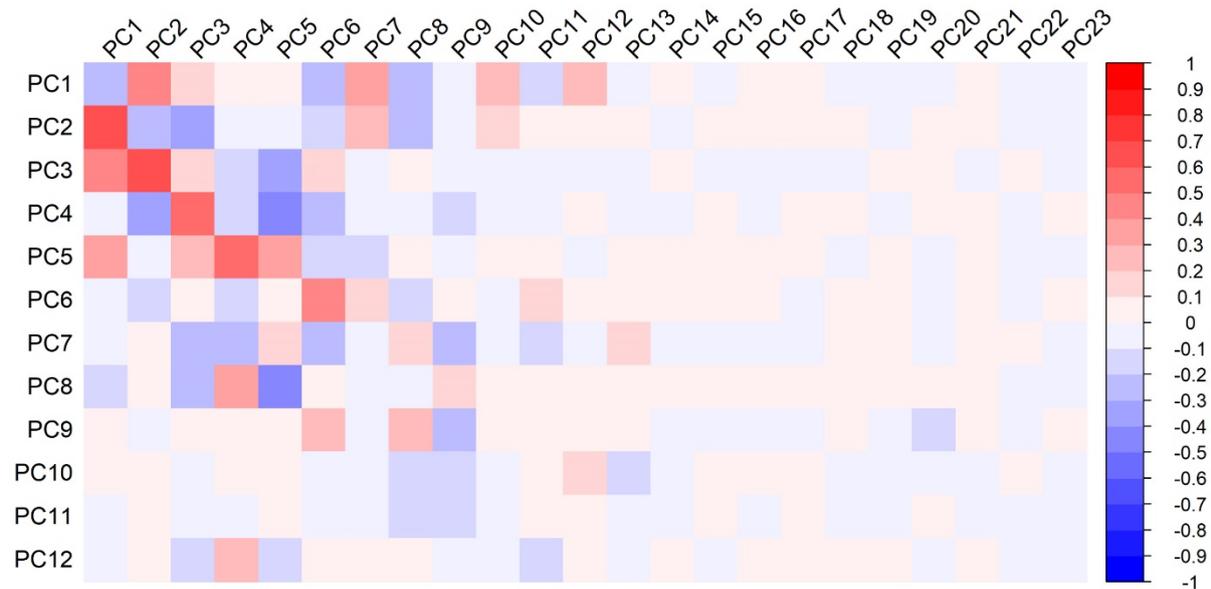
Mediation analyses were performed using *medflex*<sup>19</sup> R package.

## References

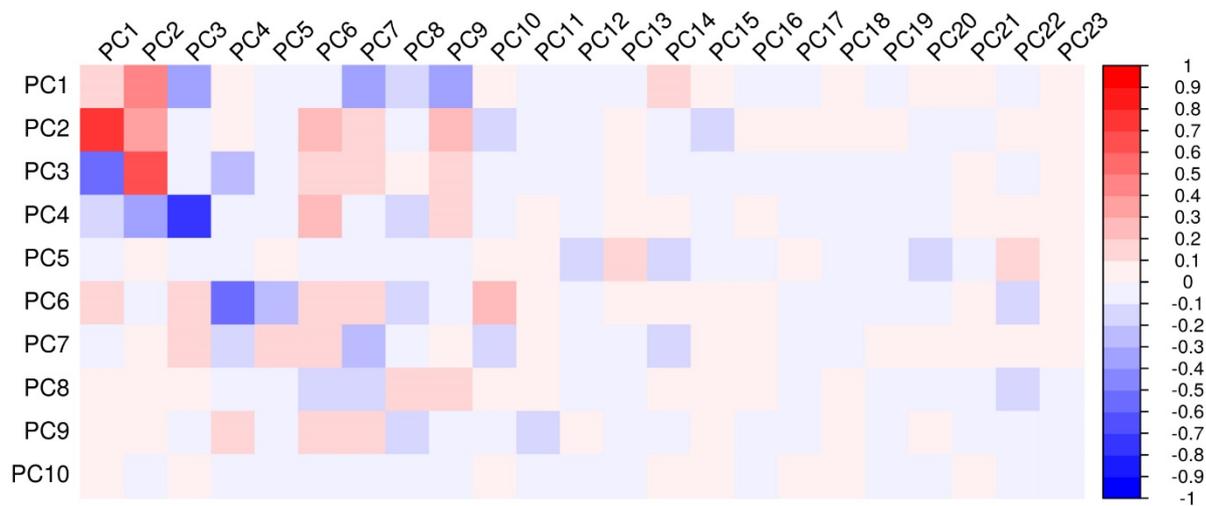
1. B. Galobardes, J. Lynch, G. Davey Smith, Measuring socioeconomic position in health research. *Br Med Bull* **81-82**, 21-37 (2007).
2. Swiss Federal Statistical Office. Data from "Activité professionnelle, tâches domestiques et familiales". Available at <https://www.bfs.admin.ch/bfs/fr/home/statistiques/population/familles.html>
3. J. Boss, *et al.*, Estimating outcome-exposure associations when exposure biomarker detection limits vary across batches. *Epidemiology* **30**, 746-755 (2019).
4. B. Lehne, *et al.*, A coherent approach for analysis of the Illumina Methylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biology* **16**, 37 (2015).
5. S. Marco-Sola, M. Sammeth, R. Guigó, P. Ribeca, The GEM mapper: fast, accurate and versatile alignment by filtration. *Nat Methods* **9**, 1185-1188 (2012).
6. O. Delaneau, *et al.*, A complete tool set for molecular QTL discovery and analysis. *Nat Commun* **8**, 15452 (2017).
7. M.E. Ritchie, *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
8. U. Võsa, *et al.*, Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. bioRxiv:447367 (19 October 2018).

9. S. Ligthart, *et al.*, Genome Analyses of > 200,000 Individuals identify 58 loci for chronic inflammation and highlight pathways that link inflammation and complex disorders. *Am J Hum Genet* **103**, 691-706 (2018).
10. J. Bowden, *et al.*, Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I<sup>2</sup> statistic. *Int J Epidemiol* **45**, 1961-1974 (2016).
11. S. Burgess, A.S. Nutterworth, S.G. Thompson, Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* **37**, 658-665 (2013).
12. J.D. Storey, The positive false discovery rate: a Bayesian interpretation and the *q*-value. *Ann Statist* **31**, 2013-2035 (2003).
13. G. Hemani, *et al.*, The MR-Base platform supports systematic causal inference across the human phenome. *eLIFE* **7**, e34408 (2018).
14. D.P. MacKinnon, J.H. Dwyer, Estimating mediated effects in prevention studies. *Eval Rev* **17**, 144-158 (1993).
15. T.S. Lange, S. Vansteelandt, M. Bekaert, A simple approach for estimating natural direct and indirect effects. *Am J Epidemiol* **176**, 190-195 (2012).
16. J. Steen, T. Loeys, B. Moerkerke, S. Vansteelandt, Flexible mediation analysis with multiple mediators. *Am J Epidemiol* **186**, 184-193 (2017).
17. T.J. VanderWeele, S. Vansteelandt, Mediation analysis with multiple mediators. *Epidemiol Methods* **2**, 95-115 (2014).
18. S. Vansteelandt, M. Bekaert, T. Lange, Imputation strategies for the estimation of natural direct and indirect effects. *Epidemiol Methods* **1**, 131-158 (2012).
19. J. Steen, T. Loeys, B. Moerkerke, S. Vansteelandt, medflex: An R Package for Flexible Mediation Analysis using Natural Effect Models. *Journal of Statistical Software* **76**, 1-46 (2017).

## YFS

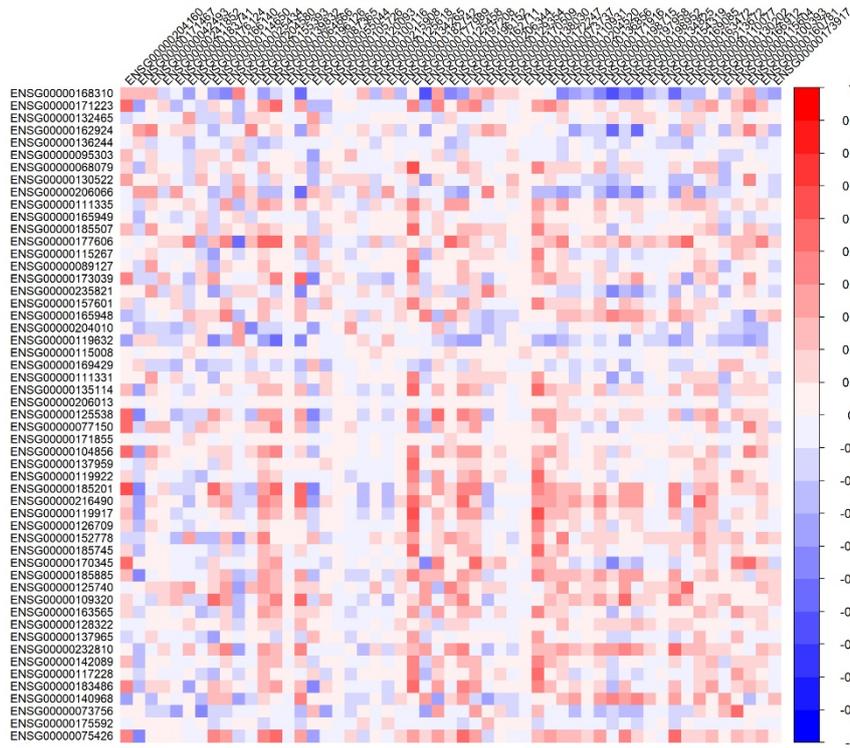


## SKIPOGH

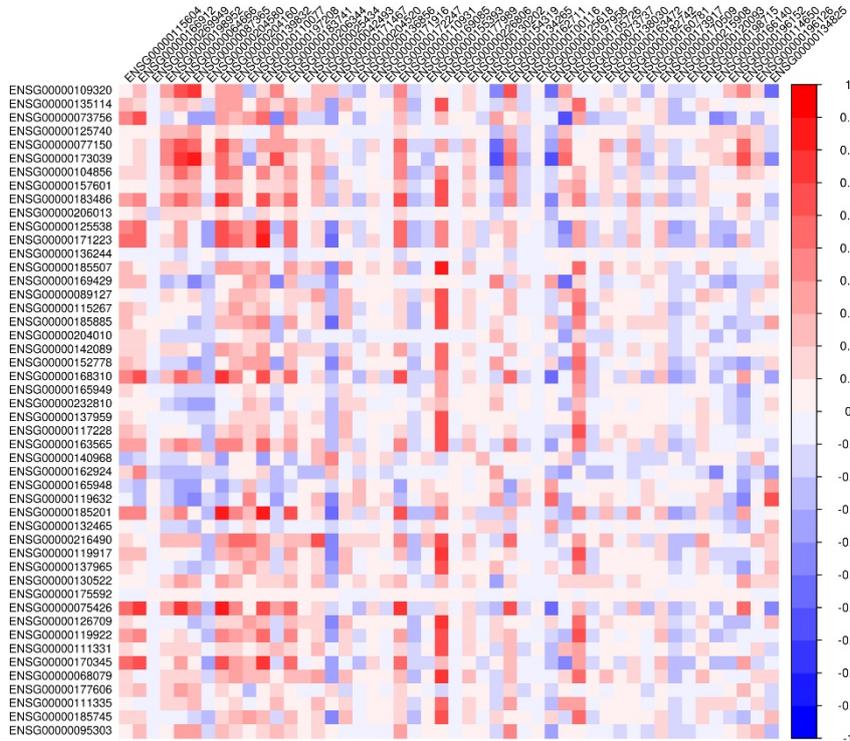


**Figure S1.** Heatmap of Pearson's correlation between principal components of transcription levels of 25MR genes (columns) and of CTRA genes (rows) in CRYFS (top) and SKIPOGH (bottom) participants. Warm colours are for positive correlation values, while cold colours are for negative values. The PCs summarize about 50% of the transcription levels variance in the 25MR (23 PCs in both CRYFS and SKIPOGH) and CTRA (12 PCs in CRYFS and 10 PCs in SKIPOGH) sets, respectively. The PCs are ordered according to decreasing amounts of explained variance.

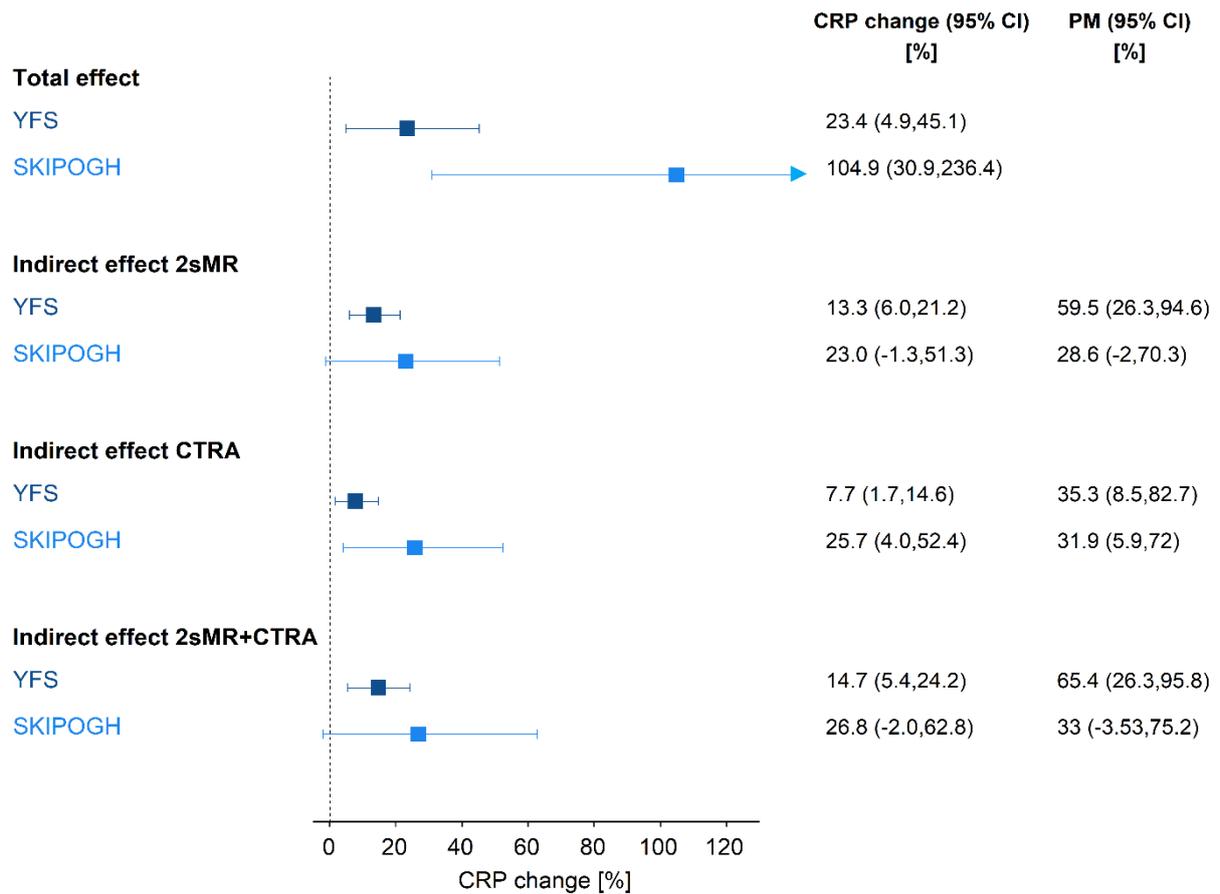
## YFS



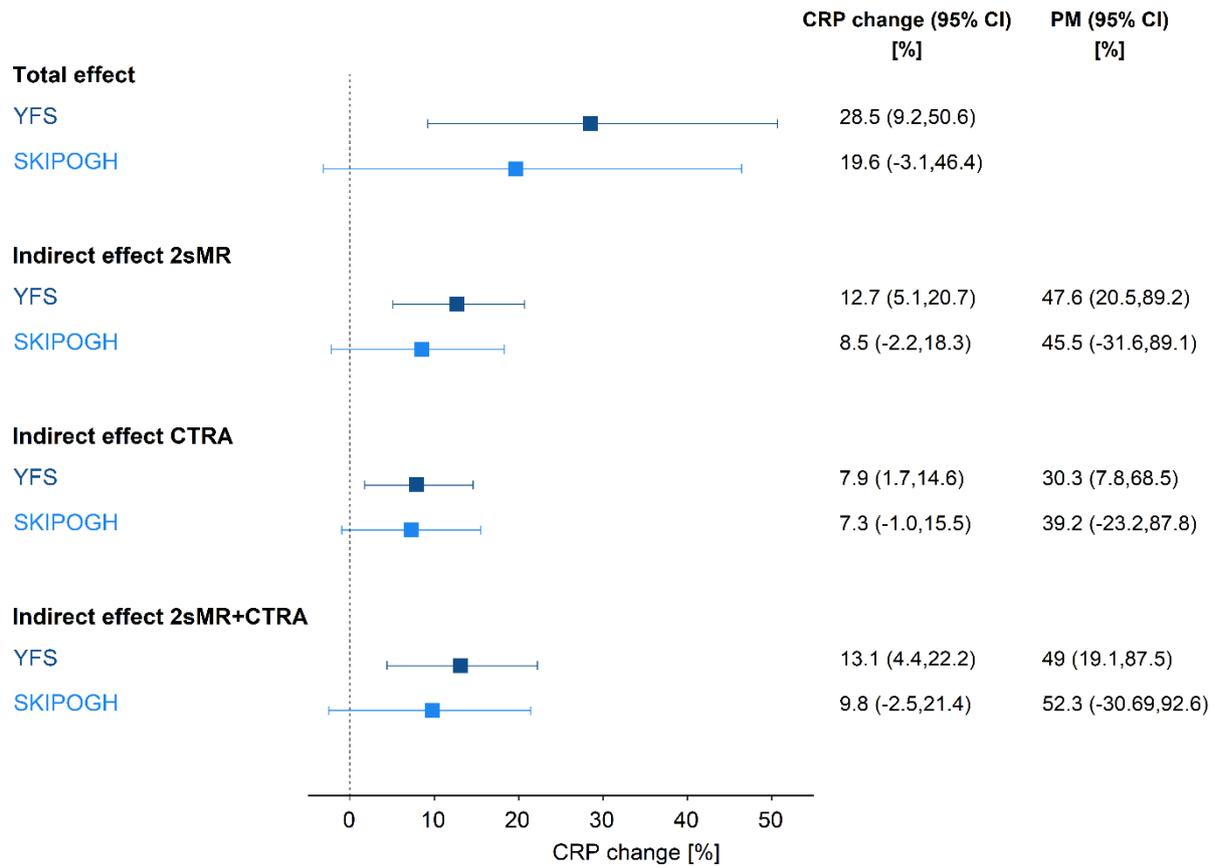
## SKIPOGH



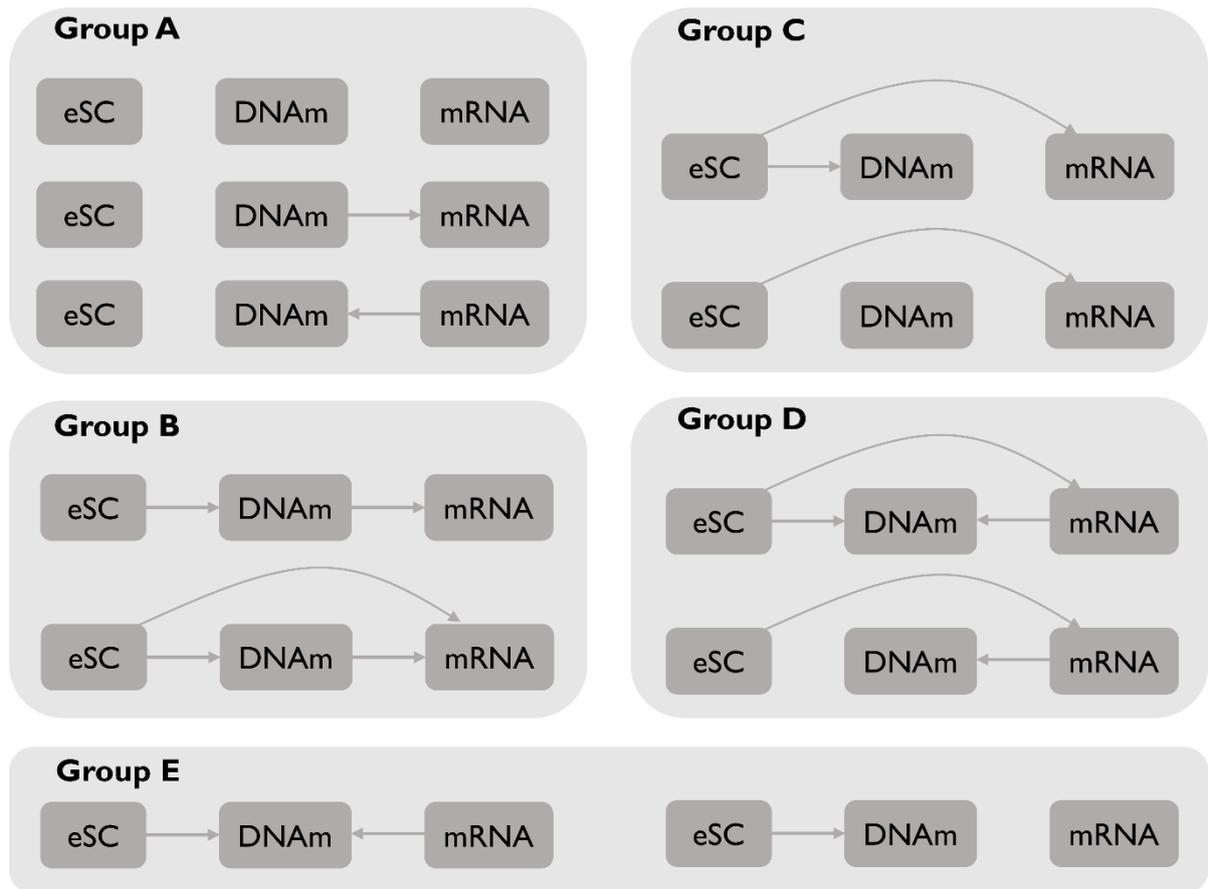
**Figure S2.** Heatmap of Pearson's correlation between transcription levels of 2sMR genes (columns) and of CTRA genes (rows) in YFS (top) and SKIPOGH (bottom) participants. Warm colours are for positive correlation values, while cold colours are for negative values. Genes are ordered according to their similarity in transcription levels based on hierarchical clustering run separately in both sets. Genes are reported according to their Ensembl identifier. In YFS, four 2sMR genes are not reported as they were not measured. In SKIPOGH, five CTRA genes are not reported as they had an average value of counts per million reads smaller than 0.25.



**Figure S3.** Size of effects and proportion mediated (PM) estimated by mediation analysis in each population when participants with CRP>10 mg/L were excluded (N=34 in YFS and N=11 in SKIPOGH). CRP change (%) and 95% confidence intervals (CI) are reported for the total effect of low (vs high) parental occupational position on CRP levels in adulthood, and for indirect effects through transcription levels of pro-inflammatory genes *en-bloc*. Joint indirect effects via 2sMR, CTRA or 2sMR+CTRA genes were estimated by summarizing transcription levels with principal components corresponding to about 50% of the transcription variance in each genes' set.



**Figure S4.** Size of effects and proportion mediated (PM) estimated by mediation analysis in each population when parental education was the exposure (N=1,433 in YFS and N=721 in SKIPOGH). CRP change (%) and 95% confidence intervals (CI) are reported for the total effect of low (vs high) parental education on CRP levels in adulthood, and for indirect effects through transcription levels of pro-inflammatory genes *en-bloc*. Joint indirect effects via 2sMR, CTRA or 2sMR+CTRA genes were estimated by summarizing transcription levels with principal components corresponding to about 50% of the transcription variance in each genes' set.



**Figure S5.** Competing groups of causal structures (DAGs) to investigate the relationships between early life socioeconomic conditions (eSC), methylation at DNA sites (DNAm) and gene transcription (mRNA) in leukocytes in adulthood. For the sake of simplicity potential confounders (age, sex, centre of blood collection) are not drawn.

**Table S1.** Annotation (HGNC symbol and Ensembl name, genomic location in GRCh37) of 2sMR and CTRA genes.

**Table S2.** Annotation (identifier, genomic position in GRCh37, and Ensembl identifier of nearby gene) and winning model (5 models, see Figure S5) of 4,076 CpGs selected for the Bayesian network scoring analysis.

**Table S3.** Joint indirect effects via 2sMR-implicated genes and CTRA genes (% of CRP change) when transcription levels are summarized via different numbers of principal components corresponding to about 30%, 40%, 50% or 60% of transcription variance.

Explained transcriptional variance	YFS	SKIPOGH
30%	12.0% (3.9%,20.8%)	10.4% (0.2%,19.9%)
40%	14.0% (5.1%,24.0%)	11.1% (-0.7%,23.1%)
50%	17.2% (7.1%,28.5%)	13.9% (0.0%,27.2%)
60%	17.3% (6.7%,29.1%)	16.0% (1.0%,32.2%)

**Table S4.** Size of effects (odds ratio OR) and proportion mediated (PM) estimated by mediation analysis in SKIPOGH when binarizing CRP values ( $\geq 3$  mg/L for heightened inflammation). OR and 95% confidence intervals (CI) are reported for the total effect of low (vs high) parental occupational position on heightened inflammation in adulthood, and for indirect effects through transcription levels of pro-inflammatory 2sMR+CTRA genes *en-bloc*. Transcription levels were summarized with principal components corresponding to about 50% of the transcription variance in the genes' set.

Effects	OR (95% CI)
Total effect	2.22 (1.33,3.74)
Indirect effect 2sMR+CTRA	1.47 (1.06,1.95); PM = 48.1% (8.9%,89.0%)

**Table S5.** Cardinality of the most likely group of model structures across the five tested competing groups (see Figure S5) for a total of 4,076 CpG sites when parental education was used as exposure in up to N=1,211 YFS and N=677 SKIPOGH participants. The most likely group of causal structures is selected when its posterior probability is at least roughly three times larger than any other group's posterior probability. None corresponds to CpGs for which none of the five tested groups reached winning evidence or models could not be estimated as the transcription level was not available in a few genes (37 and 33 CpGs in YFS and SKIPOGH, respectively).

Causal structures		N of CpGs	
		YFS	SKIPOGH
A	Parental education does not influence methylation nor transcription levels	3,599	3,527
B	Parental education drives changes in methylation and transcription levels; methylation also causes changes in transcription	0	0
C	Parental education drives gene transcription, but methylation and transcription are not related	47	4
D	Parental education drives transcription that in turn influences methylation levels	1	0
E	Parental education influences methylation but not transcription levels	4	28
None		425	517
Total		4,076	4,076