1. Online Materials and Methods

1.1 Discovery population.

Analyses were conducted on data collected in the UK Biobank project[31] under project 19655. All subjects gave written informed consent. UK Biobank has approval from the North West Multi-Centre Research Ethics Committee (MREC), In Scotland, UK Biobank has approval from the Community Health Index Advisory Group (CHIAG). We included only subjects who completed the food frequency questionnaire and were defined by UK Biobank as genomically British and unrelated.

1.2 Phenotype modelling

Quantitative food and drink intake phenotypes were all converted to weekly consumption, e.g. consuming 3 cups of tea a day was converted to 21 cups/week. Semi-quantitative descriptors: never, 2-4 times a week, 5-6 times a week, once or more daily were converted to 0, 3, 5.5 and 7 respectively. All “Prefer not to answer” and “Do not know” answers were excluded from the analysis.

All coffee traits were stratified by type (instant, ground, decaffeinated) to account for differences in consumption patterns such as cup size and caffeine concentration. Participants who did not specify the type of coffee they usually consume were excluded from the analysis.

Coffee consumption (any type of coffee, including unspecified type of coffee) was treated as a covariate for water consumption due to a very high negative phenotypic correlation between water and coffee consumption. Some semi-quantitative traits do not directly refer to the amount of food or drink consumed but to its type. Fat content in milk was calculated as the fat percentage and non-dairy types of milk (e.g. soy) were removed. Drink temperature (very hot, hot and warm) were converted to an arbitrary 3-unit scale (3, 2, 1). People who did not report consuming hot drinks were excluded from the analysis. Summary statistics for the traits are reported in Supplementary data 10.

All traits including the binary ones were treated as being quantitative. For binary traits this is the same as running a trend test. We did this because standard software cannot perform logistic regression on such a large sample in a timely fashion. Residuals for each trait using a linear model were first estimated in R using age and sex and then used as phenotype for the association analysis using BOLT-LMM. Coffee and tea consumption were added as covariates for the analysis of water consumption as described before.

In order to verify the best function for trait normalisation in the association analysis and to ensure that most quantitative traits demonstrated a right-tailed distribution, we applied both log10(x+1) and sqrt(x) transformations to the traits and then regressed them against the covariates. The best transformation was chosen by visually inspecting the Q-Q plots of the residuals from these regressions and checking which one better approximated a normal distribution: in all cases the residuals were properly normalised. Supplementary data 3 gives full details of the phenotype modelling.

1.3 Genome wide association study (GWAS).

Association analyses were conducted on SNPs imputed to the HRC panel[1], as provided by the UK Biobank, using BOLT-LMM[2]. Population stratification was assessed using LD-score regression as implemented in ldsc[3] both for both GWAS and after meta-analysis using the LD scores provided with the software: no evidence of residual stratification was observed. Table S15 reports the LD regression intercept and h^2 estimation using ldsc. Given that we identified 5 main clusters of traits we set the genome-wide significance threshold at $1x10^{-8}$.

1.4 Replication Analysis

Replication analyses were conducted independently by using genetic and dietary data from the EPIC-Norfolk Study[4] and the Fenland Study[5]. Both are on-gonig population based cohort studies conducted in the East of England. At baseline of the EPIC-Norfolk Study (1993-1997) and the Fenland study (2005-2015), the same food-frequency questionnaires (FFQs)[4,6] were administered to participants. Each participant was requested to report frequency of consumption of 131 food

items by selecting one of nine categories of frequency of food consumption ('never or less than once/month', '1-3 per month', 'once a week', '2-4 per week', '5-6 per week', 'once a day', '2-3 per day', '4-5 per day', and '6+ per day'). As performed in the analysis of the UK BioBank, the quantitative information on frequency of consumption was assigned to each response for 131 food items. We summed up frequencies of consumption of multiple food items in each food group (e.g. margarine and butter for bread spread; different types of vegetables for total vegetable consumption), so that resulting variables were comparable with those used in the UK BioBank.

Participants were genotypes with different genotyping arrays. In the EPIC-Norfolk Study, Affymetrics Axiom UKBiobank was used (n=21,044). In the Fenland Study, three arrays were used: Affymetrics Axiom UKBiobank (n=8,994), Illumina Metabochip (n=3,217), and Illumina ExomeChip v1.0 Human Exome-12v1-B (n=1,650). Further details are available elsewhere[5]. Missing information of SNPs was imputed to the HRC and UK10K panels by population and genotyping array. After we excluded participants without either dietary or genetic information, 32,779 participants in total became available for the replication analysis of EPIC-Norfolk Study (n=21,337; $n_{original}$=25,639) and the Fenland Study (n=11,442; $n_{original}$=12,731). Association analyses were conducted with BOLT-LMM in each of the four population-array strata.

1.5 Univariate MR

Univariate MR to measure the causal effect of health-related traits on food was conducted using the TwoSampleMR[7] R package. We decided to focus on traits for which dietary medical advice is given due to medical conditions in particular: body mass index (BMI), low density lipoprotein (LDL) Cholesterol, high density lipoprotein (HDL) Cholesterol, Total Cholesterol, Triglycerides, Diastolic and Systolic blood pressure, Educational attainment, Type II diabetes and Coronary Heart Disease.The full list of studies from which the summary statistics were derived is given in Table S6.

For each trait we selected all SNPs with $p<5 \times 10^{-8}$ and $r^2<0.001$. We then performed stepwise heterogeneity pruning, first estimating heterogeneity using the Q statistic, if $p<0.05$, we removed

one SNP at a time, looking at which removal would improve the statistic more. This procedure was repeated until $p>0.05$. We then tested if the intercept from the MR-Egger regression was different from zero ($p<0.05$). If this was the case, MR-Egger was used for the MR analysis otherwise the Inverse Variance method was used. We considered as significant those relationships in which the Benjamini and Hochberg FDR<0.05.

1.6 Estimation of prior expected effect through bGWAS and genome-wide mediation analysis.

One of the main issues in GWAS studies is to decide which covariates to apply in the regression model. When deciding to include a covariate or not, depending on the causal relationships between the traits, we may risk creating different types of biases. One approach could be to include in the model just the non-heritable covariates (i.e. sex and age) which will avoid spurious results due to collider bias. The problem with including heritable covariates is that the GWAS will also detect those SNPs which are associated with the covariates which are causally related to the trait, making the interpretation of the results harder. For example if educational attainment causally influences BMI, if the study is powered enough, it is possible that the genes from educational attainment show up on the BMI GWAS. Moreover, this limits the generalisability of at least part of the results to other populations in which the phenotypic architecture of the trait may be different. For example, following the previous example educational attainment was causal to the trait in the European population and not in East Asia the SNPs causal from Educational attainment will not be replicated in East Asian populations affecting also the generalisability of results.

It is thus extremely important to identify a technique which allows to distinguish between those SNPs which are directly causal of the trait of interest from those which are associated through other mediators.

There are 3 possible scenarios:

1) The covariate is causing the trait of interest in which case it would be correct to include it in the model (i.e. diet and socioeconomic status).

2) The trait of interest is causing the covariate in which case including the covariate in the model could result in collider biases.
3) The trait and the covariate are causing each other. In this case, using the covariate in a regression framework will correct for the overall effect while in truth we are interested in correcting only for the effect of the covariate on the trait and not vice versa.

So, in order to properly correct our analyses, we need to determine if a covariate is causing the trait of interest and at the same time estimate the size of the effect. In this respect, we can use two-sample MR to establish both causality and effect size, using a multivariable MR approach if there are multiple covariates. In principle, once this is done we can correct the phenotype given the covariates and run the GWAS based on this new corrected trait. However, given that in many cases information on covariates may be missing, a method which exploits available GWAS summary statistics would be more desirable. Such a method would need to first estimate which covariates are causal to the trait of interest and then based on their causal estimates, perform mediation analysis for each SNP. In the second step, for each SNP an expected mediated effect is estimated, combining the effect of the SNP on each of the causal traits with their multivariable causal effects. The expected effect can be subtracted from the observed effect on the trait of interest to get the “direct causal effect” of the SNP on the trait.

For the estimation of the prior expected effect, we used a Bayesian GWAS (bGWAS)[8]. The bGWAS approach leverages information from studies of related traits to carefully build informative priors for each SNP. To analyse food choices, we decided to include information from the same traits used for the univariate MR plus Crohn’s disease and ulcerative colitis. Given that the traits were meant to be used in a multivariable model, total cholesterol was removed to avoid strong collinearity with LDL and/or HDL cholesterol. MR was used to derive multivariate causal effects of the set of related traits on the different food choice phenotypes, using independent instruments (association p-value below 5×10^{-8} for at least one related trait, LD pruned $r^2<0.001$). For each food choice phenotype, a stepwise selection approach was used to select only the related traits

significantly affecting the focal phenotype. To calculate the prior for SNPs on a given chromosome, we first apply multivariate MR (masking the focal chromosome) using the significantly related traits identified by the stepwise selection to estimate causal effects. We next use the causal effect estimates in combination with GWAS summary statistics of the related traits to estimate the prior effects. The prior effect of a SNP *i* ($\hat{\mu}_i$) is calculated using the observed standardised effects (Z-scores) for the *T* different related traits ($Z_{i,t}$) and the causal effects estimated masking one chromosome ($\widehat{\beta_t}$):

$$\widehat{\mu_i} = \sum_{t=1}^{T} Z_{i,t}\,\hat{\beta}_t \quad (1)$$

The prior estimated by bGWAS is on the scale of the z-score of the GWAS from the trait of interest, so the non-mediated z-score can be easily then estimated as the difference between the original z-score and the prior. The prior can be thought of as the total indirect effect while the corrected effect as the pure direct effect from mediation analysis[9]. Keeping the standard error constant, it is then easy to derive the corrected beta as z-corrected*se. It is important to note that when we estimate the prior expected effect we do not take into account the error of the multivariable estimates and we use the point estimates directly. This is because in MR the standard errors linked to each beta estimate are relatively large and if taken into account would lead to a very large final standard error and thus all pure direct effect estimates would have extremely large errors making them not usable. Although this is of course an approximation this is not unlike the estimation of polygenic risk scores where the SNP point estimates are used as weights for the score. It is important to note that for the further causal inference we used uncorrected betas and thus this will not influence the effect estimation.

This approach has several advantages compared to correcting the phenotype directly. First, it allows the GWAS to be corrected for covariates which have not been measured directly on the same samples. This is a great advantage in terms of the models that can be explored, for example in our case we have corrected the GWAS for LDL although LDL had not been measured in UK

biobank at the time of the analyses, and this is also useful for phenotypes such as Crohn's disease, for which relatively few cases are present in UKB.

Moreover, it is possible to compare the effects before and after correction, giving us information on the likelihood the observed effect is directly associated with the trait of interest or is mediated. it is important to remember that conditioning on a phenotypic covariate will not necessarily completely correct the mediated effect, due to noise, and thus the comparison of the two effects is more informative. Finally, we should be able to trace back the path of mediation looking at the different components of the prior for each SNP thus helping greatly in interpreting the results and planning subsequent studies.

All the exposure traits GWAS have been first imputed using SSimp[10] (https://github.com/zkutalik/ssimp_software) and the UK10K genotypes as the reference panel. Finally, all A/T or G/C SNPs were removed to avoid errors in the harmonisation of effects coming from different GWAS, due to strand errors. The proportion of genetic variance of the food traits explained by the health-related traits was measured by taking the squared genetic correlation between the expected Z-score and the observed one.

1.7 Genetic correlations with other traits and stratified LD-score regression.

Genetic correlations of the self-reported food consumption traits against 844 traits have been estimated using LD-score regression as implemented in LD Hub [3,11]. Genetic correlations were estimated both with the corrected and uncorrected traits using the bivariate LD-score regression model. Stratified LD-score regression[12] was run using ldsc and the annotation files available on the ldsc website.

1.8 Identification of the SNPs directly associated with the food traits.

One of the main objectives of GWAS is to identify genes directly responsible for the trait of interest, in our specific case, however, we have shown that looking just at the genome-wide hits is not sufficient and does not exclude SNPs truly associated with other causal heritable traits, due to

vertical pleiotropy. In order to distinguish between these two types of SNPs, we decided to look at the ratio between the corrected trait and the original trait or corrected-to-raw ratio (CRR).

To understand this choice let's suppose we have a trait of interest (Y) and a second heritable trait (X) which is causal with effect, $\beta_{X\to Y}$. We will call SNPy, the SNPs directly causing Y with effect, $\beta_{SNPy\to Y}$ and SNPx, the SNPs which are directly causing X with effect, $\beta_{SNPx\to X}$. If the whole effect of SNPx on Y is mediated through X its effect will be given by

$$\beta_{(SNPx\to Y)expected} = \beta_{SNPx\to X}\times\beta_{X\to Y} \quad (2)$$

$\beta_{X\to Y}$can be estimated through MR, while $\beta_{SNPx\to X}$ can be retrieved from the GWAS of X. Assume we measure the effect of a SNP for which it is unknown if the effect is mediated through X or not (as is the case in real data). We define $\beta_{SNP\to Y}$, the observed effect of the SNP on Y

if $\beta_{SNP\to X}$ is truly 0, then we can estimate the expected mediated effect of the SNP through X as

$$\beta_{(SNP\to Y)expected} = \beta_{SNPx\to X}\times\beta_{X\to Y} \approx 0 \quad (3)$$

and

$$\beta_{(SNP\to Y)corrected} = \beta_{SNP\to Y} - \beta_{(SNP\to Y)expected} \approx \beta_{SNP\to Y} \quad (4)$$

thus

$$CRR = \frac{\beta_{(SNP\to Y)corrected}}{\beta_{SNP\to Y}} \approx 1 \quad (5)$$

On the other hand if

$$\beta_{SNP\to X} \neq 0 \quad (6)$$

then

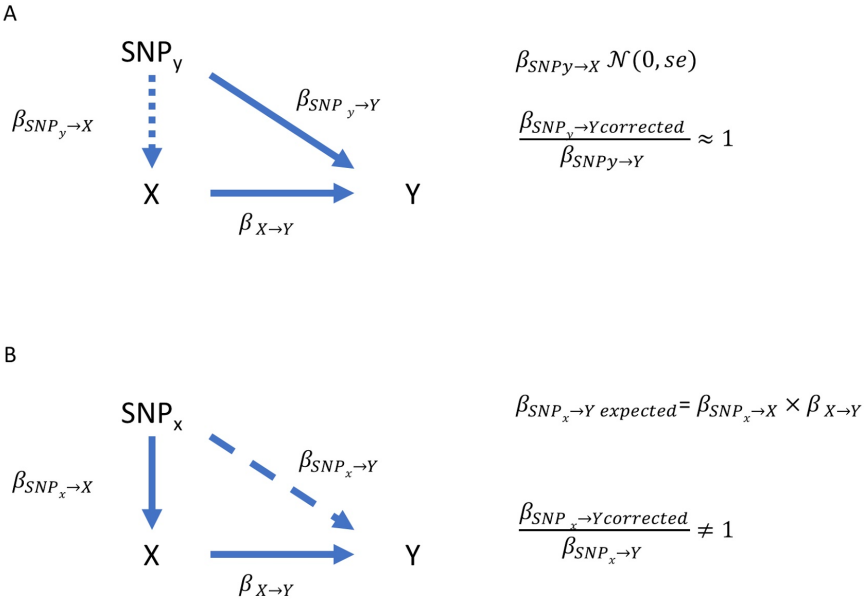
$$\beta_{SNP\to Y\ expected} = \beta_{SNP\to X}\times\beta_{X\to Y} \neq 0 \quad (7)$$

$$\beta_{(SNP\to Y)corrected} = \beta_{SNP\to Y} - \beta_{(SNP\to Y)expected} \neq \beta_{SNP\to Y} \quad (8)$$

then

$$CRR = \frac{\beta_{(SNP\to Y)corrected}}{\beta_{SNP\to Y}} \neq 1 \quad (9)$$

Figure S1. Directed acyclic graph explaining the two possible scenarios for the effect of a SNP on the trait of interest Y. (a) *The SNP has a direct effect on Y not mediated through X. Then the estimated effect of SNP on X will be normally distributed around 0, thus the corrected and uncorrected effects will be similar and their CRR will be close to 1.* ***(b)*** *The SNP effect is mediated through X, thus the corrected effect will deviate from the observed one and CRR will deviate from 1.*



In real-life situations, the betas are estimated and thus will carry an error due to chance and it is thus important to understand what values the CRR may assume under different scenarios. We can, however, summarise three types of relationship between the trait of interest and its causal factors:

1) X→Y

2) X→Y and Y→X

3) U→Y and U→X

Where U is a heritable confounder responsible for the relationship between X and Y. We thus use simulations to understand how the relationships between the traits influence the CRR and our ability to use it to distinguish between SNPs, the effects of which are mediated through X or U, and SNPs which are causal to the trait of interest without mediation mediation. The details and results of the simulations are reported in Supplementary Data 2.1.

1.9 Clustering of food and drink consumption traits.

Genetic correlations between the food traits were estimated using LD-score regression as implemented in the ldsc software[35] separately for the corrected and uncorrected GWAS. Hierarchical clustering of the two genetic correlation matrices was performed using two different algorithms: "complete" as implemented in the hclust() function from R and the ICLUST algorithm[17] from the R package psych. ICLUST assigns items to the same cluster based on the loadings of an underlying common factor. Items are then iteratively added to the clusters only if they increase the internal consistency of the cluster. The algorithm also allows for addition of traits in case of strong negative correlation. This has a compelling advantage for food consumption, as for example the intake of fatty foods has a strong negative correlation with eating healthy food, thus both can contribute to the same grouping. Differences in clusterings were compared graphically using a tanglegram in both cases. Given that ICLUST seemed to give more stable results compared to the "complete" clustering algorithm, the clusters produced with this algorithm were used for further analyses.

1.10 Multi-trait genome-wide association analysis

For each of the three main clusters of phenotypes (Meat/Fat, Healthy foods and Psychoactive Drinks), we performed multi-trait genome scans using a MANOVA-based multivariate analysis method implemented in the MultiABEL package[37] (https://cran.r-project.org/package=MultiABEL). The method can take genome-wide summary association statistics to infer phenotypic correlation coefficients and conducts a MANOVA test for each variant across the genome. This overcomes the issue of non-overlapping samples (e.g. it would be impossible to directly combine people drinking different type of coffee). The phenotypic correlation coefficient of any two traits can be estimated in an unbiased manner via the correlation of the genome-wide z-scores, and for binary outcomes, this is proportional to the phenotypic correlation on the liability scale. The MultiABEL package also calculates the best linear combination of multiple phenotypes that is associated with each variant.

1.11 Locus definition and prioritisation of genes.

To define a locus, we first selected all SNPs with p-value $<1x10^{-5}$ and then estimated the distance between each consecutive SNP located on the same chromosome. Two consecutive SNPs were identified as belonging to different loci if they were more than 250 kb apart. A locus was then considered significant if it contained at least one SNP with p-value below the previously described significance threshold: we thus identified 582 significant locus-phenotype associations. Given the high pleiotropy between different traits, we merged overlapping loci, which resulted in 302 independent loci.

In order to define for each locus which gene is more likely to be responsible for the observed association, we proceeded with custom prioritisation according to the following criteria. We first ran Haploreg v4.1[38] using r^2=0.8 as threshold (Supplementary Data 11). We also ran SMR[39] on each locus in order to identify eQTLs compatible with the observed association pattern. We used the tissue-specific significant eQTL from the Gtex Project website[26] (Supplementary data 12). We then proceeded to prioritise the genes according to the following criteria; if the locus met one of them the following ones were not tested:

1) The sentinel SNP is itself or is in strong LD (r^2>0.8) with a non-synonymous SNP
2) There is evidence of an eQTL (as tested with SMR) and the sentinel SNP and the eQTL are in strong LD (min r^2=0.5). Given the high number of significant eQTLs detected by SMR we used a dynamic selection starting from r^2≥0.99 and decreasing by 0.05 at each step until an eQTL was found or until r^2≤0.5.
3) The sentinel SNP is itself or is in strong LD (r^2>0.8) with a coding SNP (synonymous or in the untranslated region of the gene)
4) The top SNP is intronic or is in complete LD with an intronic SNP in the gene.
5) The top SNP is in strong LD (r^2>0.8) with an intronic SNP in the gene.
6) The closest gene.

The category and prioritised gene for each locus is annotated in supplementary table 3.

1.12 Prioritised gene annotation and network construction

Tissue enrichment using MAGMA[13] was run using FUMA[14]. For each available SNP, we chose the lowest p-value from all corrected analyses. For the functional annotation of the prioritised genes, we focused only on those coming from the "direct effect only" loci. First, we used the gene2Function tool from FUMA to identify enrichment in the same tissue used for the analysis with MAGMA. For both analyses, only tissues with Bonferroni corrected $p < 0.05$ were considered significant (Table S11-S12 report full results).

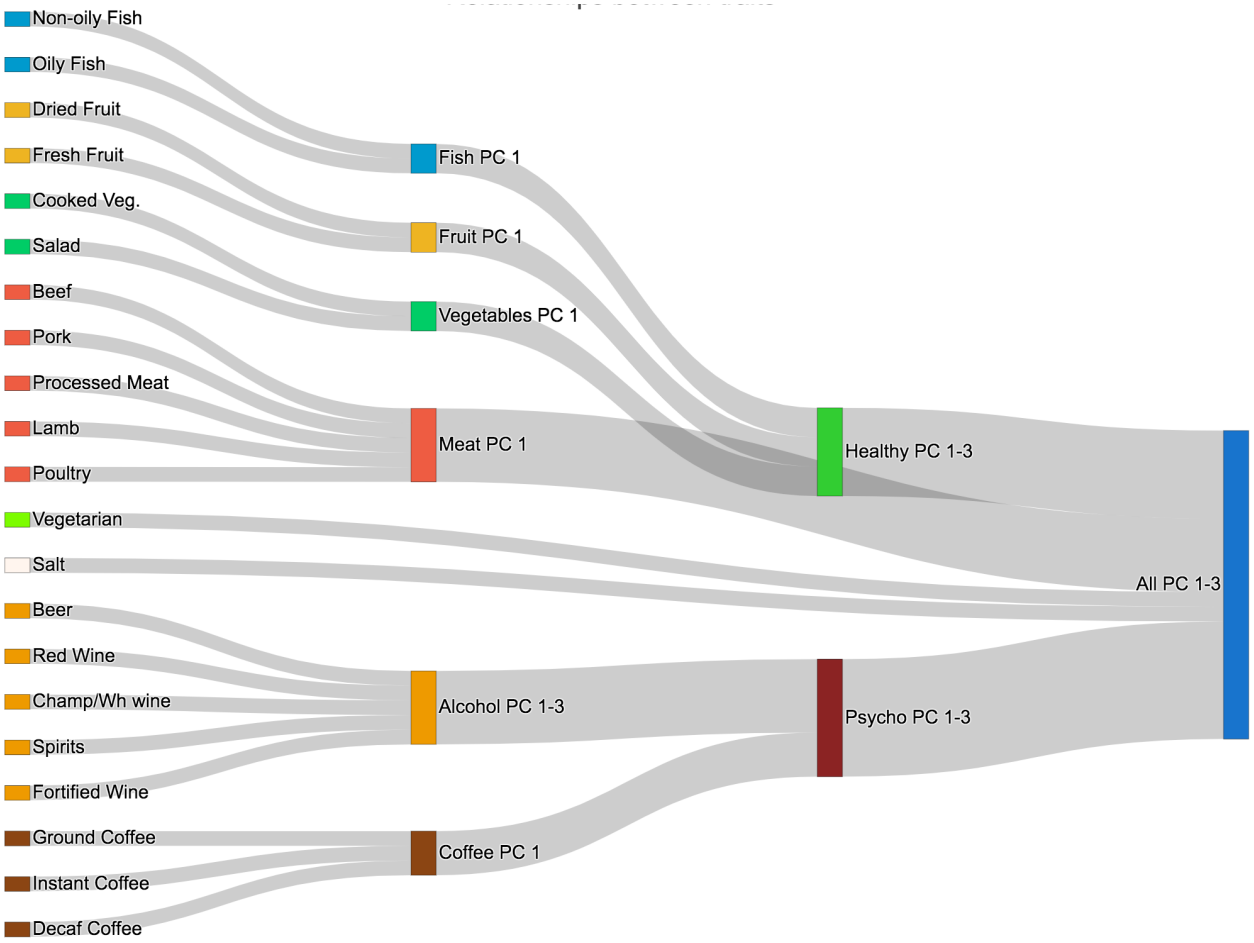
We then constructed an interaction network using STRING[15] (Table S13). After removing the genes which were not connected with any of the others, we ran community detection using Leuvain's method (Table S14 for membership). Tissue enrichment analysis was then performed for each community as done for the full gene set, focusing only on the overexpressed tissue analysis. Given the much higher number of tests performed, we used Storey's q-values to define significant tissues. Gene ontology enrichment was performed for each community using the compareCluster() function from the clusterProfiler R package[16]. We considered significant those terms which had a FDR<0.05 using the Benjamini and Hochberg method.

1.13 Mendelian randomisation to assess causal relationships between food and health

MR was conducted using the food traits as exposures and 79 traits available in MR-base[7] as outcomes. The genetic instruments were chosen from those with $p<5x10^{-8}$ and pruning for LD ($r^2<0.01$). We selected two sets of instruments: the first using raw p-values for thresholding (Raw) and the second using bias-corrected p-values and the filter with values comprised between 0.95 and 1.05 (CRR). Both sets of instruments were run using both raw effect estimates and corrected ones for comparison purposes (uncorrected/corrected). We considered the main analysis the one with the CRR-uncorrected analysis and multiple test correction was applied to these results. All other analyses were used for comparison purposes to show the difference in applying the CRR filtering. In order to maximise the number of SNPs available for analysis, the instrumental variables (IVs) were selected from those available for the specific outcome, thus the IVs change from outcome to outcome even if the exposure trait is the same. As exposures, we used 26 specific food items: adding butter to bread, and percentage fat in milk were not analysed as none of the

significant SNPs passed the CRR filter; fortified wine was also excluded as no significant SNPs were detected in the GWAS. Principal components (PC) traits were also estimated using as rotation matrix the eigen decomposition of the genetic correlation matrix. Betas from the GWAS for the interested traits were then projected on the rotated PC space. IVs for the PC analyses were selected by merging first all SNPs which could be selected as IVs in each trait of the group and assigning the lowest p-value in case of overlap. Then the SNPs were pruned for LD and the betas projected onto the PC space. Only PCs which could be interpreted were retained (not more than 3). Groups were created based on the groups created with ICLUST but also grouping the foods which had a clear common origin (e.g. oily and non-oily fish). A Sankey diagram of the relationships amongst the different traits is reported in Figure S2.

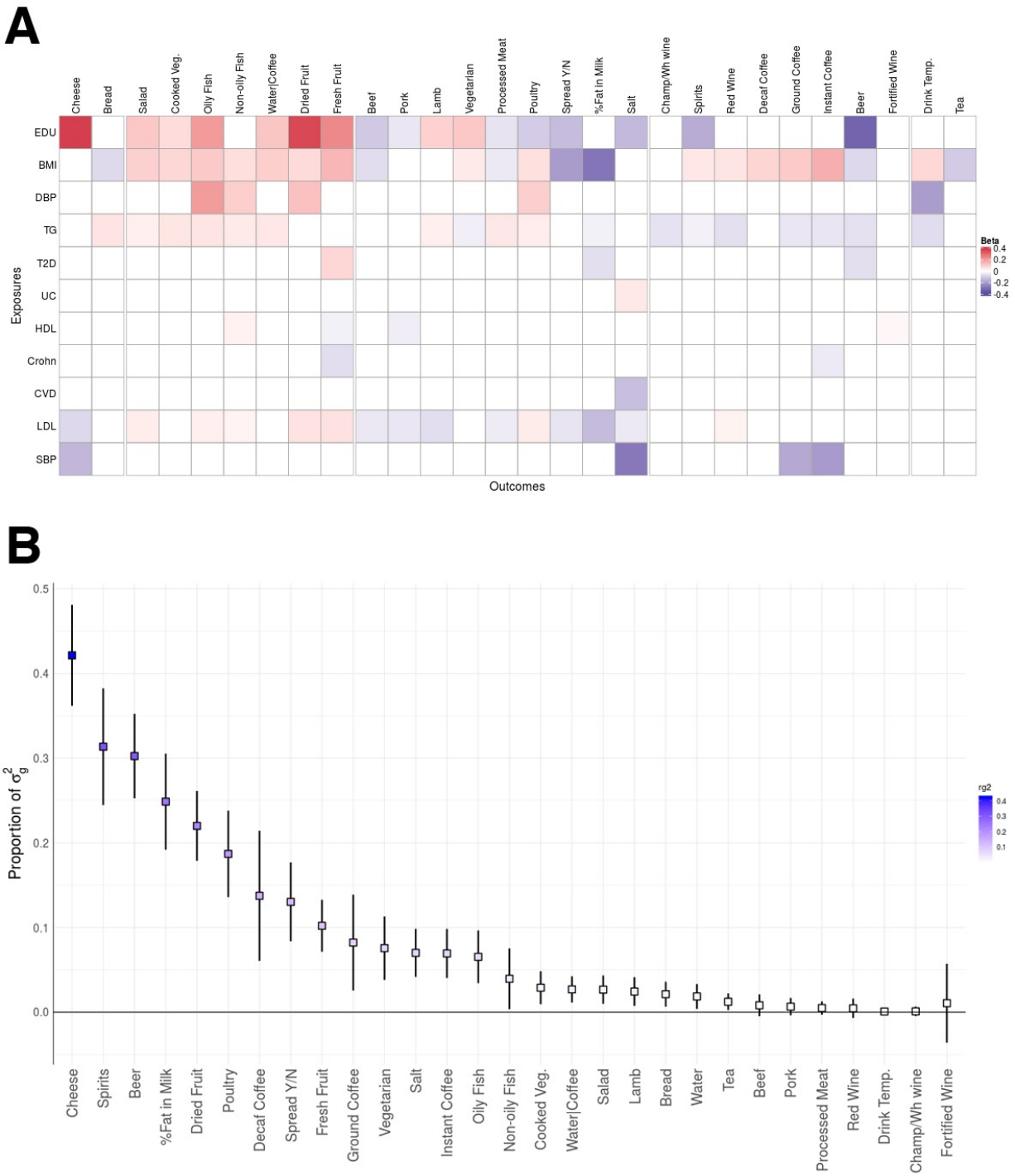
Figure S2. Sankey diagram describing the relationships between the single food traits and the composite ones.



All analyses were run using the same pipeline. After selecting the SNPs to be used as IVs, exposure and outcome data were harmonised. Only non-palindromic SNPs were used to avoid any misalignment between the two datasets. The IVs were tested for heterogeneity and outliers were removed using the MR-Radial method[17]. Inverse variance-weighted (IVW) MR was then used as the main analysis method, using random effects IVW if the residual heterogeneity had a p-value less than 0.05/79. We then tested for the presence of directional pleiotropy using the intercept from the MR-Egger regression. Finally, MR median and MR-Raps were used as sensitivity analyses. All results have been made available through an online app (https://npirastu.shinyapps.io/Food_MR/) which allows interrogation of the results both visually and in tabular form.

2. Extended Figures/Results

Fig S3. Results for the Mutlivariable MR. *Panel A The heatmap represents the effect of the health related traits on each food trait using from the multivariable model. The color is proportional to the effect size. Panel B. The plot represents the proportion of genetic variance which is explained by the effect of the health related traits on the food traits. Clearly some of the food traits are extremely biased having up to 40% of genetic variance due to the mediation of the health related traits.*



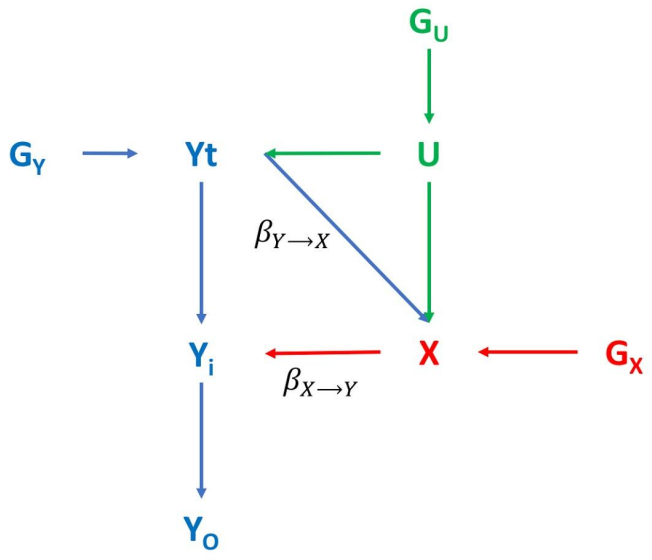
2.1 Simulation results for optimal parameter tuning for selecting SNPs directly associated with the trait of interest.

The objective of the simulations was to understand if the corrected/raw beta ratio limits chosen based on the genes for which biology is well known are correct. The simulations are particularly complex to set up since the DAG of the studied relationships involves bidirectional causal effects. Another important limitation is the fact that the exposure trait is not directly and correctly observed but it is the result of a FFQ in which the noise is extremely high with a test-retest correlation that can be as low as 0.5 (r^2=0.25).

The simulations include 4 different normally distributed traits:

Y_t which is the true trait of interest (food consumption in our case) without the effects of the outcome. X and U represent each the sum of all traits causal to Y. The difference between X and U is that U traits are also causal to X (so they act as a confounders) while X traits may also be subject to reverse causality by Y_t. Y_o is the observed trait which thus includes all causal effects and the noise due to the use of the questionnaire.

Figure S4. Diagram describing the relationships between the simulated traits and their relative parameters. Gy refers to the genetic variants which directly affect Y before any influence of confounding or other mediated traits (Yt). Gu represents the genetic component of a confounder trait U which causally affects both Y and X. Gx represents the genetic component of the outcome trait X which is in turn causally affecting the trait Yi. Yo represents the actual observed trait to which we add noise to reflect the test-retest correlation in FFQ data.



For simplicity each of the first 3 traits is determined by 10 SNPs each of which has a frequency of 0.3 and explains 1% of variance. Overall 30 SNPs have been used for each simulation and they are denoted as SNP_Y, SNP_X and SNP_U depending if their direct effect is through Y, X or U. The relationships between the different traits are summarised in Figure S4.

Where:

$\beta_{U->Y}$ represents the effect of the confounder on Y_t,

$\beta_{U->X}$ represents the effect of the confounder on X,

$\beta_{Yt->X}$ represents the effect of Y_t on X,

$\beta_{X->Yt}$ represents the effect of X on Y_t.

$\beta_{Yo->X}$ represents the causal effect we would be able to measure through MR. This measure is not of interest for the scope of our simulations.

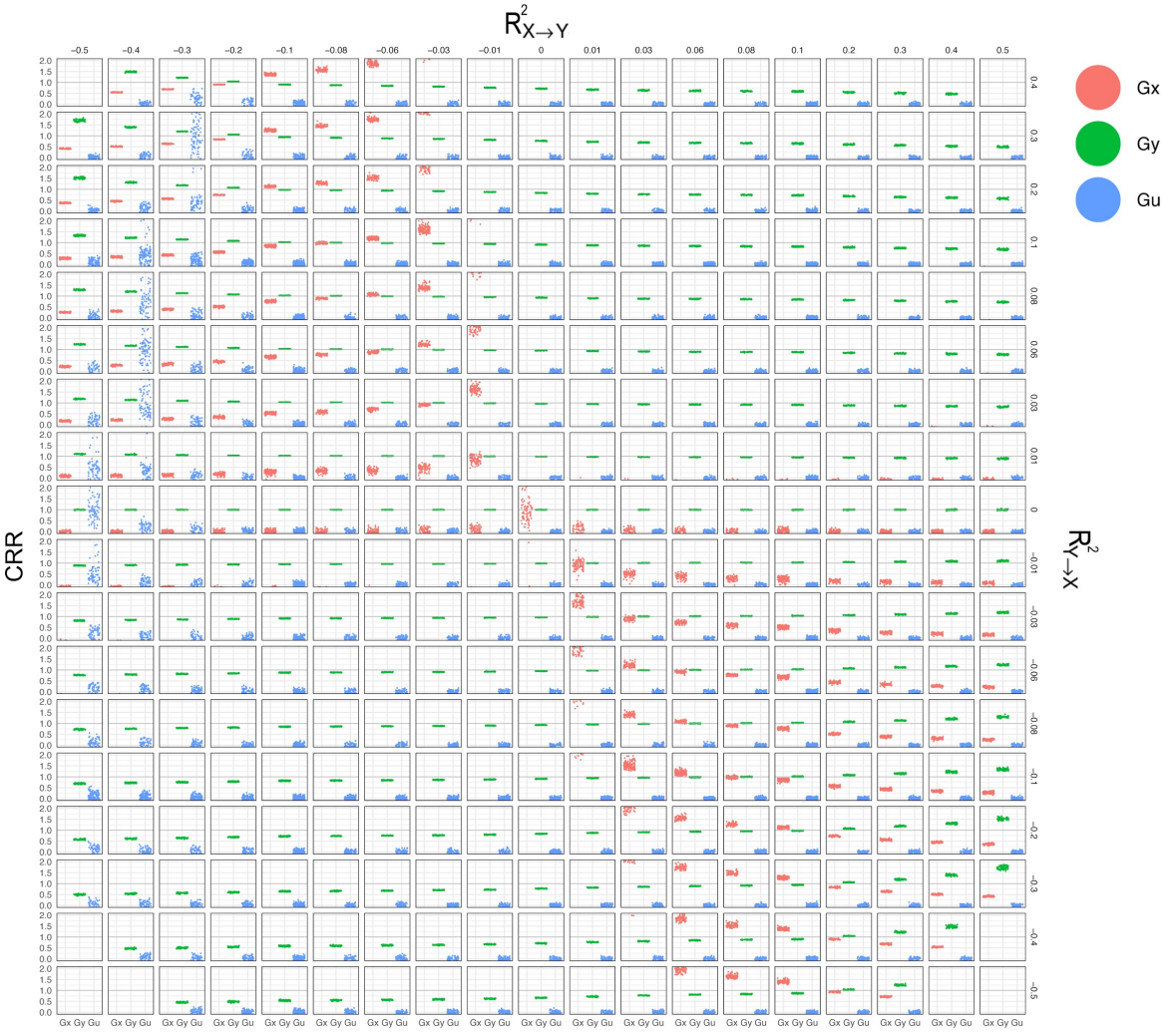
For simplicity the $\beta_{U->Y \text{ and }} \beta_{U->X}$ were both set so that the confounder explained 20% of the variance of Y and X. Simulations were then conducted for a large array of values of $\beta_{X->Yt}$ and $\beta_{Yt->X}$, which ranged from 50% of variance explained to 0%, with both positive and negative effects. Values of $\beta_{Yt->X}$=0 simulates the case where no reverse causality exists. A denser grid was used between r^2=0-0.1 to examine more closely the results at smaller effects which likely better resembles most real cases. Each set of parameters was run 10 times so that 100 SNPs were simulated for each category and set of parameters. To replicate the conditions of the paper we simulated two different independent populations so that we could apply the MV MR correction procedure to study the effects in a setting which resembles the real life scenario. Both populations were simulated so that N=400,000.

After simulating the two populations we proceeded to perform the association analysis for all 30 SNPs with all three observed traits, leaving out the original trait of interest Y_t, assuming we would not be able to directly observe it (as is the case for food consumption measured with FFQ). We then performed the MV MR of X and U on Y using as IV all SNPs which had $p<5x10^{-8}$ in either the GWAs from X or U assuming we had no way of distinguishing the source of the SNP. For this

analysis we used population 2 for the exposure betas and p values while population 1 for the effects of Y. Finally we used the betas combined with the population one Y_o association results to estimate the corrected/raw ratio (CRR).

The main objective was to verify if using the proper limits of the corrected/raw ratio allows the SNP_Ys to be correctly distinguished from the SNP_Xs and the SNP_Us. Figure S5 shows the scatterplot of the CRR for the 3 categories of SNPs for each combination of simulated parameters. R^2 refers to the amount of variance explained by the causal trait. The values are both positive and negative to reflect the direction of the correlation. The CRR range for the plot is limited to values between 0 and 2 because SNP_Ys never assumed values outside this range.

***Figure S5**. Scatterplot of the CRR values for Gx (in red), Gy (in green) and Gu (in blue) at the different values of the effect of Y on X and of X on Y.*

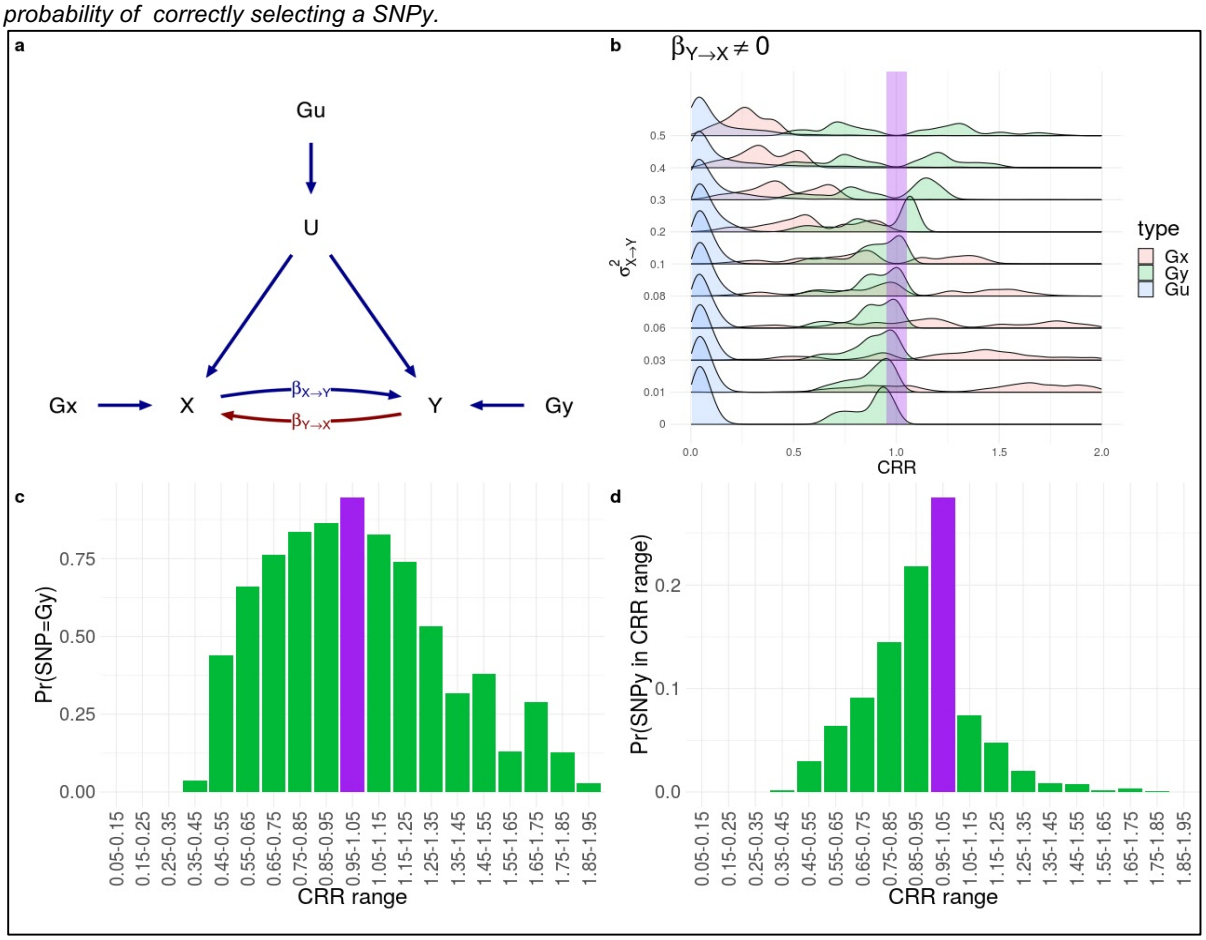


Clearly for most combination of parameters it is possible to easily separate the SNP_Ys from the SNP_Us. The task becomes slightly more complex in the case of the SNP_Ys and SNP_X in which some overlap is possible, especially when $\beta_{X->Y}=\beta_{Y->X}$ however this particular case is probably unlikely in real case scenarios as it would mean that the two effects cancel each other out.

From the previous figures, it is clear that if we were to know $\beta_{X->Y}$ and $\beta_{Y->X}$, we would be able to determine which values of CRR to use for the selection of the IVs in most cases. However without knowing these *a priori* it is however quite difficult as they would require prior knowledge of valid IVs for both Y and X, which is the objective of the method. Thus, the real question is if there is a range of values for the CRR which maximises the probability of not discarding the SNP_Ys while not including SNP_Xs or SNP_Us.

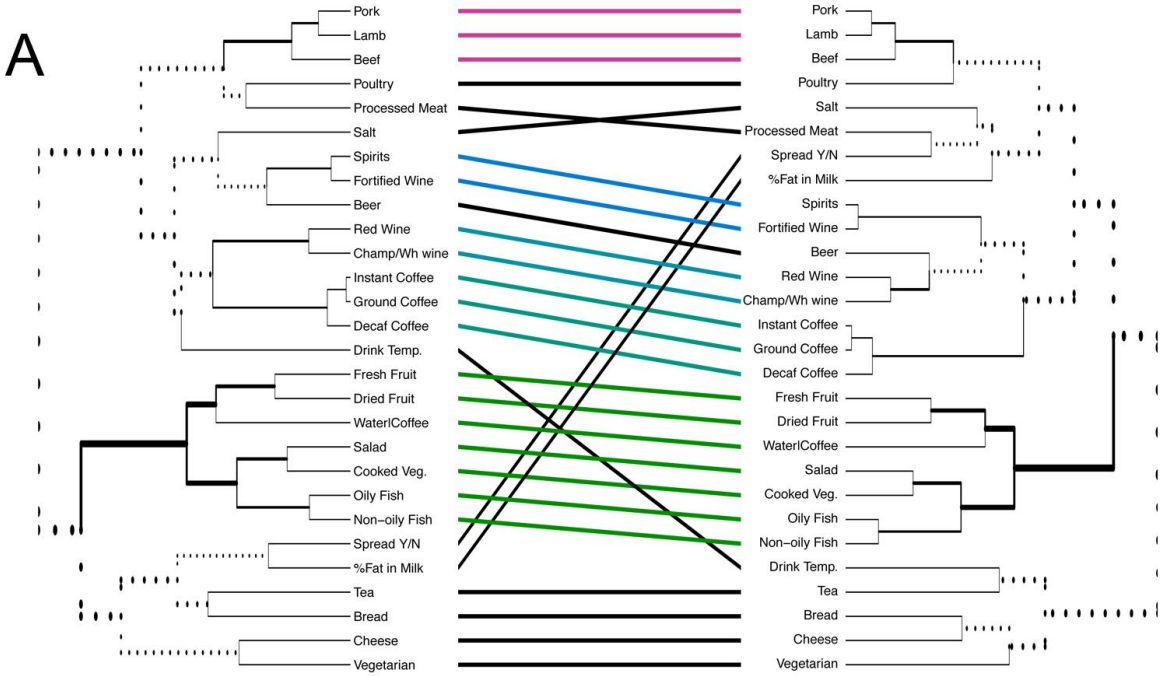
Therefore we verified how the probability of actually detecting a SNP_Y, P(detection), and the probability that an IV which met the criteria was actually a SNP_Y, P(SNP_Y), at different ratio limits. For the values selected for our study (1±0.05), the P(detection) is 0.27 overall, and 0.57 for the combination of the weak effects, while P(SNP_Y) is >80% in both cases (Figure S6). Given that, for almost all traits we have at least 1 SNP which meets these criteria and that loosening them will increase the chances of including SNP_Xs as instruments, the choice of parameters used so far seems reasonable.

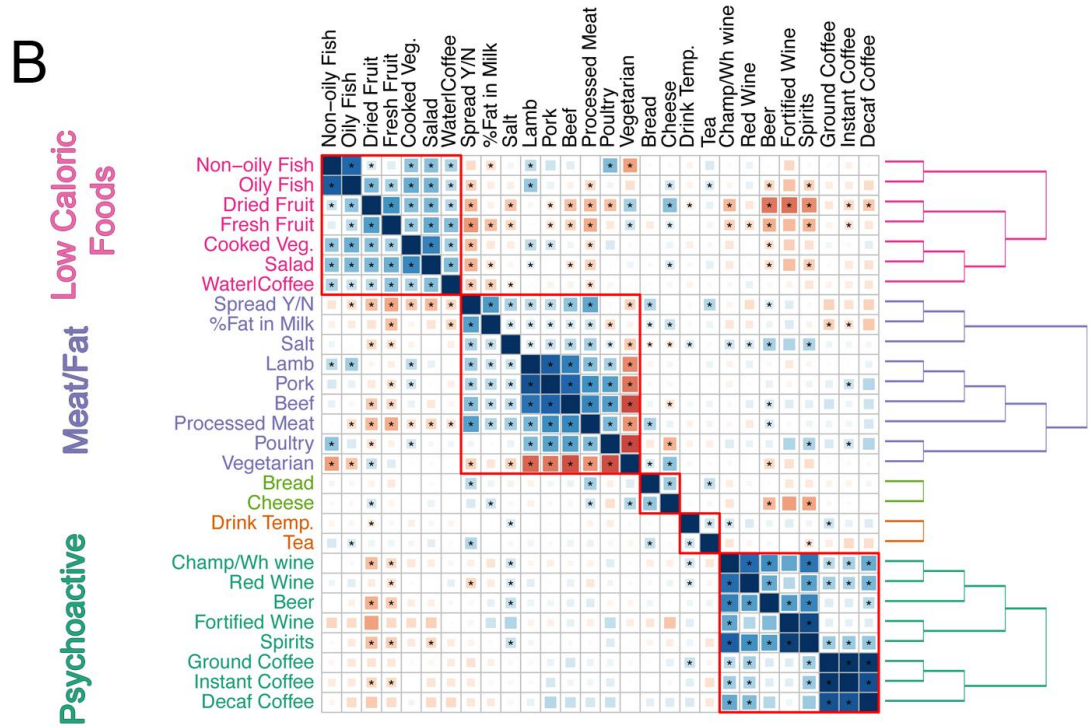
Figure S6. Corrected-to-raw ratio (CRR) successfully distinguishes mediated and non-mediated associations. (a) *Graph showing mediated and non-mediated pathways. The values of CRR that different types of simulated SNPs (Gx, Gy, Gu) assume at different explained variances (σ^2) of X->Y when (Y->X)≠0, i.e. presence of reverse causality* ***(b)****. The values we used for defining a "non-mediated" variant are highlighted in purple.* ***(c)*** *The proportion of variants that are truly Gy, that is directly associated with the trait of interest, across a range of CRR.* ***(d)*** *The overall proportion of variants directly associated with the trait (SNPy) whose CRR falls inside the specified ranges, i.e., the probability of detecting SNPy over all possible scenarios.When the effect of Y->X is equal to zero, Gy is clearly distinguishable from Gx and Gu using CRR (Fig. S5), however, when (Y->X) increases, values of CRR for both Gy and Gx start varying and overlapping (Fig. S6b). We thus determined which values of CRR would maximise the probability of correctly selecting Gy under all scenarios. Clearly the parameters we have chosen for defining a "non-mediated" SNPs maximise both the probability of correctly selecting a SNPy.*



2.3 Effects of GWAS correction on genetic correlations. To investigate how the mediation procedure affected the genetic correlations amongst the consumption traits, we compared the correlation patterns using the raw and corrected results (Table S10). Overall, the two genetic correlation matrices were very similar, but with some important differences. In particular, the number of Bonferroni-corrected significant correlations diminished in the corrected results, reflecting the fact that the conditioning on the health-related traits will weaken those correlations which are partly due to those traits. The hierarchical clustering of the traits (Fig. S7a) using the two different matrices shows an improvement in the quality of the clustering with the groups formed using the corrected r_G matrix being more interpretable than the raw ones. While the group composed of healthy foods is stable across methods, we can see that using the raw r_Gs, salt clusters with beer and strongly alcoholic beverages, while wine clusters closer to coffee than other alcoholic drinks, perhaps reflecting medical advice. Also, the fatty foods (percentage fat in milk and adding spread to bread) cluster in an unexpected way, grouping closer to healthy foods than to meat. Looking at the corrected r_Gs, clustering accords better with common sense, for example fatty foods grouping with meat and salt.

Fig S7. Clustering of food consumption traits before and after correction. (a) *Comparison between the hierarchical clustering of the food traits based on the uncorrected (on the left) and corrected (on the right) genetic correlations. Black lines connect the same traits for which the clustering has changed. Dendrograms connect the items in each case with the boldness of the line representing the strength of support for the tree nodes. Unique nodes are represented with a dashed line while shared nodes with a bold one. The thickness of the line is thicker for conserved higher level nodes.* ***(b)*** *Genetic correlation plot amongst the food traits. The lower triangle reports the corrected genetic correlation results while the upper triangle reports the uncorrected ones. The stars report the Bonferroni-corrected significant correlations. The dendrogram and the boxes represent the clustering according to the ICLUST algorithm.*





The clustering results show that the mediation correction procedure has been at least partly able to remove the genetic correlations due to shared heritable factors, rather than common biology. Such correction is extremely important, not only for creating homogeneous clusters of traits for multivariate GWAS, but also because these may change in a population- or age-dependent manner. It is reasonable to believe that some of these strong mediating factors (e.g. LDL cholesterol) are due to the older age of the samples in UKB (indeed ~27% are prescribed lipid-lowering therapy and will thus likely have been given medical advice to change their diet). It will be interesting to compare with a younger population less affected by perceived or actual medically advised lifestyle changes.

To explore the shared genetic underpinnings of food choices and a broad range of complex traits, we estimated genetic correlations with 832 traits present in LDhub[11]. We identified 6967 and 4943 significant (FDR<0.05) genetic correlations for raw and corrected traits, respectively across a large number of traits (Table S10, interactive view available at https://npirastu.shinyapps.io/rg_plotter_2/). The correction affected greatly not only the genetic correlations with the traits used for the correction but also those with many others. We can only highlight a number of examples of changes in genetic correlations here. A notable instance is that prior to adjustment, CVD and percentage fat in milk showed a genetic correlation of -0.24, i.e. decreasing the %fat increased the chances of CVD, but after correction, r_G was 0.02. Another example is again cheese consumption which has a genetic correlation with a longer paternal lifespan of 0.5 before adjustment, but only 0.2 afterwards. These results are particularly important because they suggest that the recent epidemiological findings associating higher consumption of fat in milk with protection from CVD[18,19] may be due to confounding and caution should be used when defining dietary policies.

2.4 Multivariate association analysis. Clustering of the traits using ICLUST identified five different groups (Fig 7b): one composed of increased meat, fat, salt and decreased vegetarianism (labeled as “Meat/Fat”), one made up of alcoholic beverages and coffee (labeled “Psychoactive”)

and one comprised of healthier items such as fish, fruit and vegetables (labeled “Low Calorie Foods”). Two final groups contained only two items each: drink temperature and tea; and cheese and bread; these were not used for the MV analysis. In order to explore if additional loci influence these groups we ran a multivariate GWAS using the package MultiABEL, which performs MANOVA on summary statistics. 168 additional associations, including 42 novel loci not identified at the single trait analysis, were identified in multivariate analysis of the three main food groups (Table S5) An example of these group-level loci is rs17400325, a non-synonymous variant at *PDE11A* associated with the consumption of Low Calorie Foods. When each trait was examined singularly we found that the C allele was associated with higher fresh and dried fruit consumption and a lower consumption of fish and vegetables. Mutations in this gene are responsible for primary pigmented nodular adrenocortical disease-2 (OMIM:610475), which leads to high cortisol levels, which in turn are associated with a higher consumption of highly palatable foods[20].

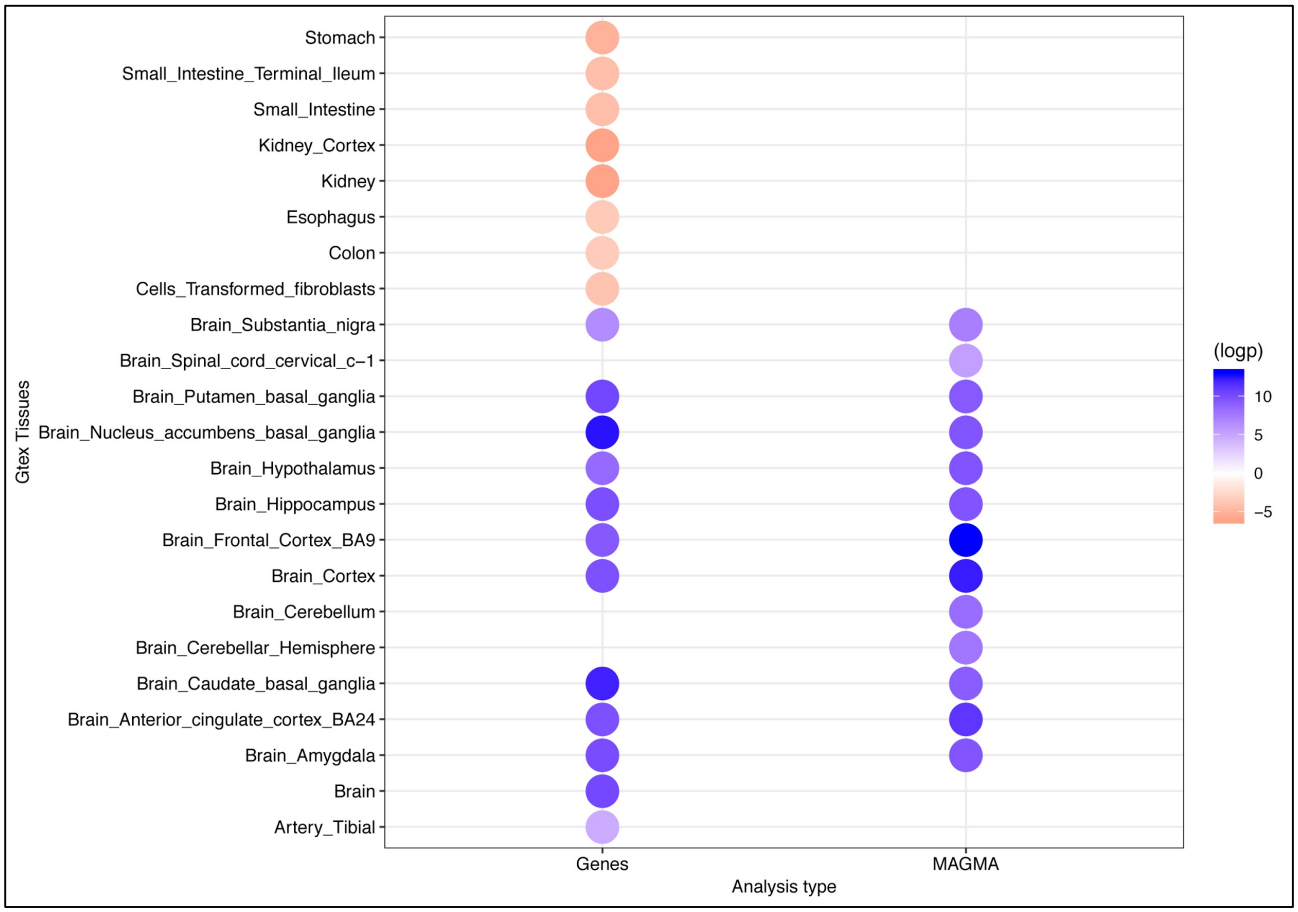
2.5 Functional annotation of food consumption genes. We used several approaches to understand the biological underpinnings of food choice. First we ran stratified LD-score regression[12] using the bias-corrected GWAS (Fig. S8). Looking at functional annotation (Fig. S8a), we found a strong enrichment for almost all food traits in conserved genomic regions with the exception of being vegetarian, decaffeinated coffee and fortified wine consumption. This is not surprising if we consider that nutrition is one of the most basic biological functions. We next looked at tissue enrichment in the Gtex[21] and Franke lab expression[22] datasets and epigenetic signatures from Roadmap[23]. There was substantial agreement across the two expression datasets with enrichment mostly limited to brain areas linked to reward and feeding, e.g. hypothalamus, nucleus accumbens, putamen. Most food traits were also enriched for epigenetic marks annotated to the male and female fetal brain, which underlies the importance of foetal development in determining food choices.

Fig S8. Heatmap of tissue and functional enrichments. *The colour is proportional for the enrichment revealed by stratified LD-score regression. Only correlations with FDR<0.05 are reported.* ***(a)*** *Enrichment among different classes of functional annotation.* ***(b)*** *Tissue enrichment from Gtex expression.* ***(c)*** *Tissue enrichment from ROADMAP epigenetics.* ***(d)*** *Tissue enrichment from the Franke lab dataset.*

Tissue over-representation analysis using MAGMA[13] (which first runs a gene-wide test and then measures enrichment using those which results significant) on the merged GWAS results (for each SNP the lowest p-value was used) confirmed the results from LD-score regression highlighting the same brain areas, e.g. substantia nigra, nucleus accumbens, hypothalamus, amygdala, known to influence food choices and reward (Fig. S9). Very similar results were obtained when the analysis was conducted using only the prioritised genes in the “direct effect only” loci. In this case, as well

as over-expression in the brain, we also detected under-expression in the kidney, stomach, pancreas, oesophagus, colon, lung and small intestine (Fig. S9).

Fig. S9. **Dotplot of the overexpression analysis run on the prioritised genes from the non-mediated loci and the overrepresentation analysis performed with MAGMA.** The overexpressed tissue involved by the two methods were highly overlapping with the analysis performed on the prioritised genes showing also the tissues in which there is evidence of underexpression.



2.6 Network analysis. The fact that the genes we prioritised show the same enrichment pattern as stratified LD-score regression and MAGMA, also suggests that the prioritisation is in most cases correct. In order to explore any interactions between the selected genes, we used STRING[24] to build an interaction network between them. A large network is revealed (Fig. 4, Table S13), including 132 genes (out of 215 overlapping the STRING database), sharing 224 edges ($p<1 \times 10^{-16}$). To identify if there were groups of genes which were more interconnected than the others, we performed community identification using Leuvain's method[24,25]. Ten communities are identified, each with specific characteristics in terms of function, cellular localisation and preferential expression (Fig. S11 for and overview and FigS 12-21 for specific comunities, Table S15 for significant Gene Ontology terms, Fig. S10, Table S16 for significant tissue enrichment). For

example, community 1 genes are linked to numerous biological processes ranging from feeding behaviour and taste to hormone binding and transport of fatty acids, and preferentially include the genes expressed in several brain areas and the liver. Genes in community 2, on the other hand, are linked to energy and glucose metabolism, are preferentially located in the mitochondrion and are over-expressed in the skeletal muscle and tibial artery. Another interesting example is community 8 which contains genes specifically over-expressed in the brain and which are involved in synaptic assembly and organisation and in neurotransmitter secretion, while other communities relate to steroid hormone response, acetylcholine receptor regulation, drug metabolism and Golgi-mediated transport. Thus, although the overall expression analysis strongly links dietary choices to the central nervous system, there are actually several different groups of genes at play, with specific functions in specific tissues.

Figure S10 Tissues which overexpress the genes in each community.

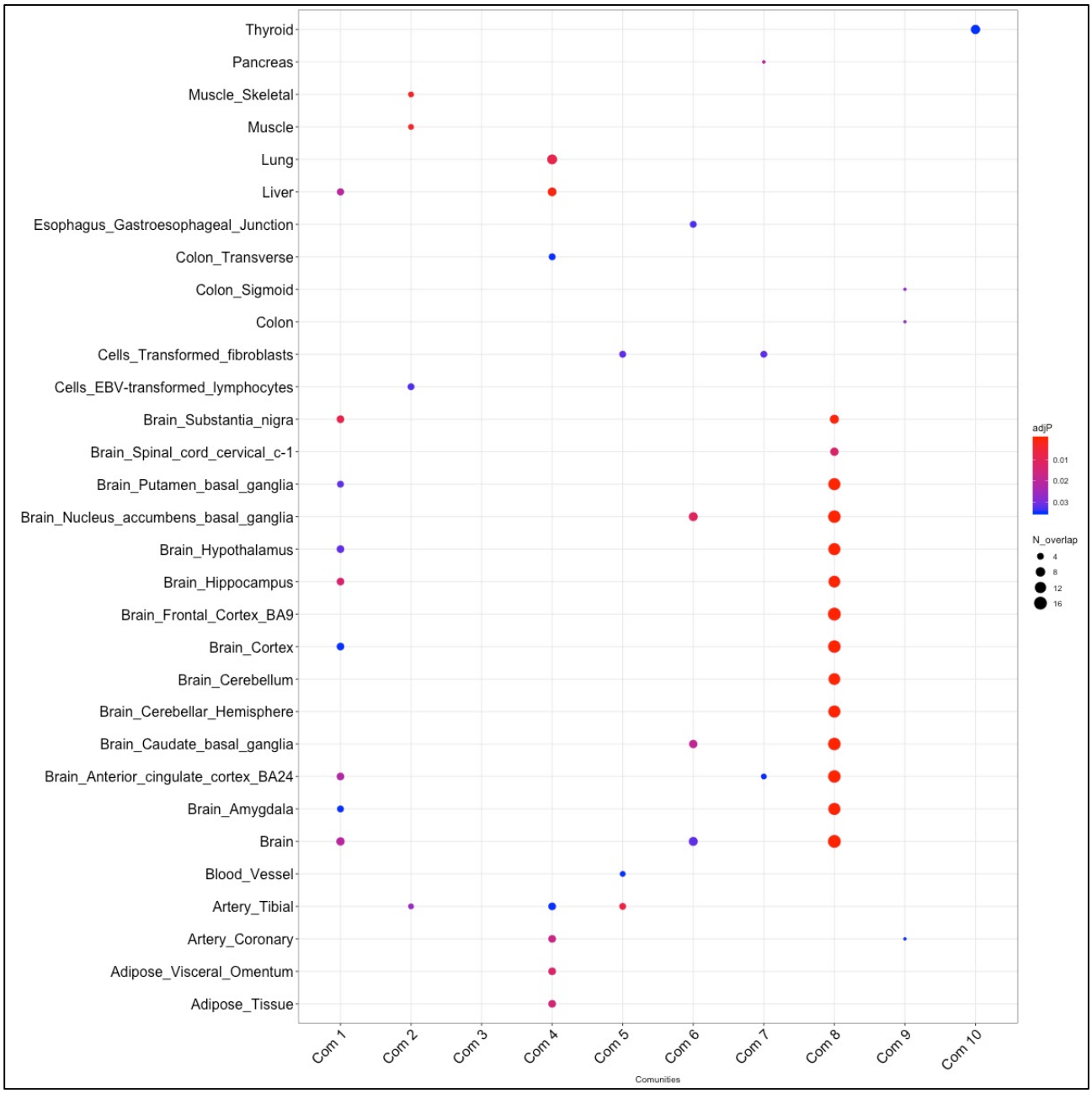


Figure S11 Overlap in Go-terms between different communities. *The figure shows that there is no overlap (with the exception of 2 terms) between the terms enriched in each community. The labels have been removed as the plot is meant to only show the overlaps. Figure S12-22 show the enriched terms for each community separately.*

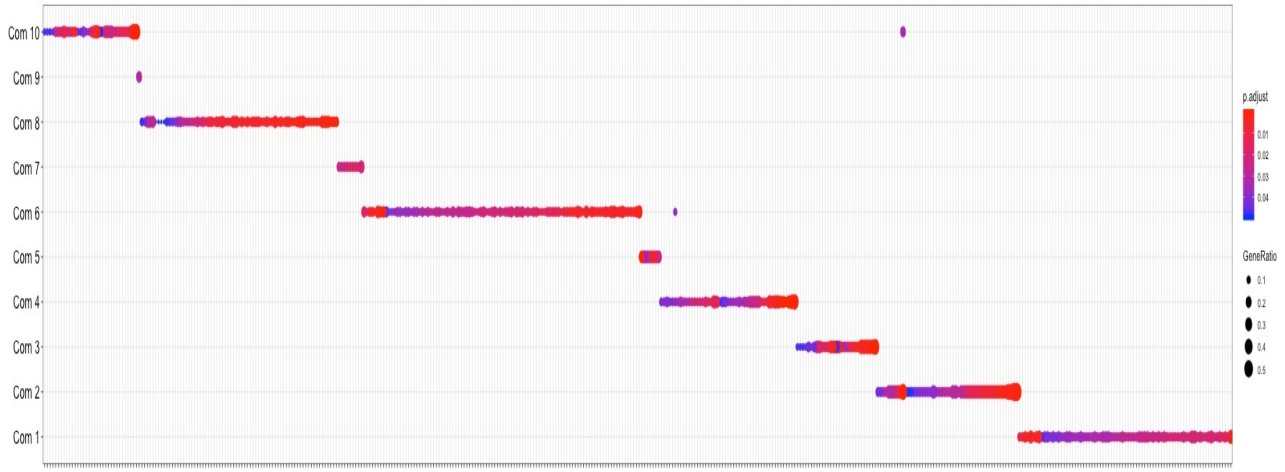


Figure S12 Enriched GO-Terms for community 1.

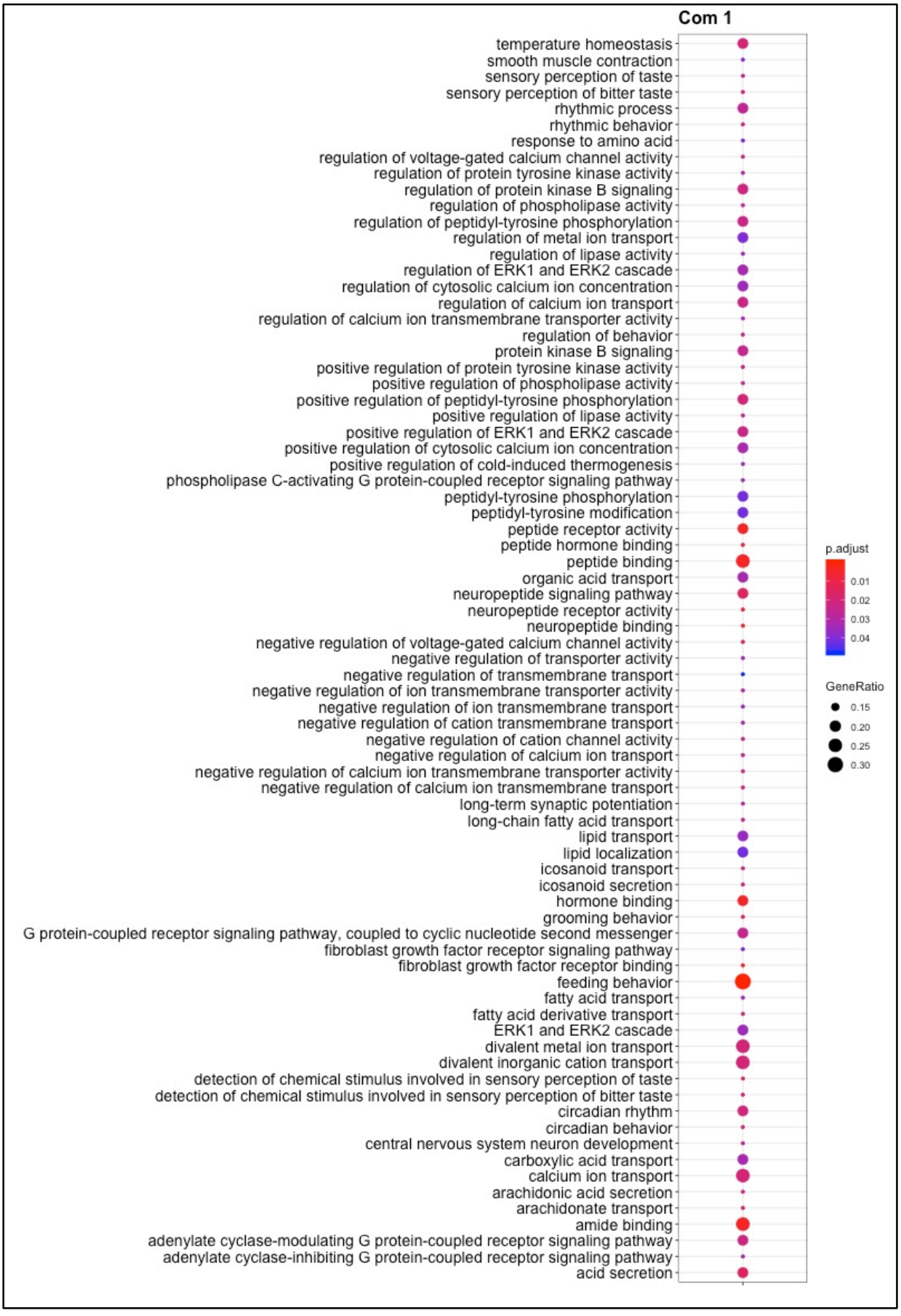


Figure S13 Enriched GO-Terms for community 2

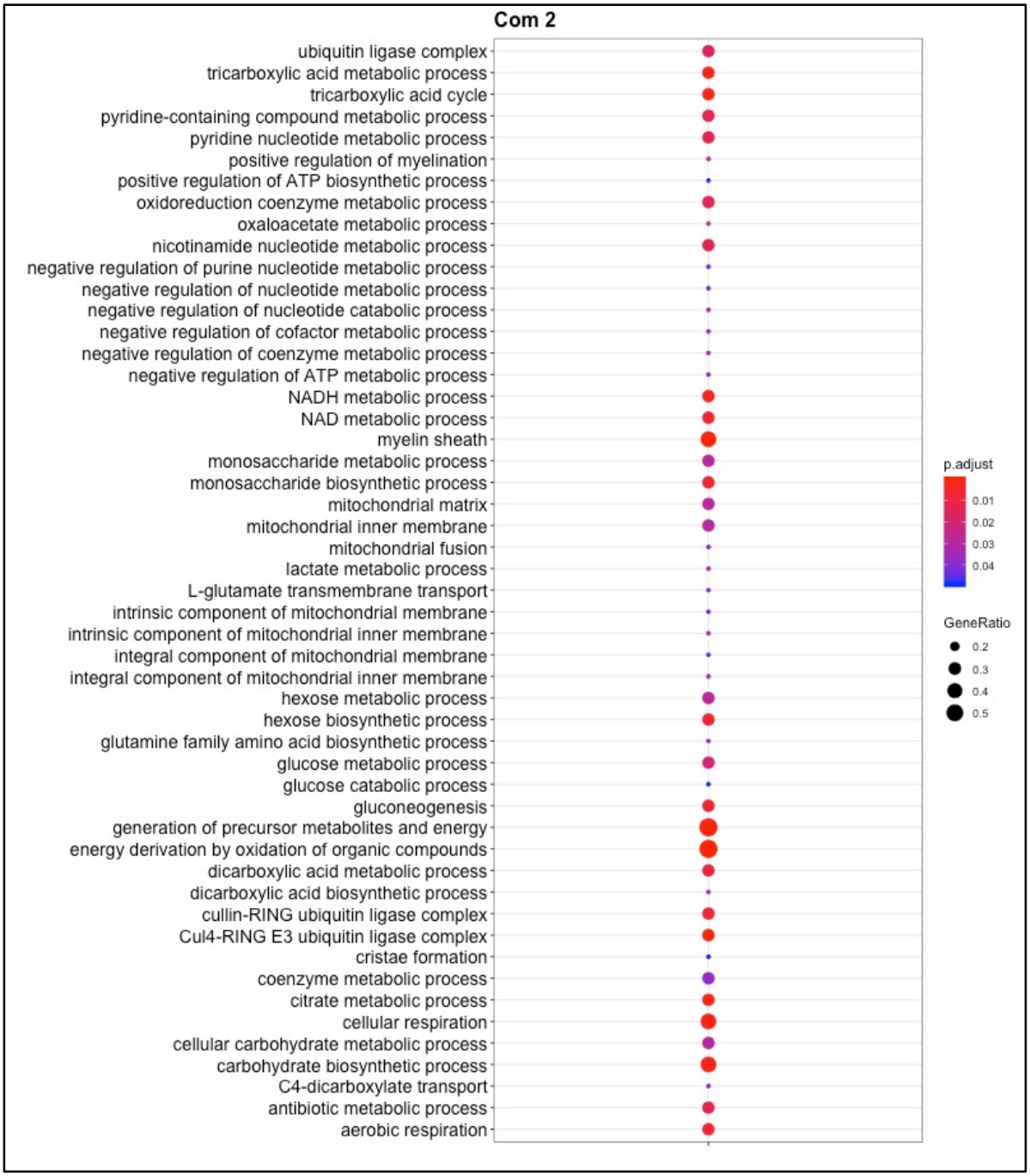


Figure S14 Enriched GO-Terms for community 3

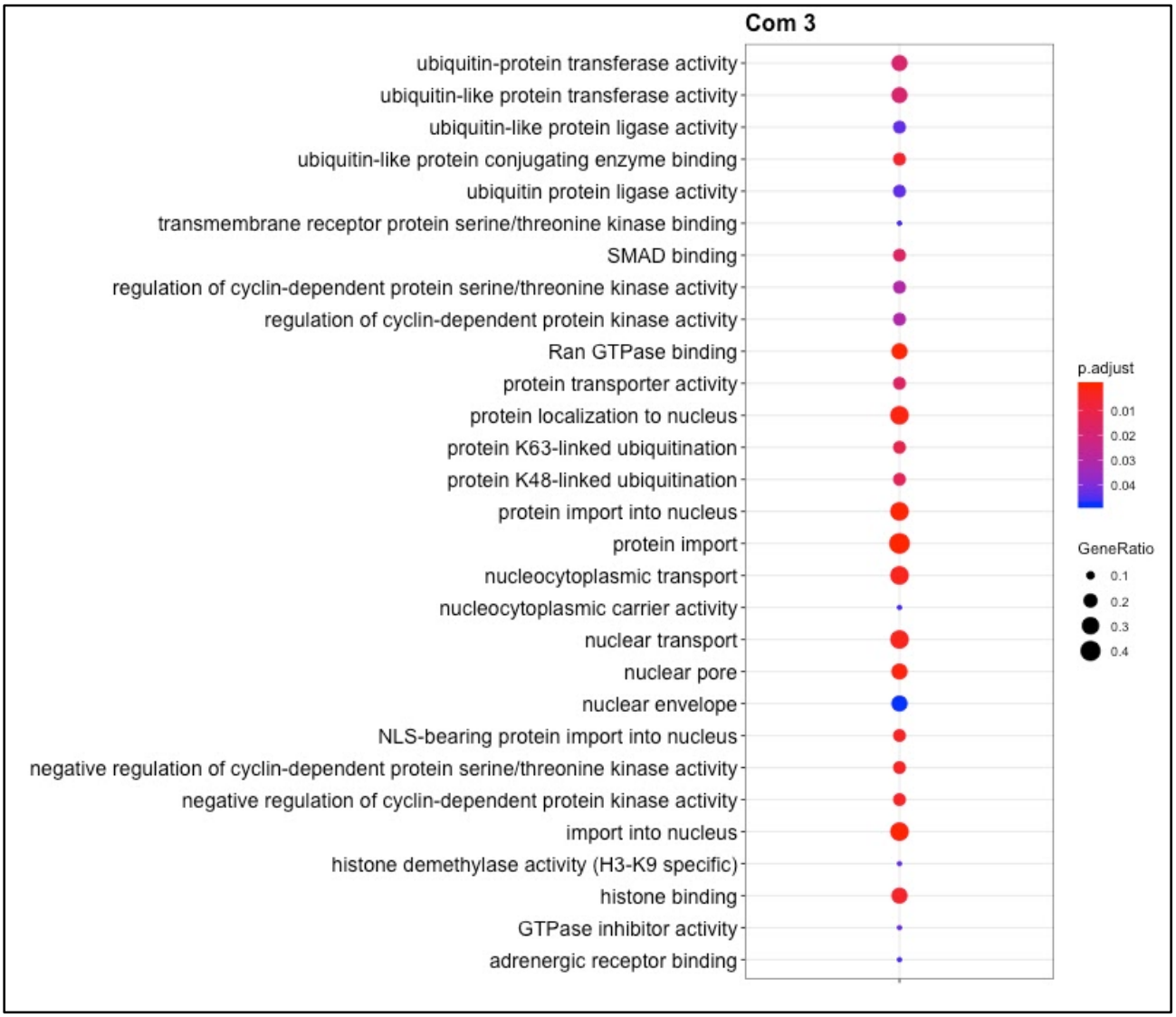


Figure S15 Enriched GO-Terms for community 4

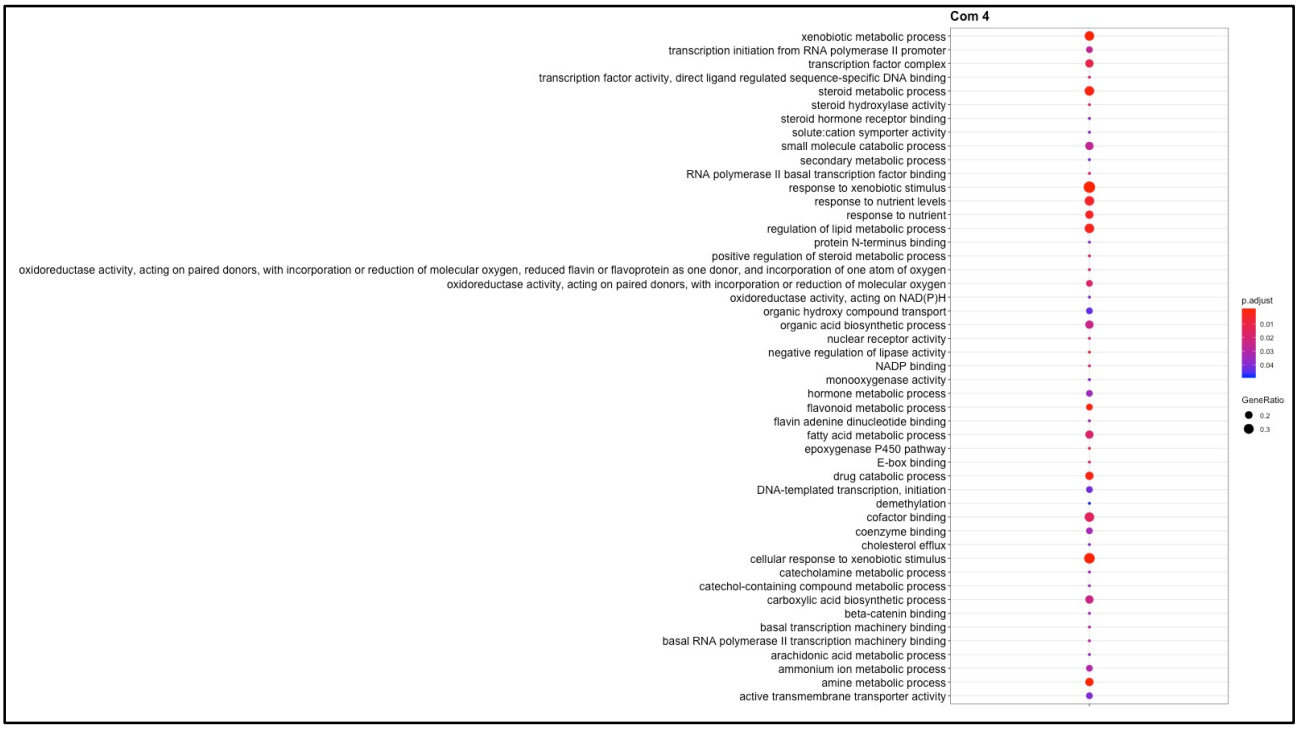


Figure S16 Enriched GO-Terms for community 5

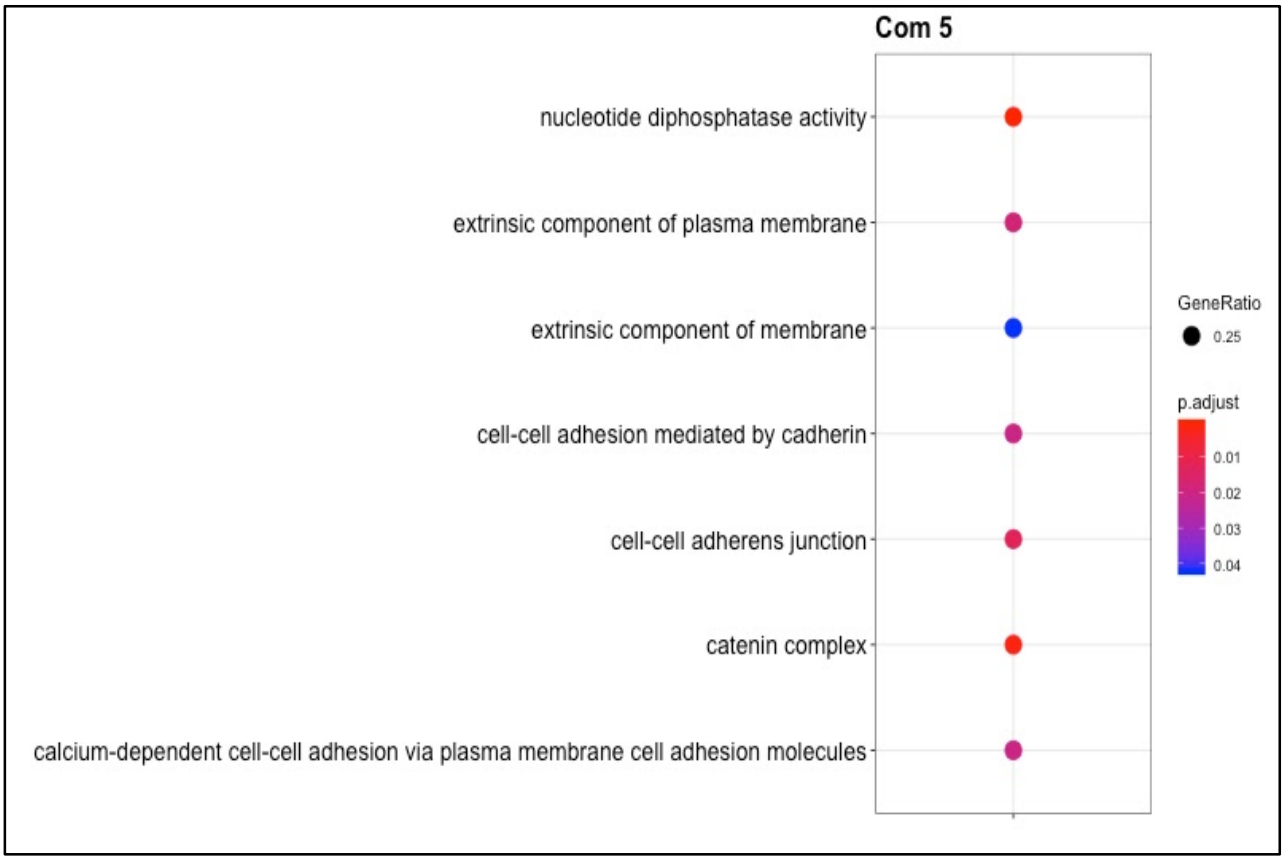


Figure S17 Enriched GO-Terms for community 6

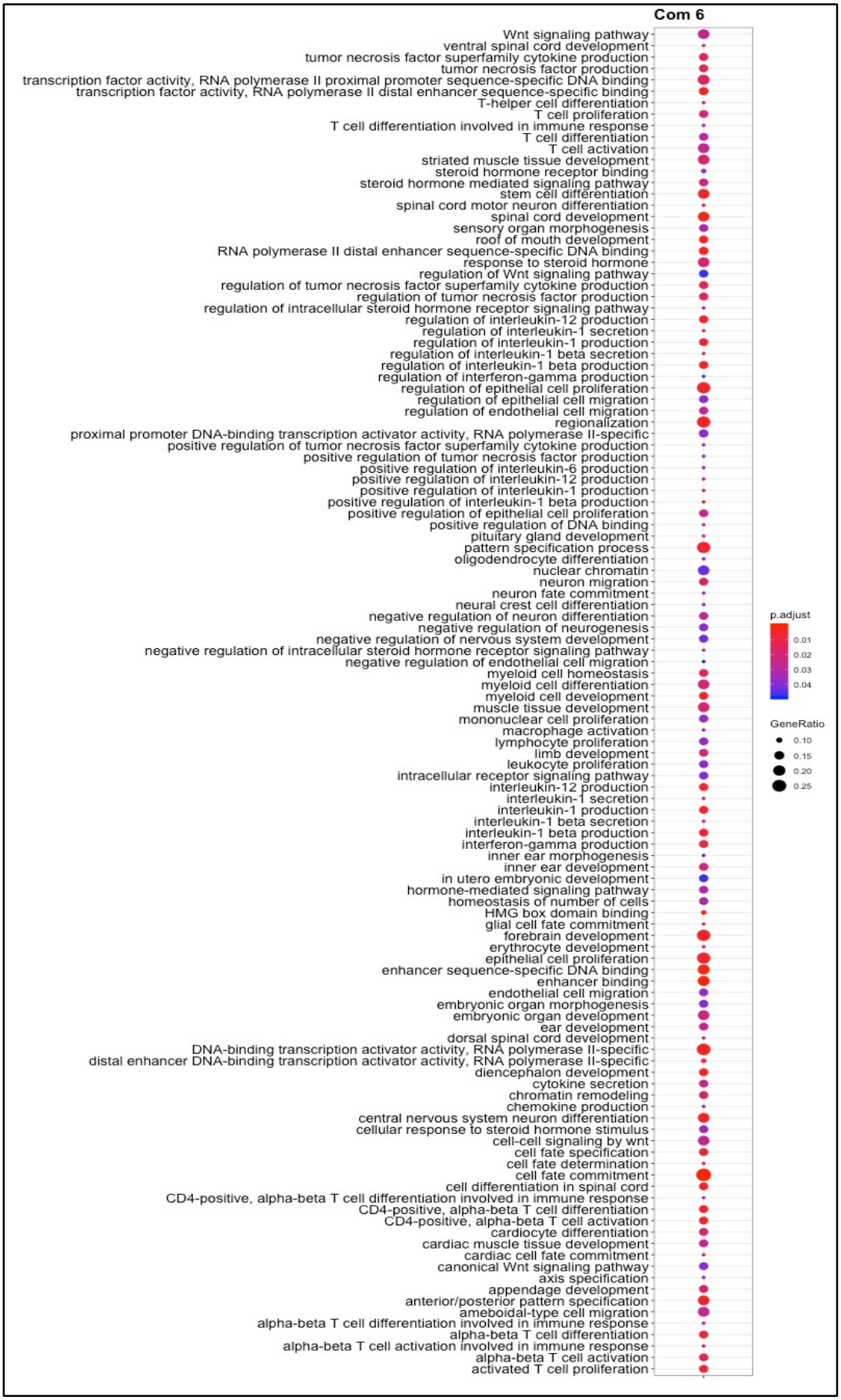


Figure S18 Enriched GO-Terms for community 7

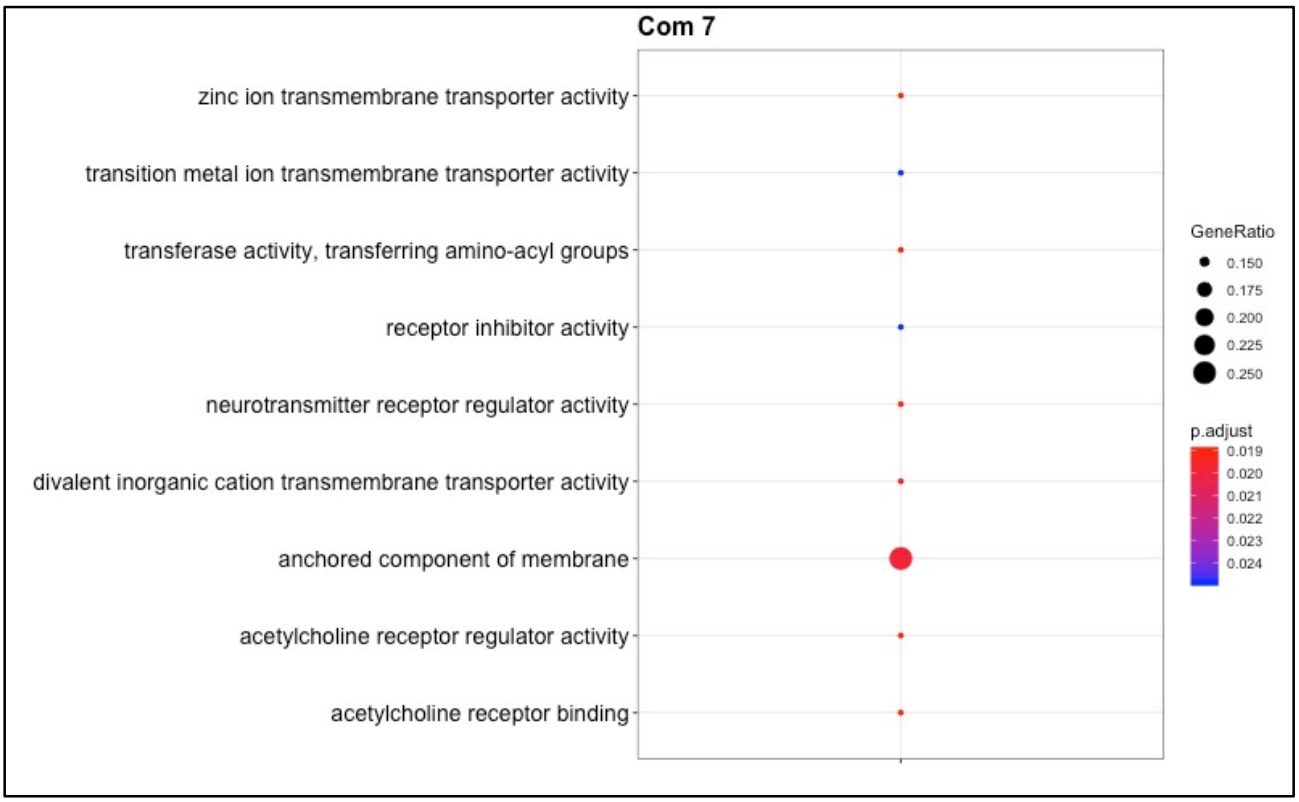


Figure S19 Enriched GO-Terms for community 8

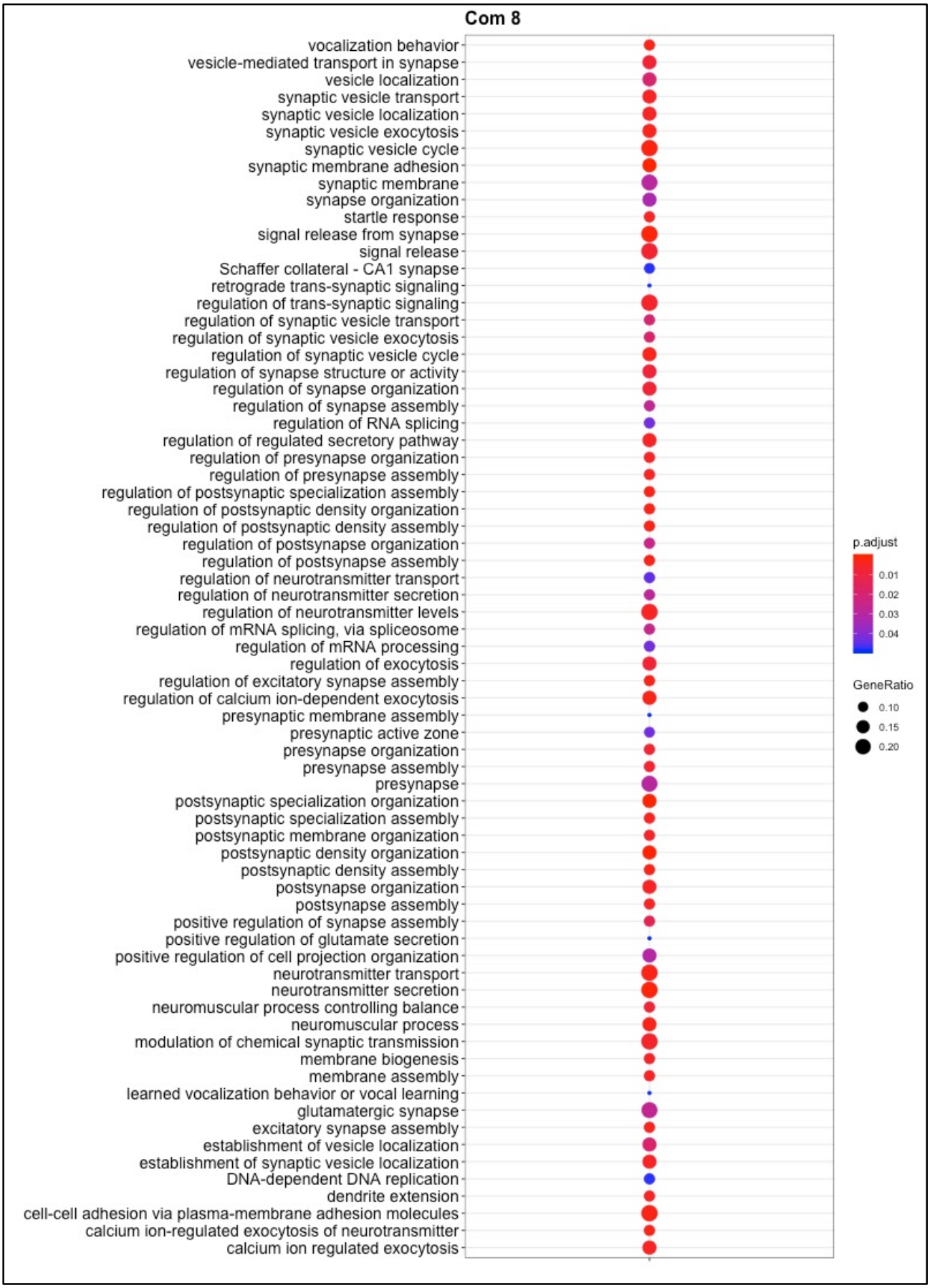


Figure S20 Enriched GO-Terms for community 9

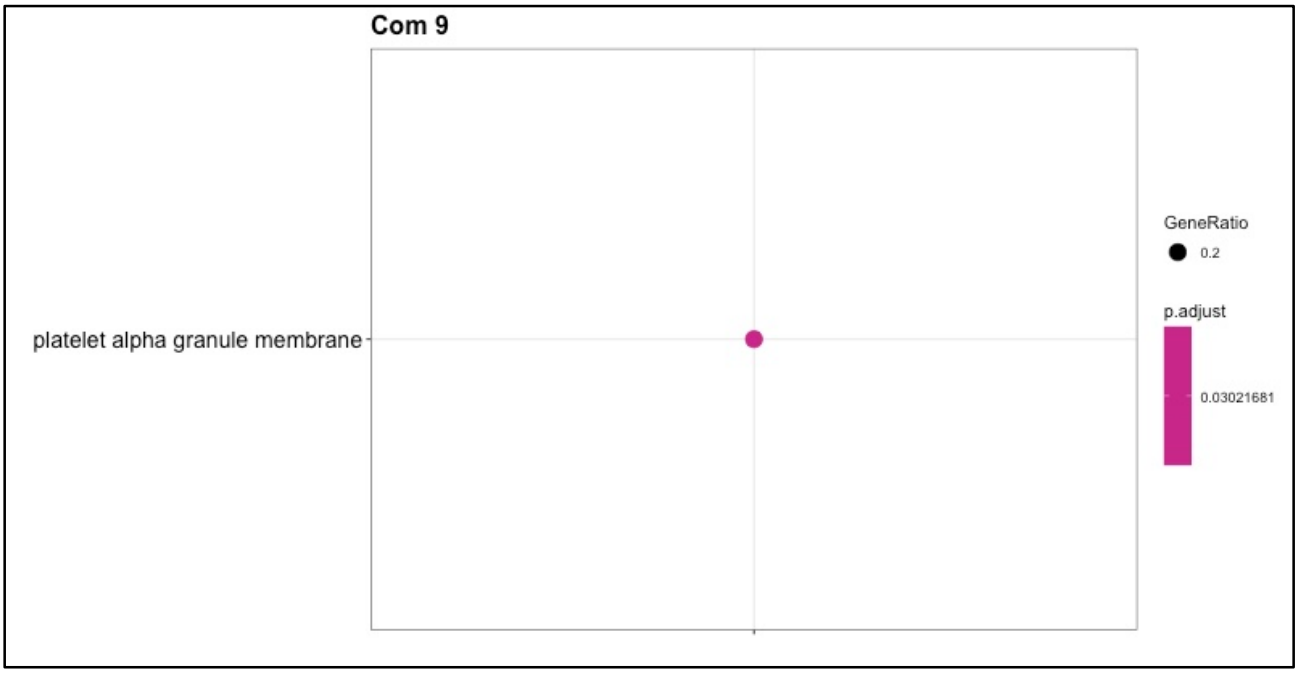


Figure S21 Enriched GO-Terms for community 10

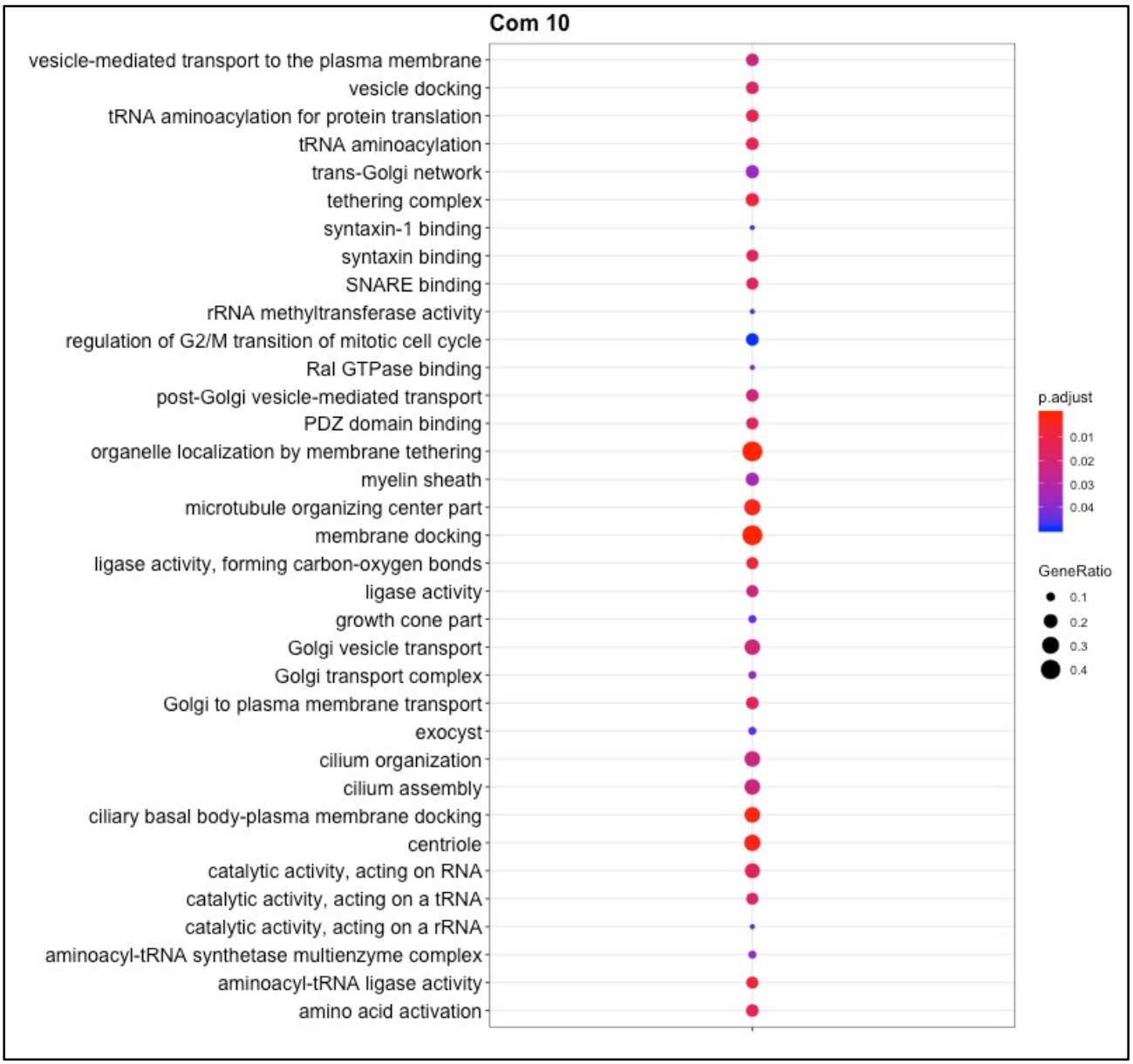
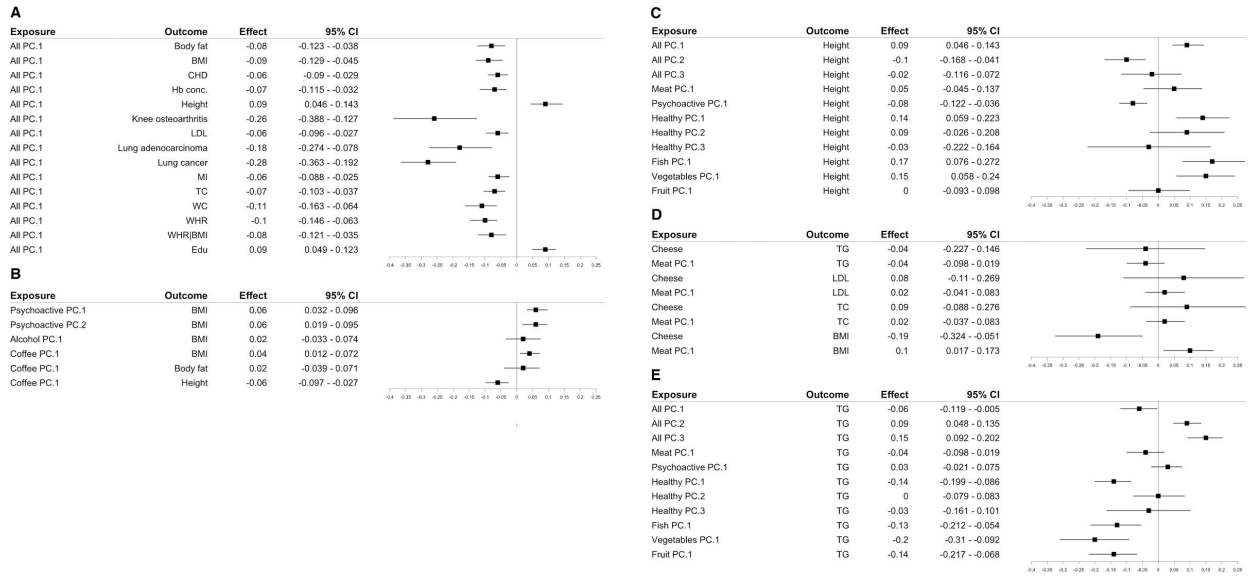


Figure S22. Selected forest plots of MR-estimated effect sizes. *For each of the six examples described in the main text. Exposure trait, outcome trait, effect size and 95% confidence intervals are reported. Abbreviations: CHD Coronary Heart Disease, BMI Body Mass Index, TG Triglycerides, MI Myocardial Infarction, TC Total Cholesterol, WC Waist Circumference, WHR Waist to Hip Ratio, WHR|BMI Waist to Hip Ratio BMI adjusted, LDL Low Density Lipoprotein*



A

Exposure	Outcome	Effect	95% CI
All PC.1	Body fat	-0.08	-0.123 - -0.038
All PC.1	BMI	-0.09	-0.129 - -0.045
All PC.1	CHD	-0.06	-0.09 - -0.029
All PC.1	Hb conc.	-0.07	-0.115 - -0.032
All PC.1	Height	0.09	0.046 - 0.143
All PC.1	Knee osteoarthritis	-0.26	-0.388 - -0.127
All PC.1	LDL	-0.06	-0.096 - -0.027
All PC.1	Lung adenocarcinoma	-0.18	-0.274 - -0.078
All PC.1	Lung cancer	-0.28	-0.363 - -0.192
All PC.1	MI	-0.06	-0.088 - -0.025
All PC.1	TC	-0.07	-0.103 - -0.037
All PC.1	WC	-0.11	-0.163 - -0.064
All PC.1	WHR	-0.1	-0.146 - -0.063
All PC.1	WHR|BMI	-0.08	-0.121 - -0.035
All PC.1	Edu	0.09	0.049 - 0.123

B

Exposure	Outcome	Effect	95% CI
Psychoactive PC.1	BMI	0.06	0.032 - 0.096
Psychoactive PC.2	BMI	0.06	0.019 - 0.095
Alcohol PC.1	BMI	0.02	-0.033 - 0.074
Coffee PC.1	BMI	0.04	0.012 - 0.072
Coffee PC.1	Body fat	0.02	-0.039 - 0.071
Coffee PC.1	Height	-0.06	-0.097 - -0.027

C

Exposure	Outcome	Effect	95% CI
All PC.1	Height	0.09	0.046 - 0.143
All PC.2	Height	-0.1	-0.168 - -0.041
All PC.3	Height	-0.02	-0.116 - 0.072
Meat PC.1	Height	0.05	-0.045 - 0.137
Psychoactive PC.1	Height	-0.08	-0.122 - -0.036
Healthy PC.1	Height	0.14	0.059 - 0.223
Healthy PC.2	Height	0.09	-0.026 - 0.208
Healthy PC.3	Height	-0.03	-0.222 - 0.164
Fish PC.1	Height	0.17	0.076 - 0.272
Vegetables PC.1	Height	0.15	0.058 - 0.24
Fruit PC.1	Height	0	-0.093 - 0.098

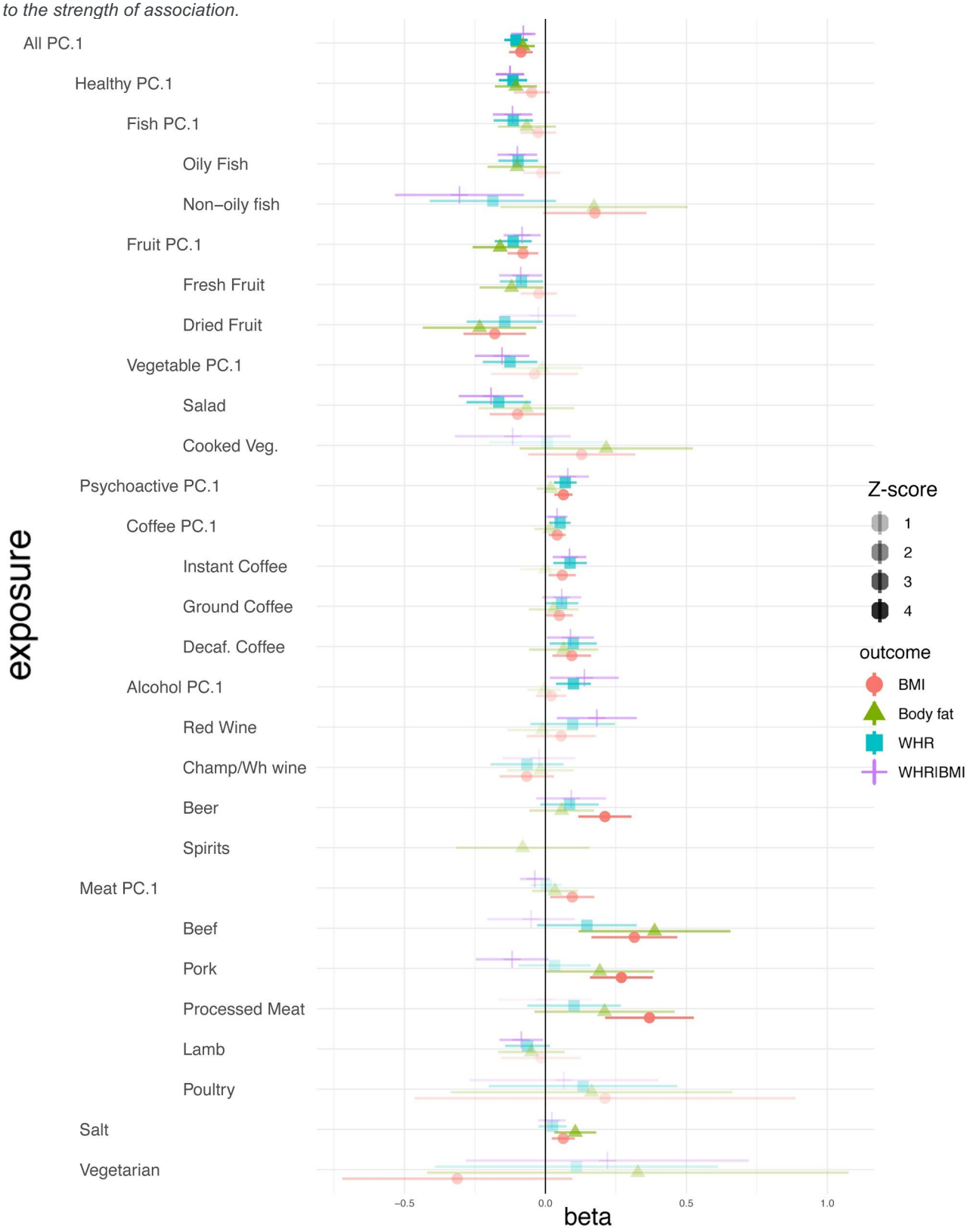
D

Exposure	Outcome	Effect	95% CI
Cheese	TG	-0.04	-0.227 - 0.146
Meat PC.1	TG	-0.04	-0.098 - 0.019
Cheese	LDL	0.08	-0.11 - 0.269
Meat PC.1	LDL	0.02	-0.041 - 0.083
Cheese	TC	0.09	-0.088 - 0.276
Meat PC.1	TC	0.02	-0.037 - 0.083
Cheese	BMI	-0.19	-0.324 - -0.051
Meat PC.1	BMI	0.1	0.017 - 0.173

E

Exposure	Outcome	Effect	95% CI
All PC.1	TG	-0.06	-0.119 - -0.005
All PC.2	TG	0.09	0.048 - 0.135
All PC.3	TG	0.15	0.092 - 0.202
Meat PC.1	TG	-0.04	-0.098 - 0.019
Psychoactive PC.1	TG	0.03	-0.021 - 0.075
Healthy PC.1	TG	-0.14	-0.199 - -0.086
Healthy PC.2	TG	0	-0.079 - 0.083
Healthy PC.3	TG	-0.03	-0.161 - 0.101
Fish PC.1	TG	-0.13	-0.212 - -0.054
Vegetables PC.1	TG	-0.2	-0.31 - -0.092
Fruit PC.1	TG	-0.14	-0.217 - -0.068

Figure S23 Effect of food on obesity related measures. *The forest plot compares the effect of each food trait on four obesity related measures: BMI, Body Fat, Waist to Hip Ratio (WHR) and BMI adjusted WHR (WHR|BMI). Each color and shape represents a different obesity related measure while the transparency of the points and error bars are proportional to the strength of association.*



620 **2.6 Significant results from the Raw uncorrected analysis.**

621

622 In order to understand what would have been the impact of performing MR without using CRR for
623 filtering the IVs, we estimated Storey q-values using only the p-values coming from the Raw
624 Uncorrected set of results. The following forest plots refer to these results and compare the Raw
625 uncorrected results with the CRR filtered IVs using the uncorrected betas.
626 Of the 115 significant exposure/outcome pairs which resulted significant after multiple test
627 correction only 26 were in common with those significant at the CRR filtered analysis. In many
628 cases this is clearly due to an overestimation of the effect size. This is particularly evident when
629 looking at Cheese which seems to have a large number of beneficial effects (7) which all disappear
630 (apart from BMI which is diminished in any case) after selecting only the IVs with non-mediated
631 effects. This is unsurprising given that Cheese was the food that had the largest proportion of
632 genetic variance explained by the health related traits (~40%).
633 These results show how risky it is to make causal claims based on the naive analysis. In our case
634 we could have used these results to make claims of a huge number of beneficial effects of Cheese
635 which do seem to be true and are likely due to the fact that people who consume a larger amount
636 of cheese have a higher education and lower cholesterol which is thus creating the confounding
637 effect. This is of particular importance as MR is generally considered (when performed properly) a
638 sturdy and reliable method of testing causal relationships. However we have shown that there can
639 be issues and particular care should be used when human behaviour is involved in the definition of
640 the exposure trait.
641 A slightly different example is the case of dried fruit were, despite none of the effects are still
642 significant after using the CRR filtered IVs, comparing the forest plots seems to suggest that this
643 difference is in some cases due to an actual difference in effect size (Years of schooling, Lung
644 Cancer and Ovarian Cancer) while in the rest of the cases the difference is due to a loss in power
645 which has led to an increase in the standard errors of the estimates.
646 It is unfortunately impossible to perform a direct test of the difference in estimates as the wide
647 confidence intervals of the MR estimates do allow to have enough power to detect the differences.

This problem will be overcome in the future through the increase in power due to the increasing size of GWAs studies but at the moment we are not able in many cases to distinguish which associations are not significant any more due to power or difference in effect size. We have however reported the forest plot comparing the two methods and have provided an online tool that each researcher can evaluate all the different estimates coming from different methods and thus make up their own mind based also on external evidence.

Fig S24-S50 Forest plots of the exposure/outcome pairs significant at the raw analysis.
The forest plots represent the estimated effect sizes for all the non CRR filtered MR analyses. The squares represent the point estimates while the bars the 95% confidence intervals. Results from the Raw analysis (raw) and CRR filtered IVs (CRR) are reported. The exposure trait is indicated in the header of the plots while the row labels refer to the outcomes. Beta's always refer to standard deviations for the exposure while for the outcomes it is standard deviations for the quantitative traits and log(OR) for the disease traits.

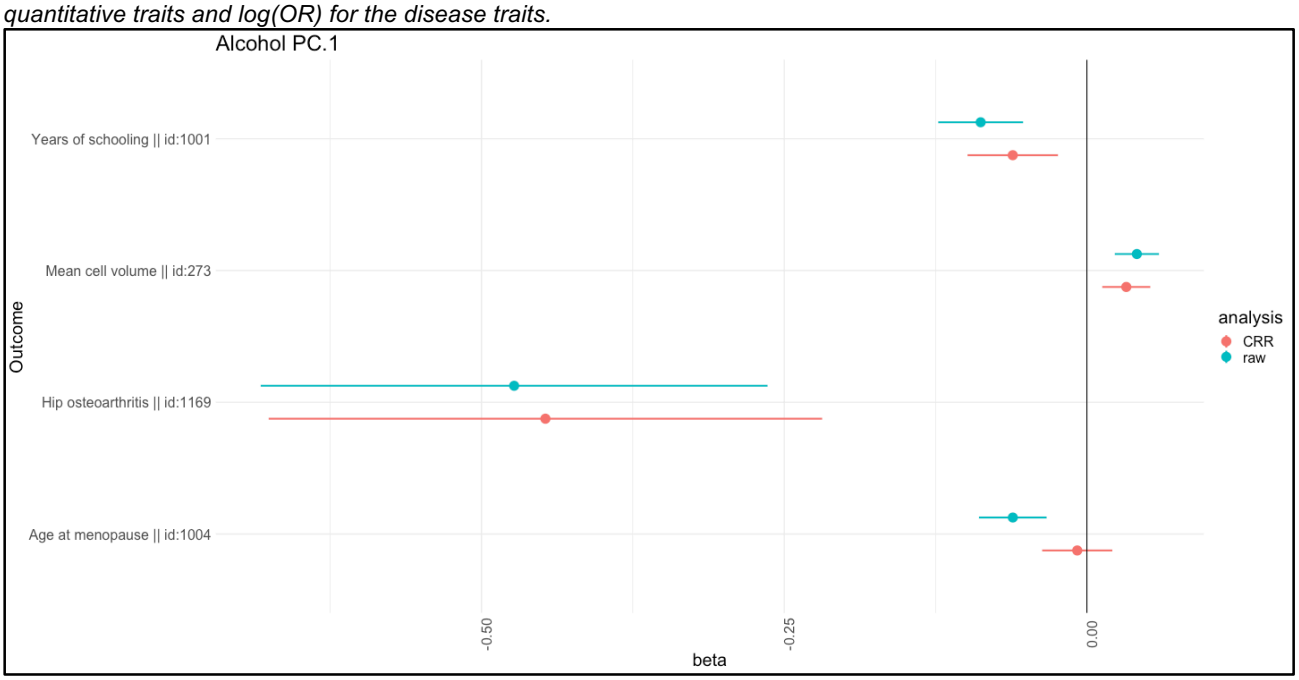


Fig S25

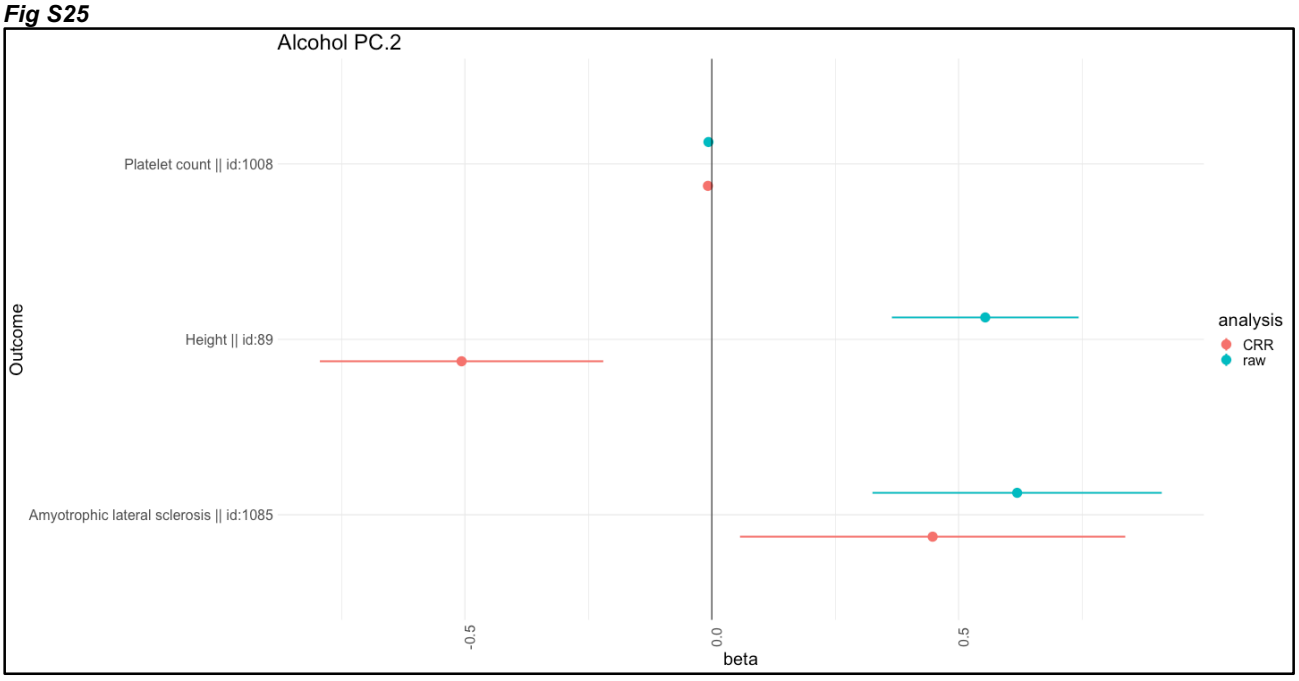


Fig S26

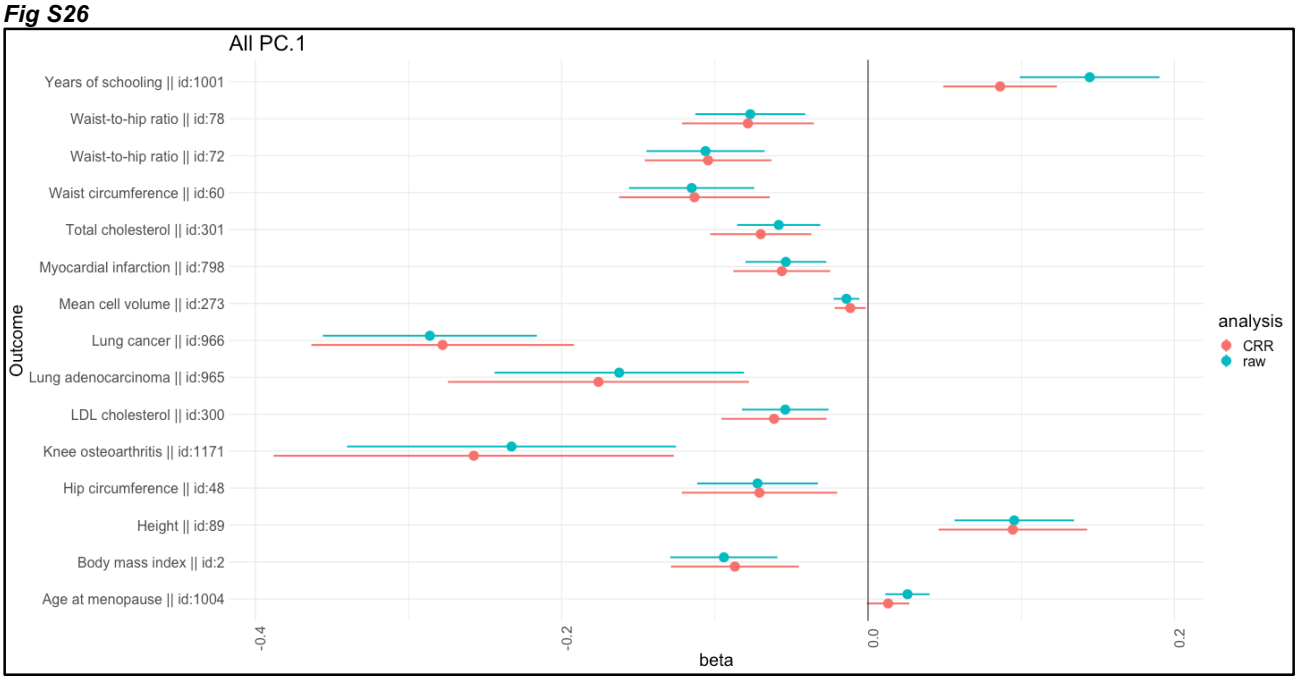


Fig S27

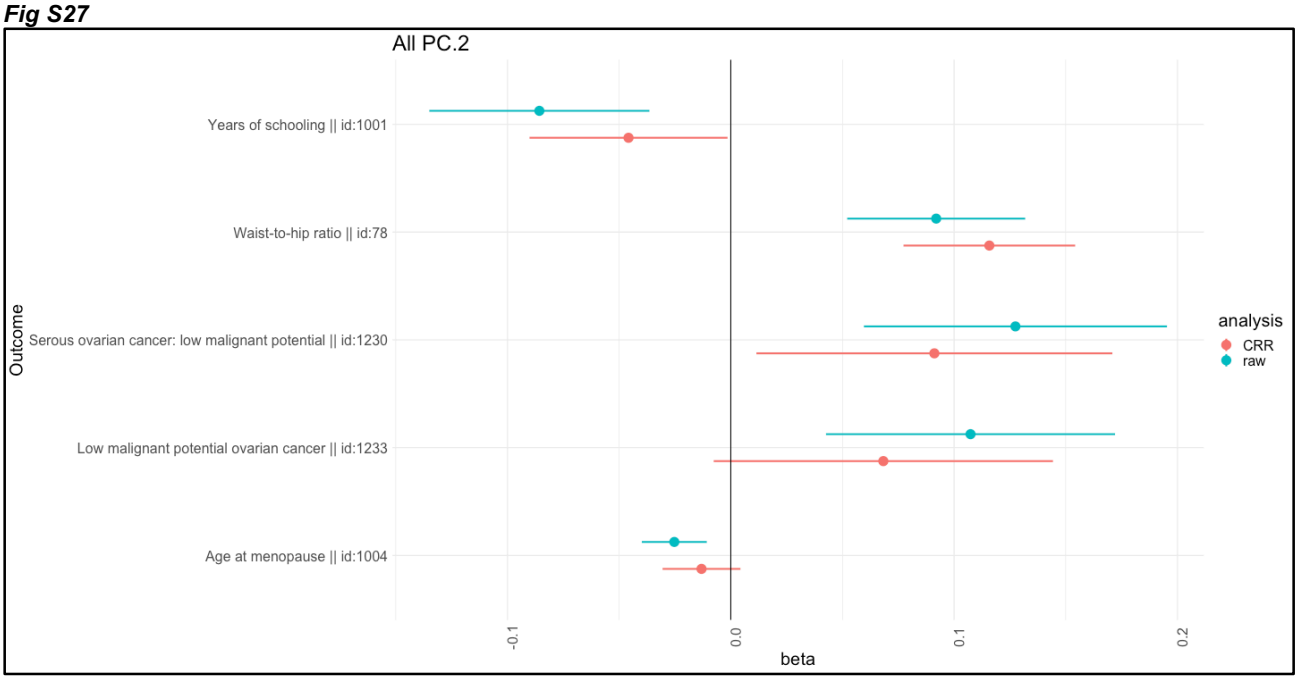


Fig S28

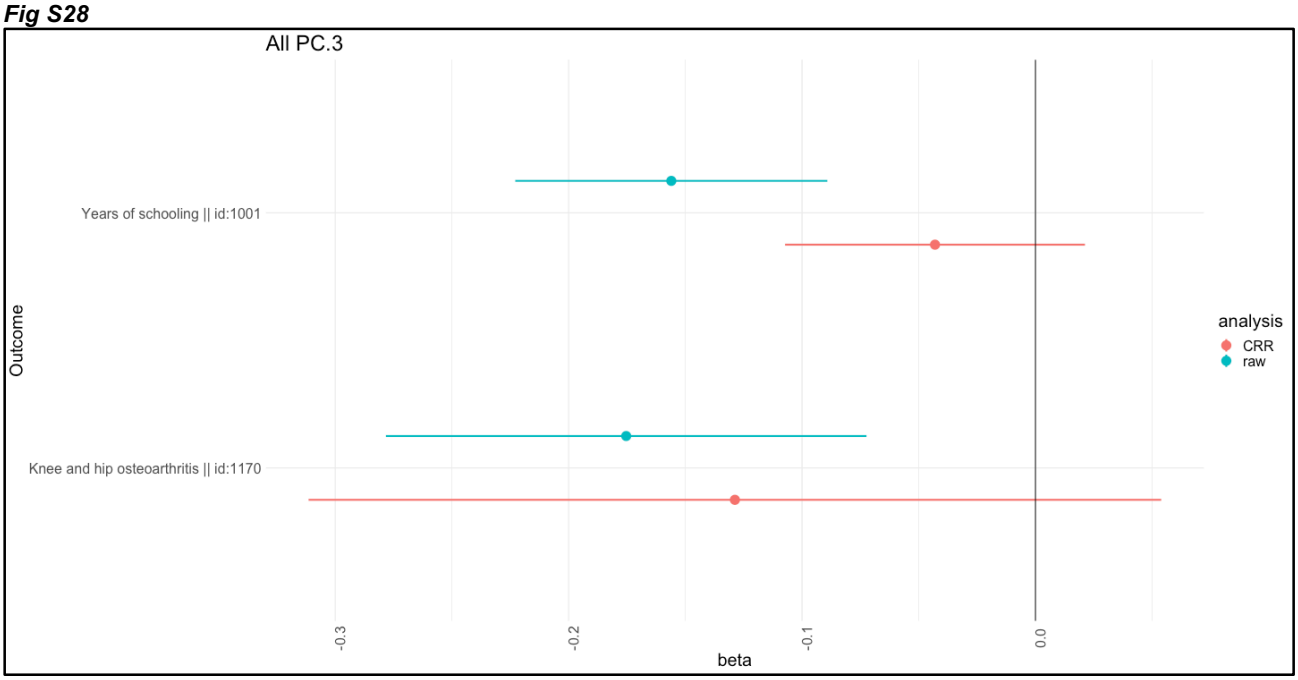


Fig S29

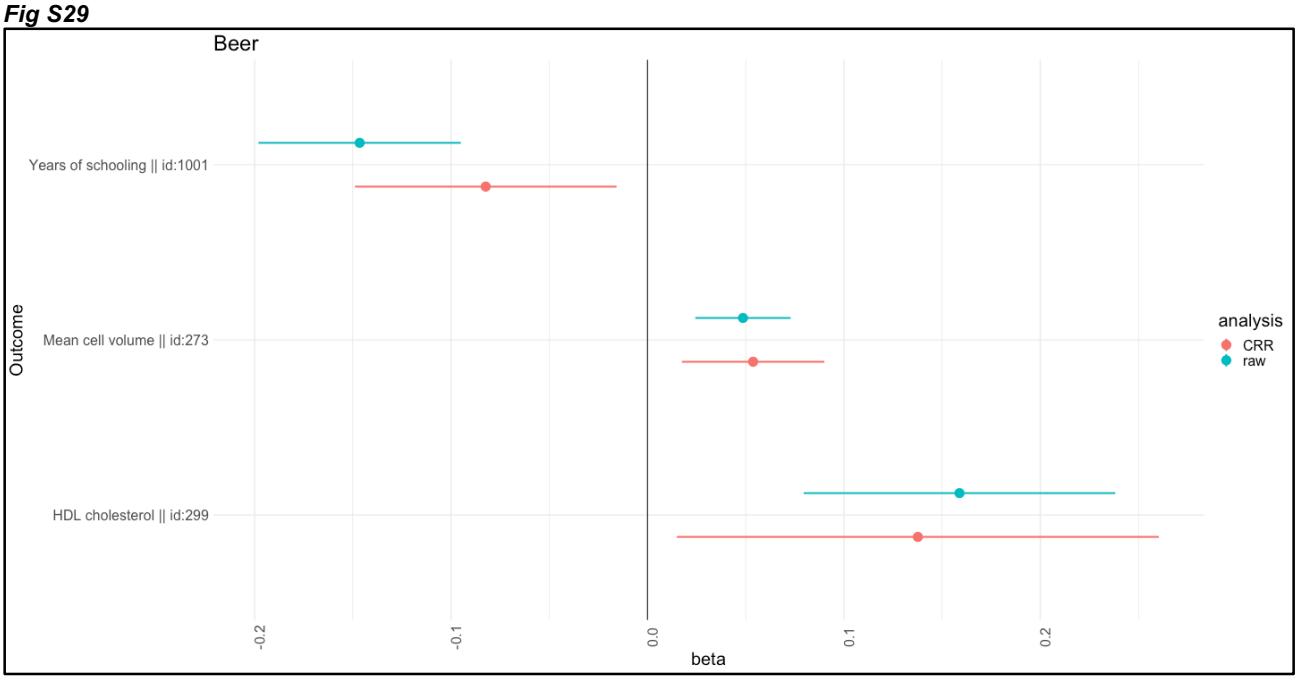


Fig S30

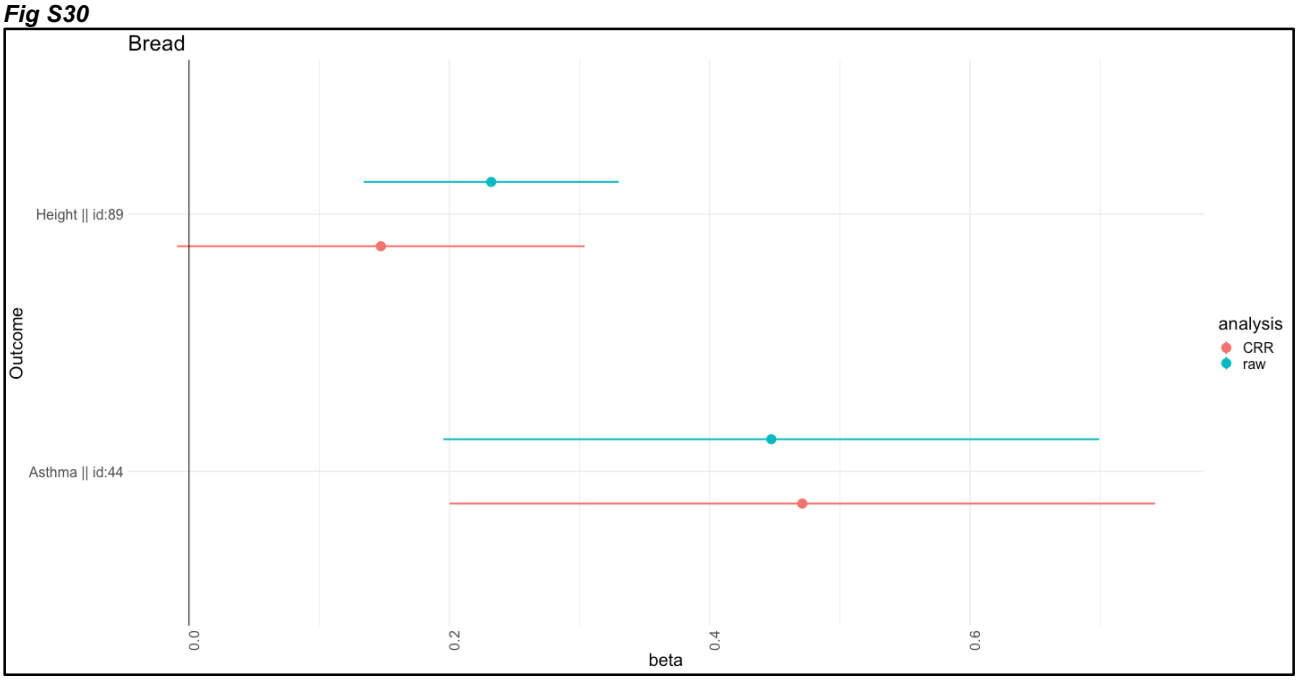


Fig S31

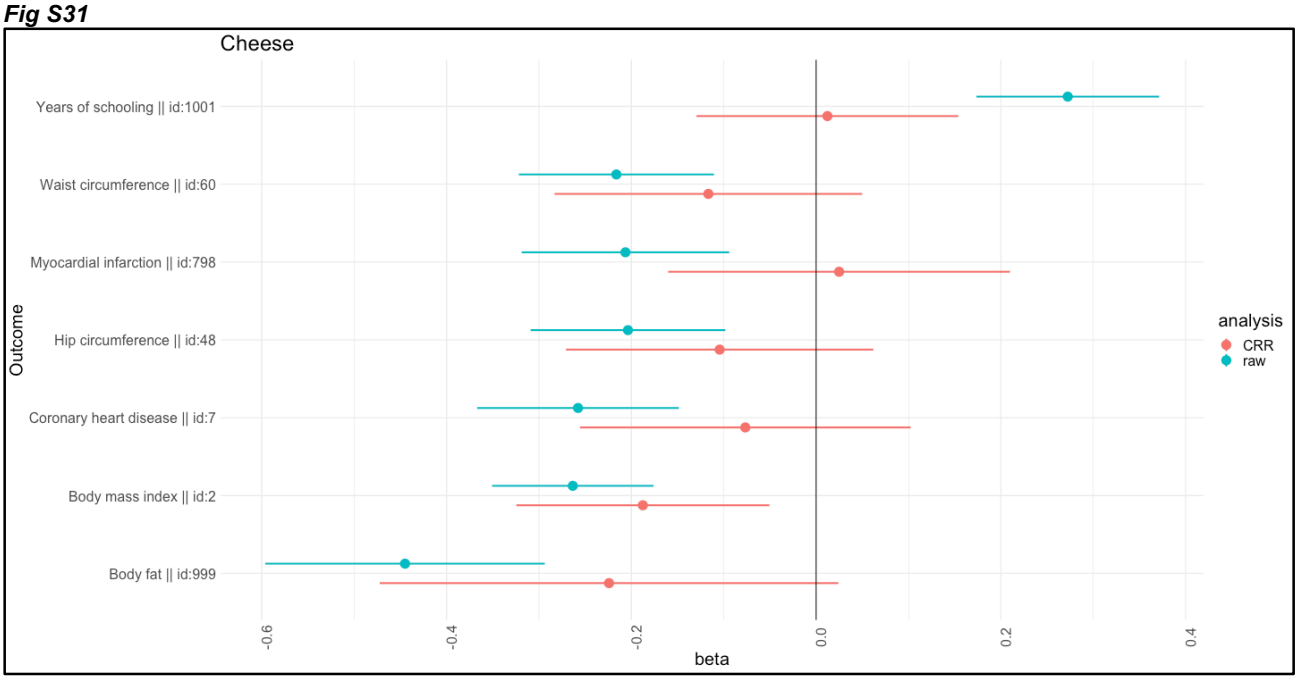


Fig S32

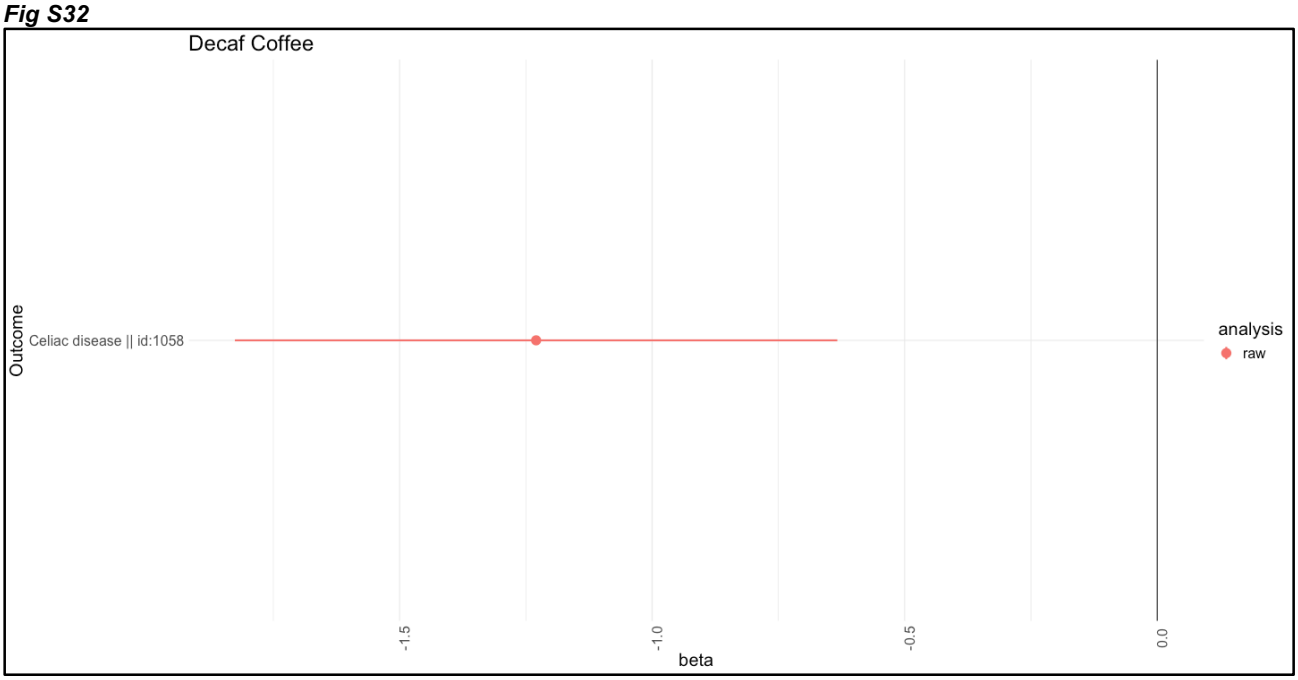


Fig S33

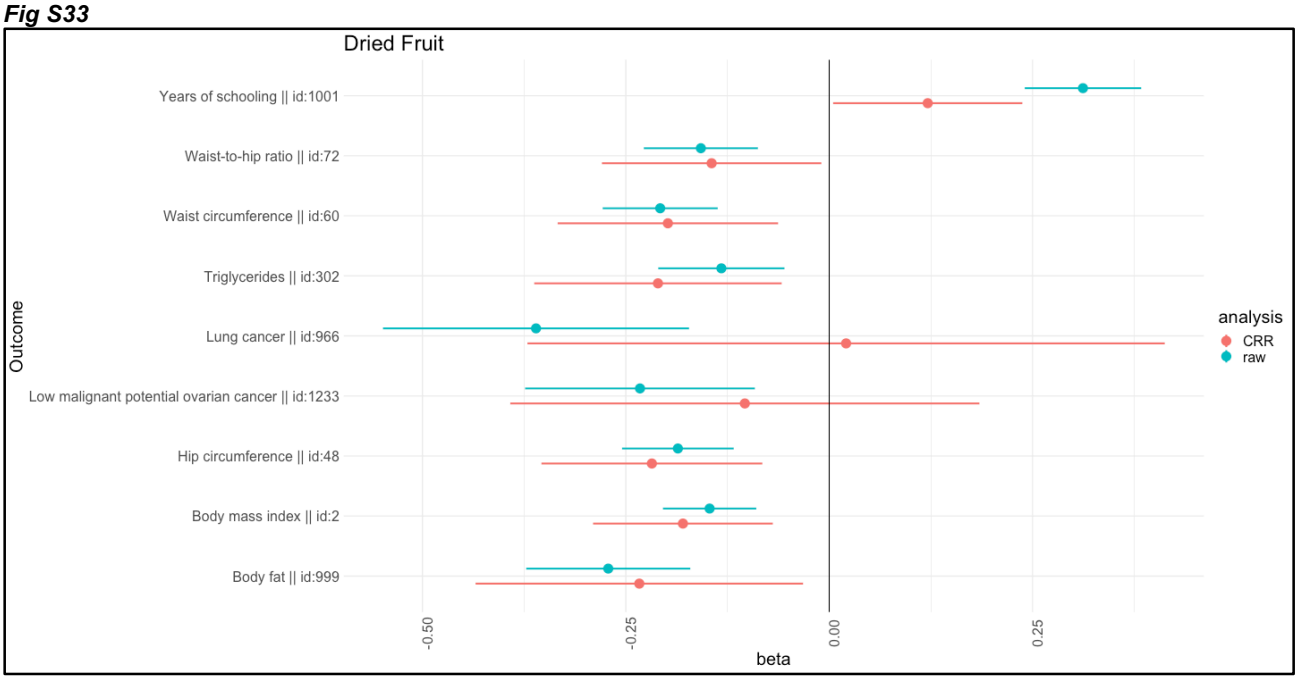


Fig S34

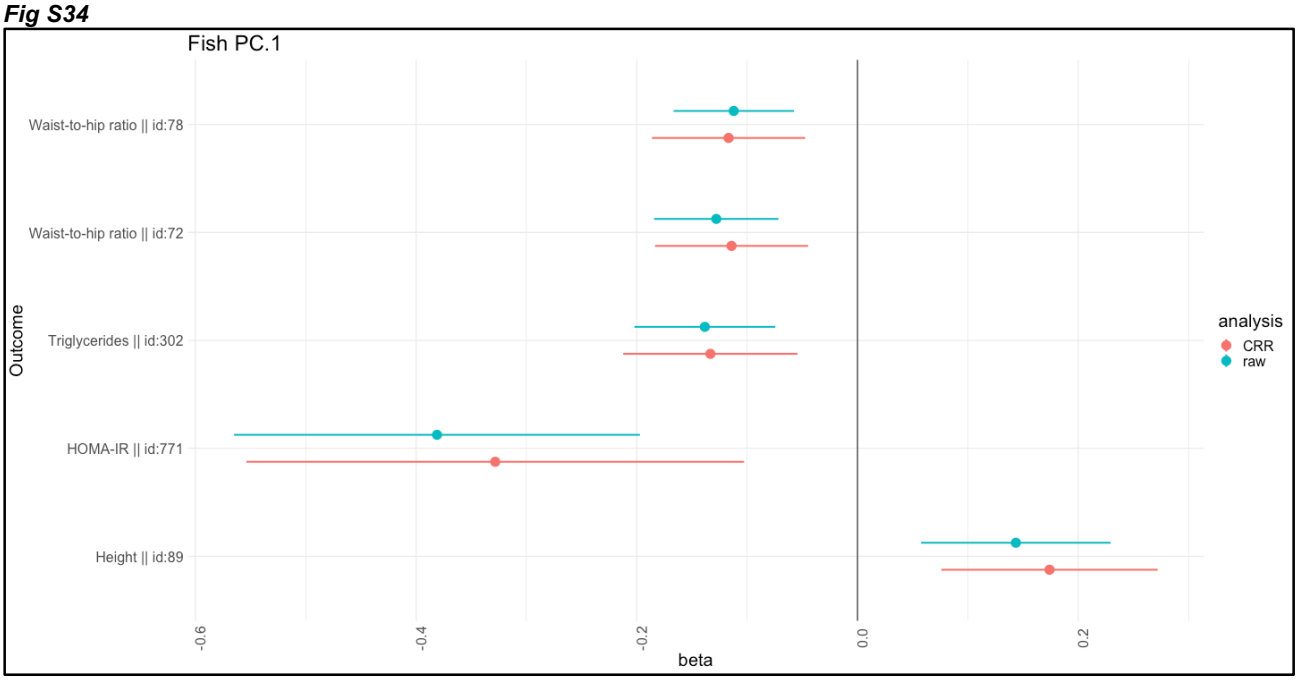


Fig S35

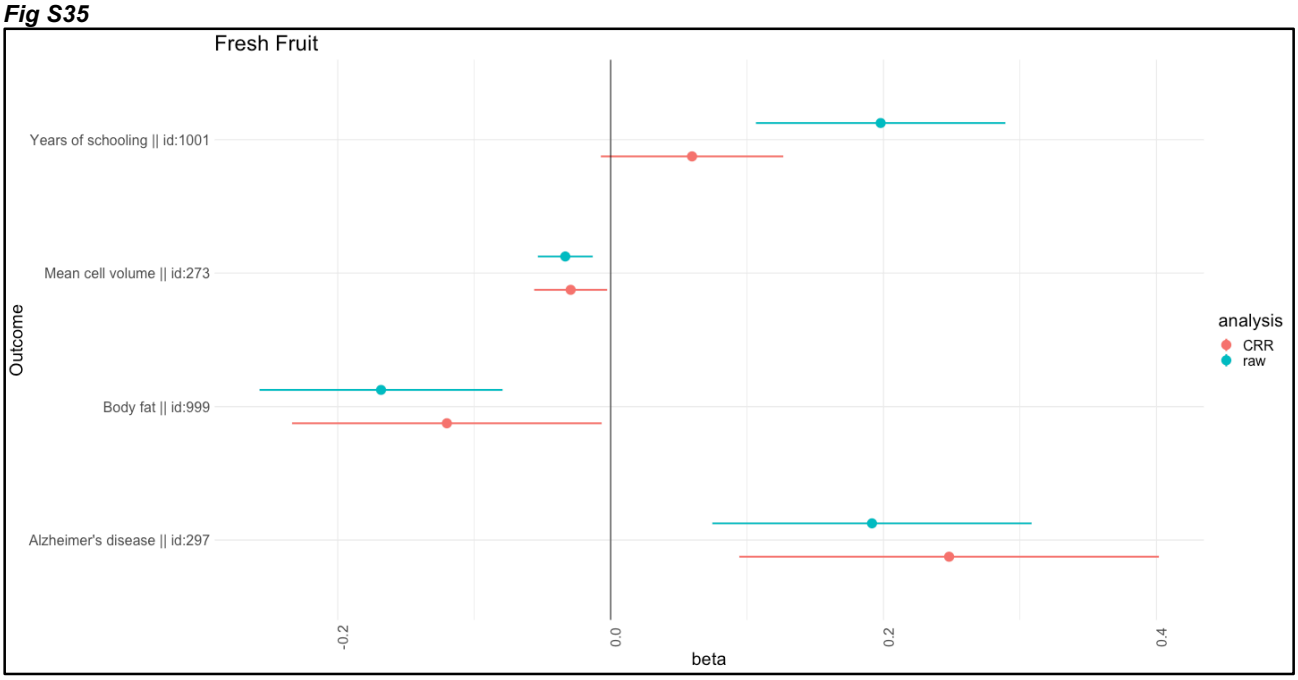


Fig S36

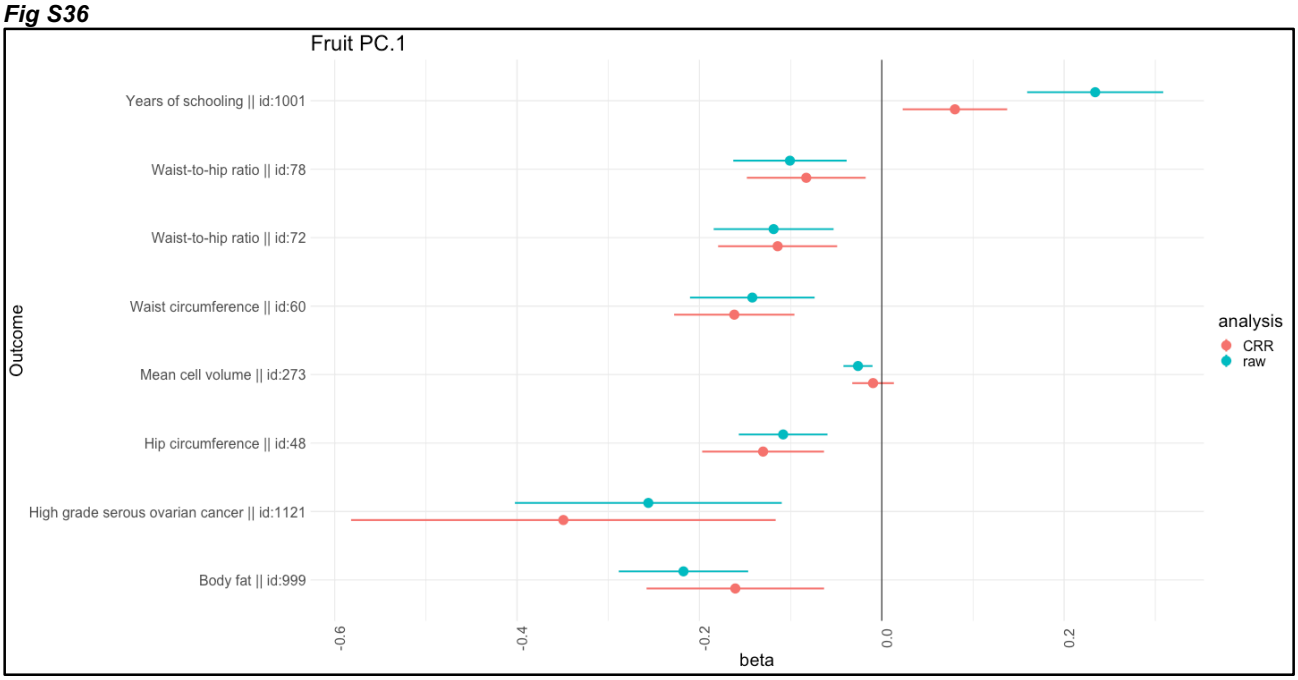


Fig S37

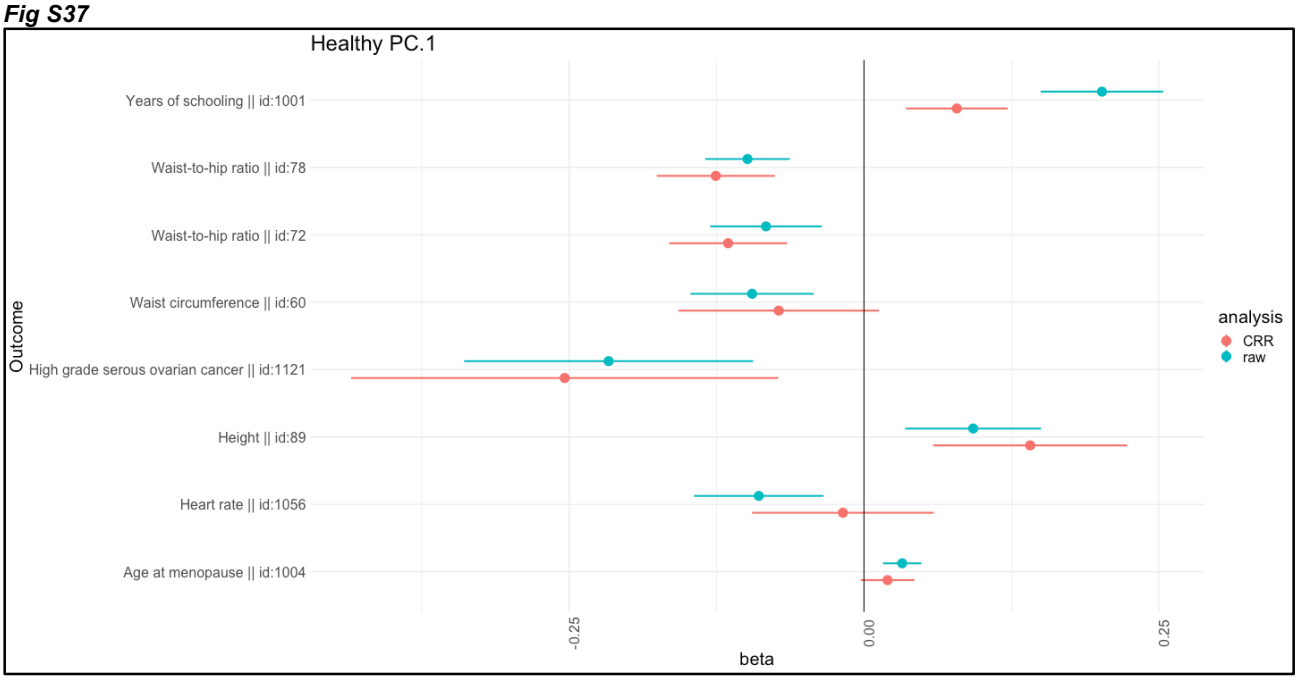


Fig S38

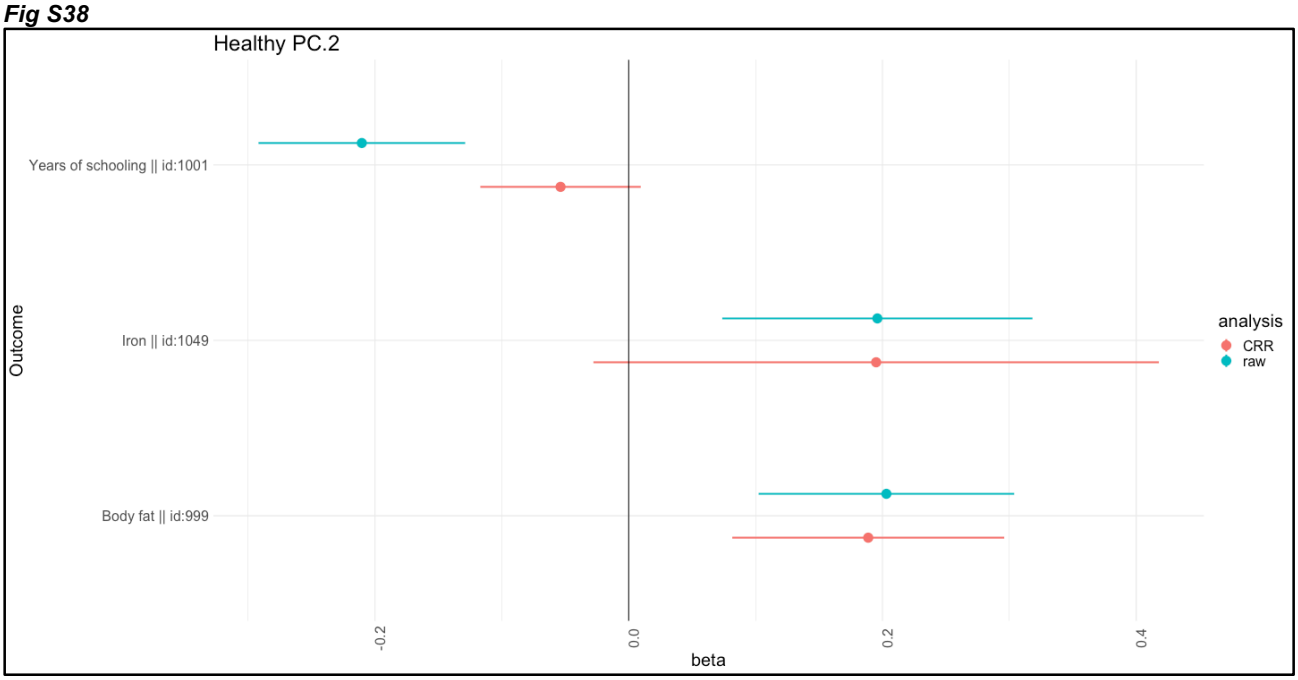


Fig S39

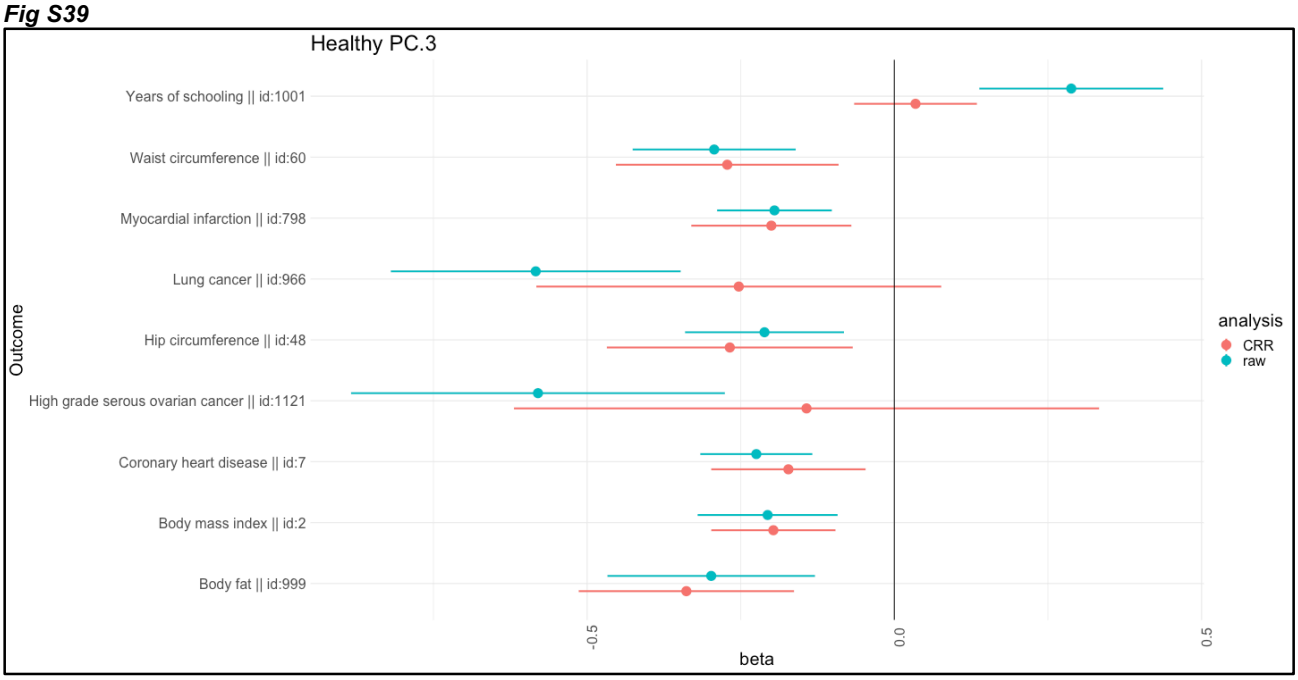


Fig S40

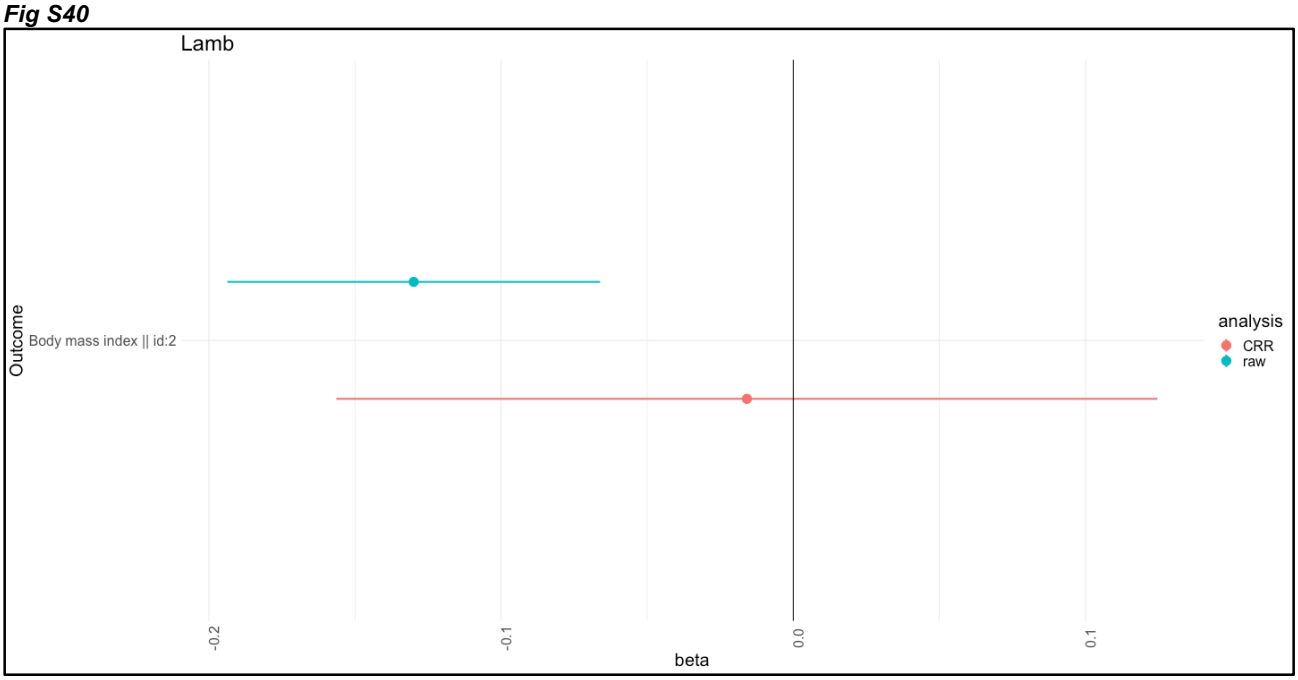


Fig S41

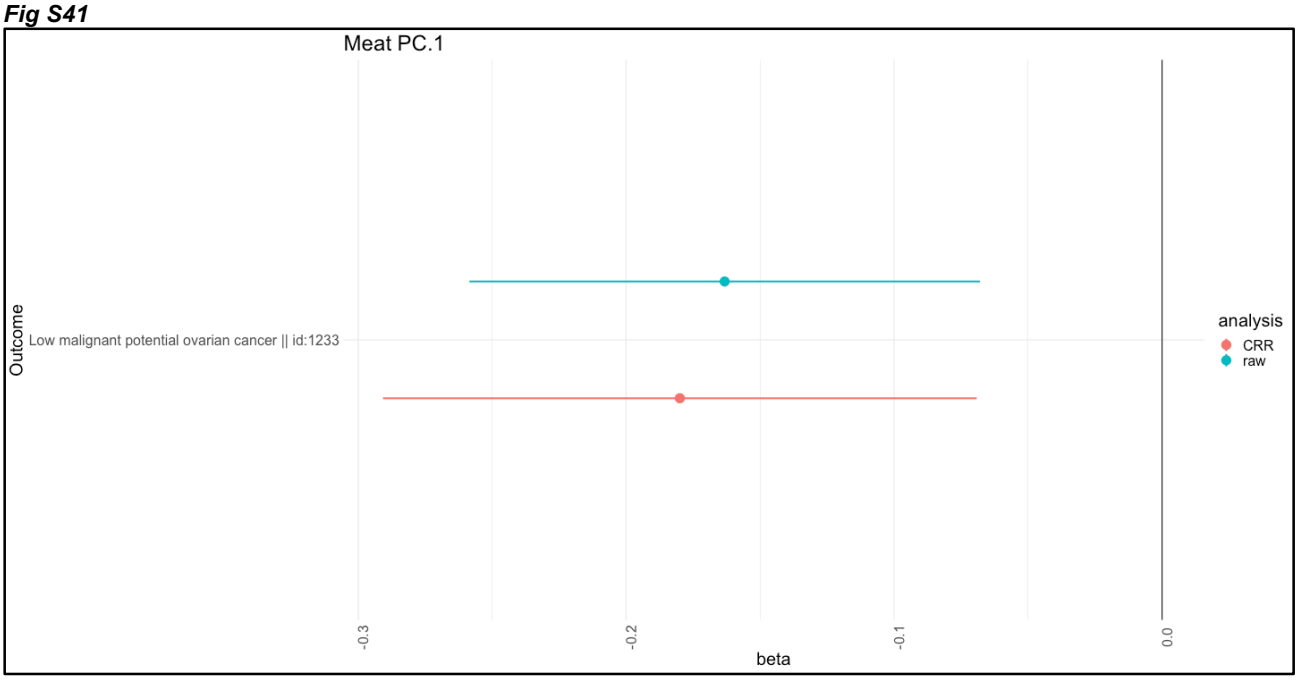


Fig S42

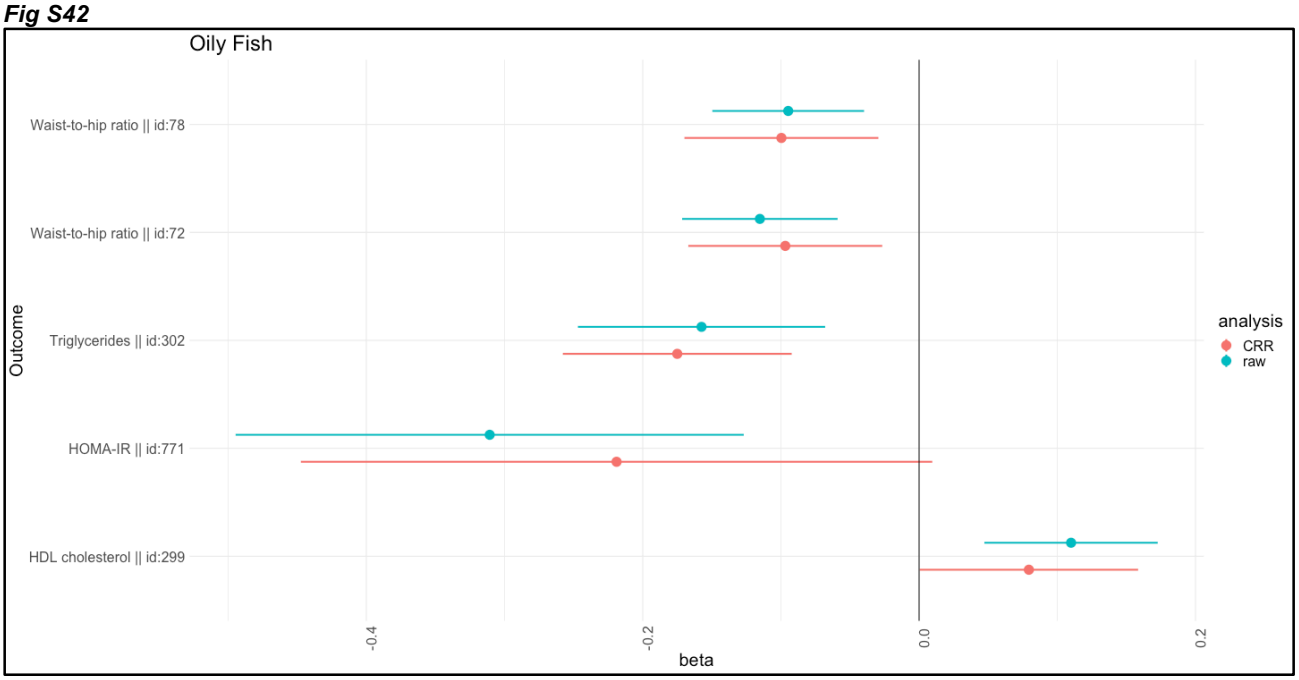


Fig S43

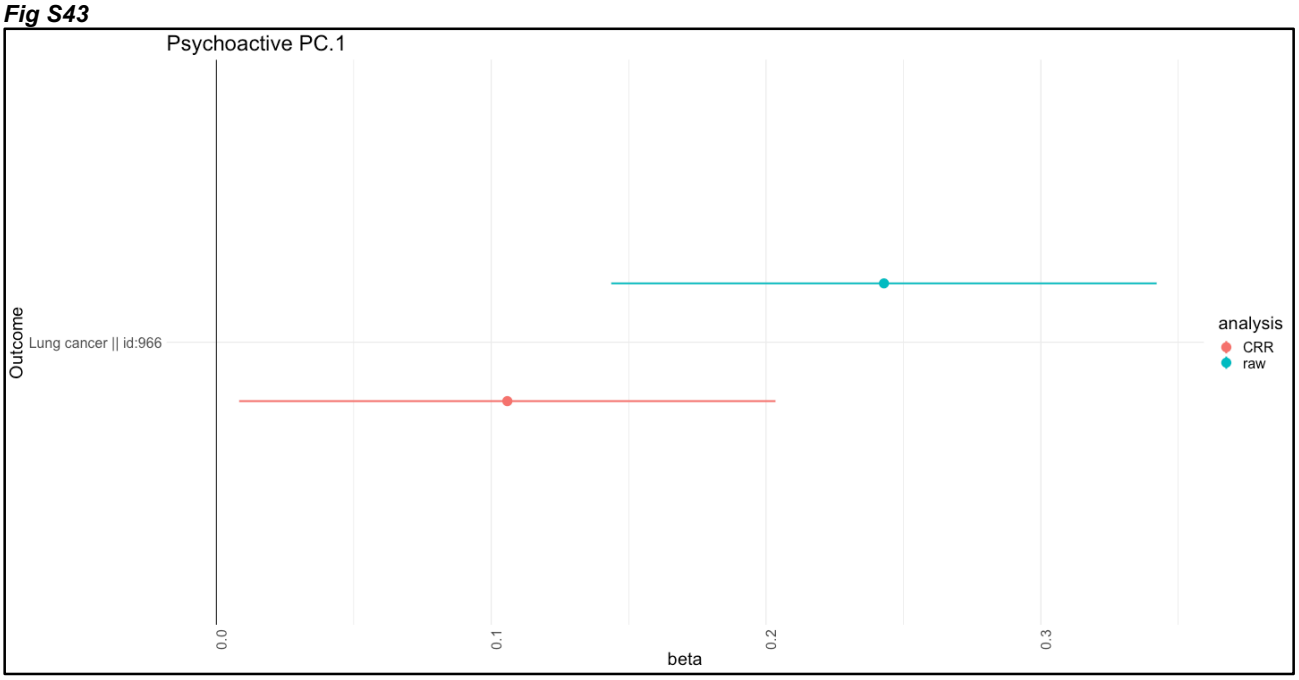


Fig S44

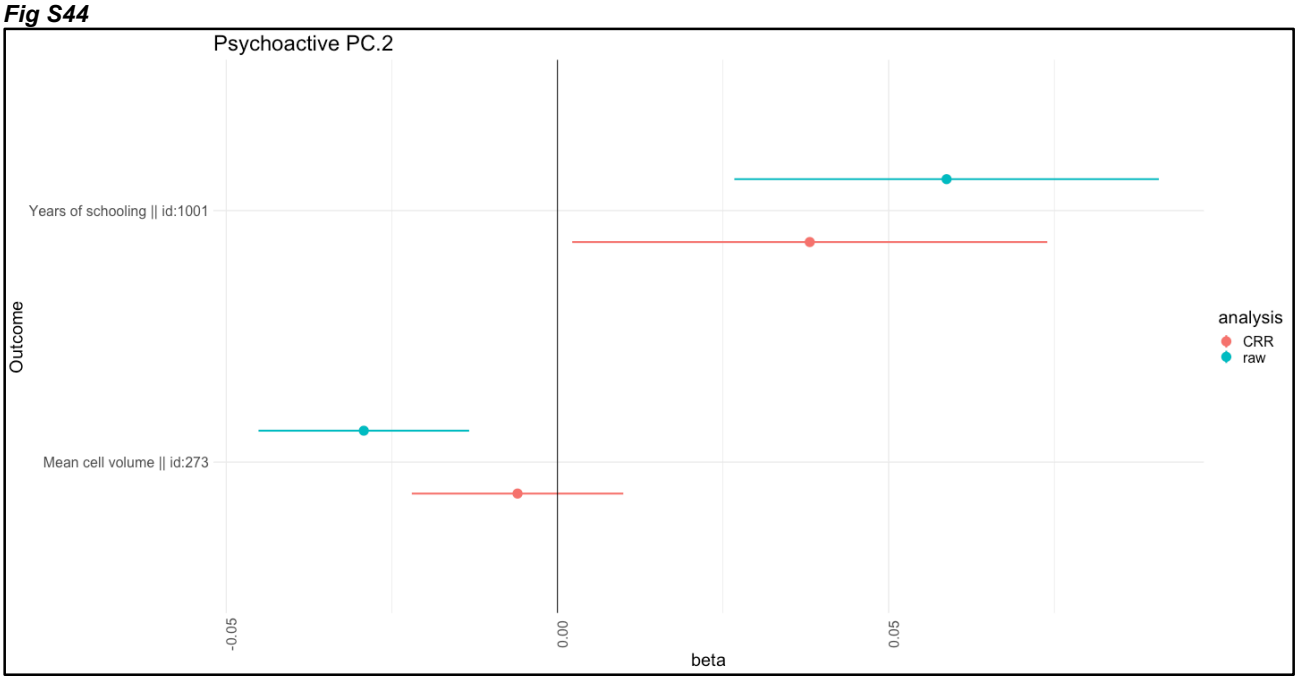


Fig S45

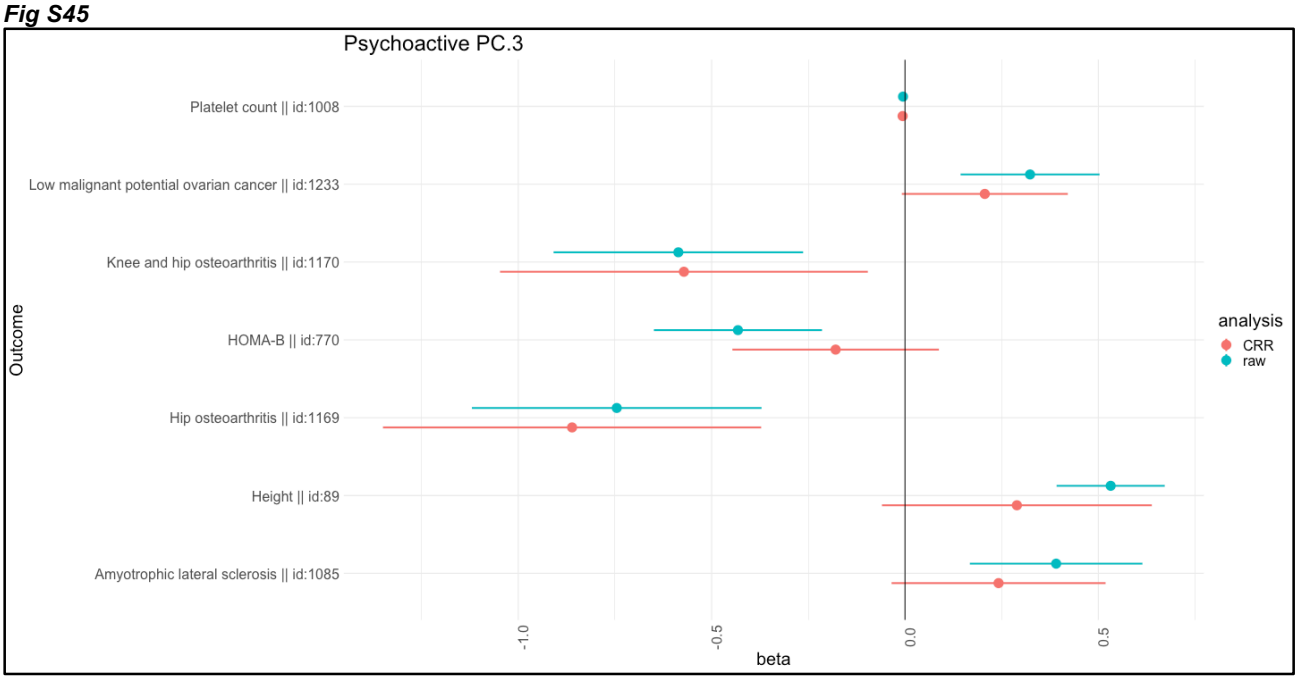


Fig S46

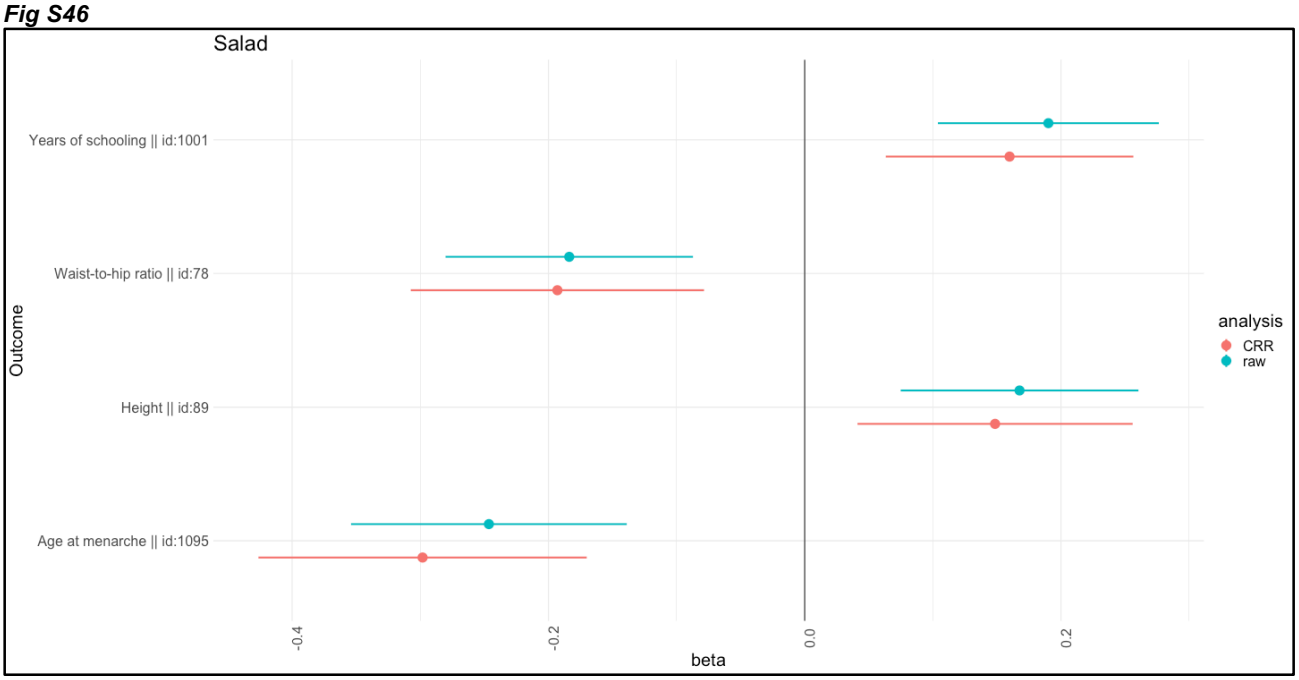


Fig S47

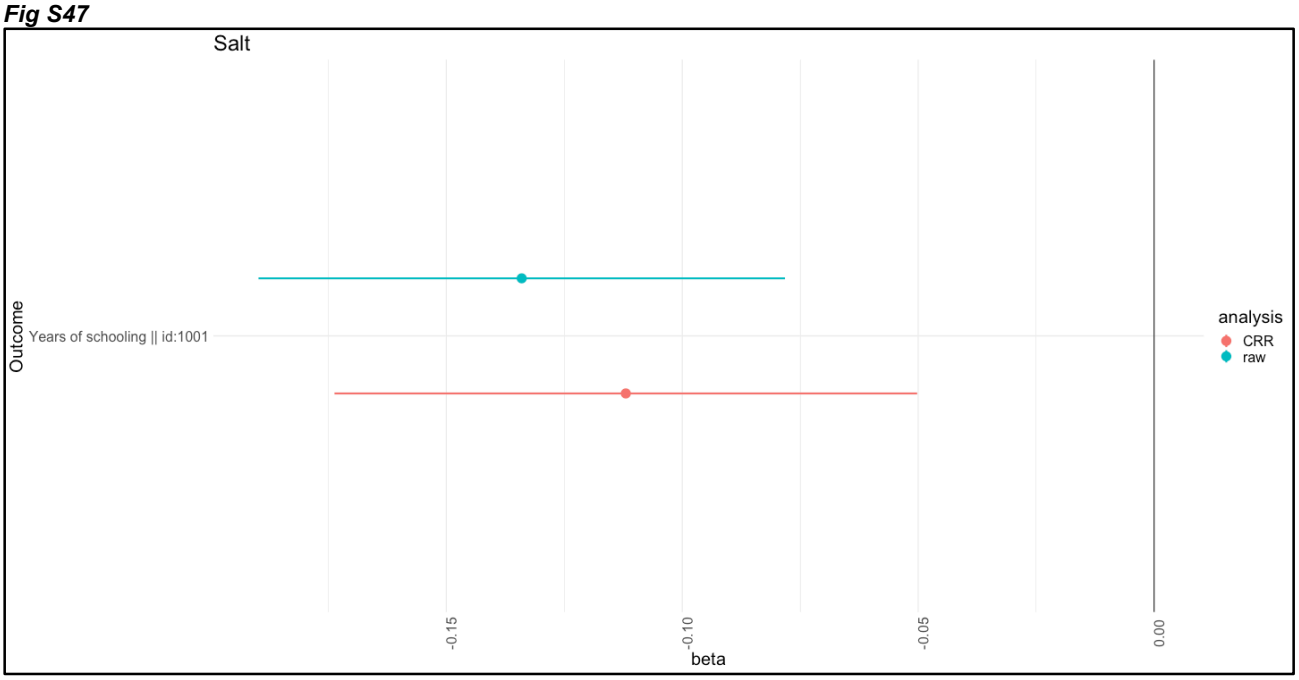


Fig S48

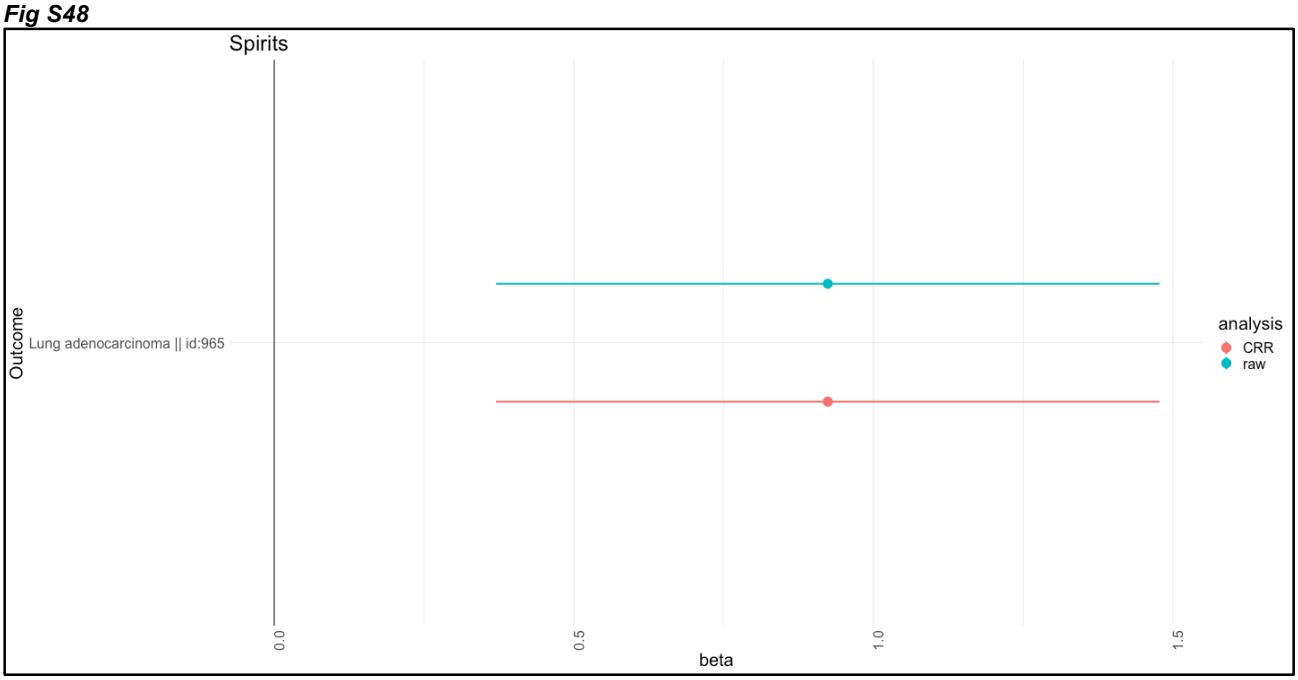


Fig S49

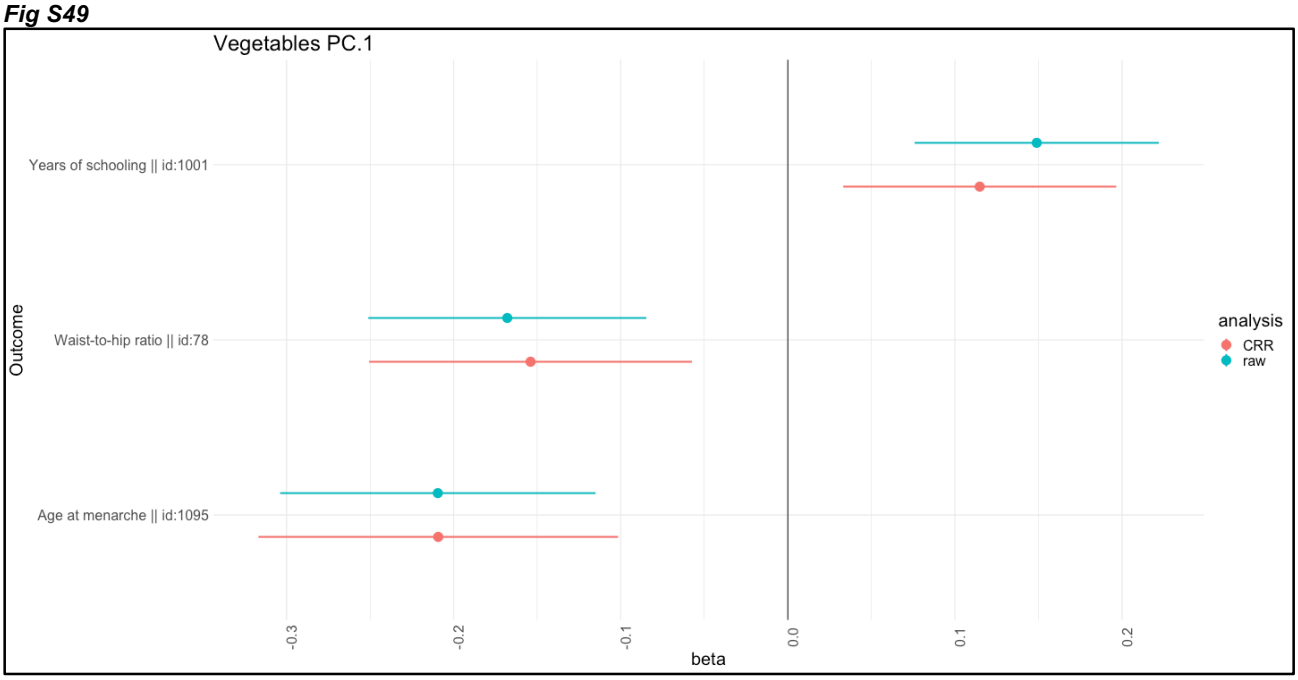
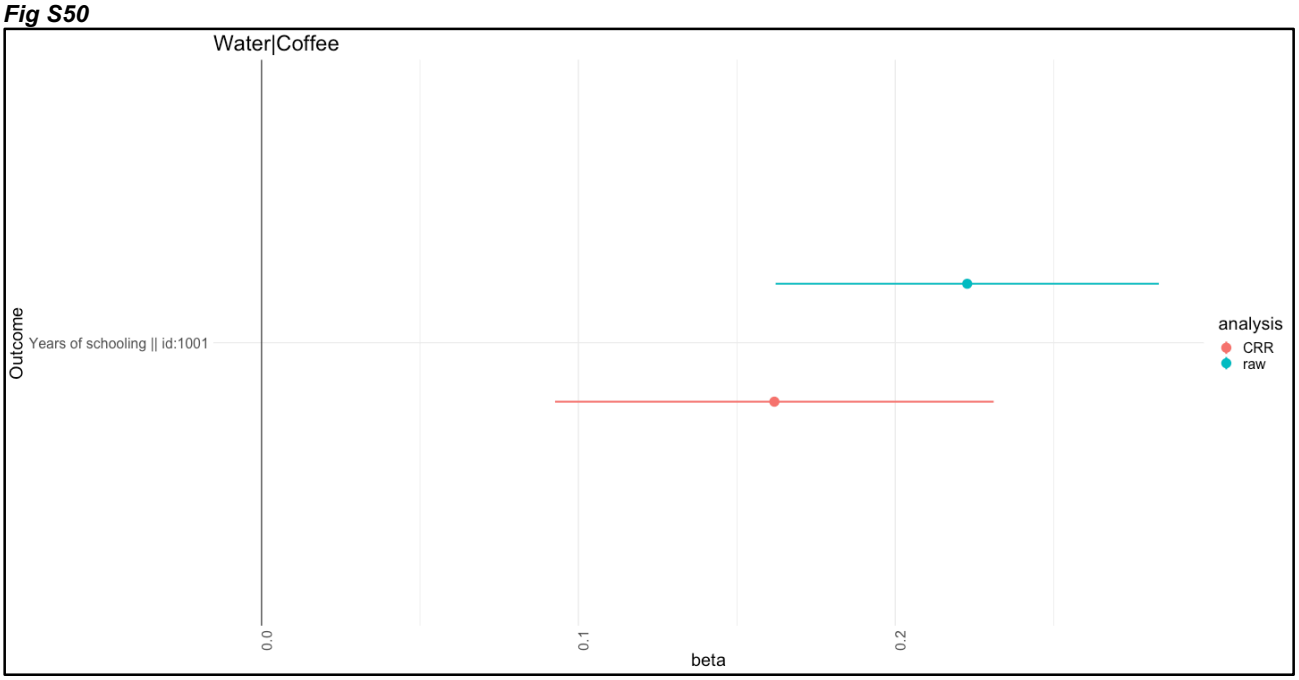


Fig S50



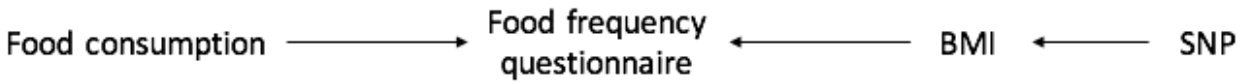
2.7 Comparison of the CRR filtering with Mendellian randomization common practice.

It is common practice in Mendellian randomization studies to determine the effect of horizontal pleiotropy through sensitivity analyses using methods such as the weighted median[26] or by evaluating the amount of heterogeneity in the effect estimates. In our case however we have shown that at least in some cases more than half of the variants can't be considered reliable. This means on one side that the assumption of the weighted median that at least half of the SNPs are valid IVs is violated for several of our food stuff and on the other that heterogeneity filtering (as in MR-Radial[27], or MR-PRESSO[28]) could potentially exclude the SNPs which are the true valid instruments. This is extremely problematic as it would lead to the wrong conclusions and results. We have in fact used the MR-Radial method in the MR pipeline and still obtained biased results. Another important point is that for the selection of the IVs it is in principle possible to use Steiger filtering[29] for understanding for each SNP if the causal pathway goes from the exposure to the outcome or the opposite. The test is based on evaluating how much variance of the exposure and the outcome is explained by the instrument, if the SNP explains more variance of the exposure than the outcome we can assume that the causal pathway goes from the exposure to the outcome, if the opposite is true, we can then infer that the causal pathway is going in the other direction.

Although the test is generally quite sturdy, in our case the extremely high noise present in the phenotype makes using Steiger's test for distinguishing the SNPs directly associated to the food trait from those associated through other traits less reliable.

To understand this issue let's imagine for simplicity that BMI is the only trait which is influencing causally either food consumption or food frequency questionnaire answers. As we cannot distinguish between the two cases not having an objective measure of consumption we can regard the two as being the same phenomenon, in fact if BMI influences how much we eat this will be transferred to the FFQ accordingly. Given we cannot distinguish between the two we need to evaluate any such effect as having a biasing effect on the FFQ response.

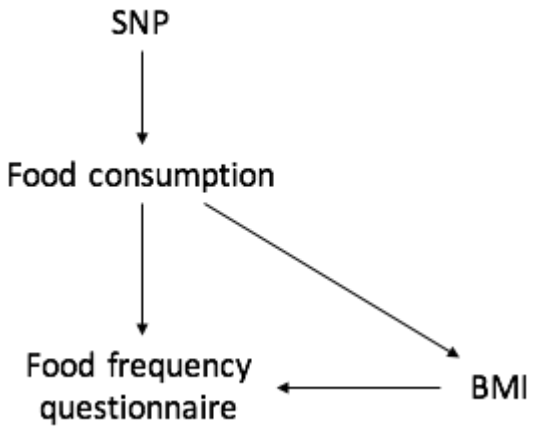
Adding a SNP which is causal to BMI the resulting DAG would look as such



In this case Steiger test will work as the variance explained by the SNP of BMI will always be greater than that explained on the FFQ as the effect is mediated through BMI, this is even truer if the effect is not directly on the FFQ but is mediated through FC.

So if we knew the factors biasing the FFQ and if they were independent from each other we could in principle use Steiger test to distinguish which are the SNPs which are mediated through these factors.

The problem arises when food consumption is causal to the confounding factor (as in many of the traits we used). For example, BMI is clearly caused by Food consumption, so after adding this information the DAG would look like.



In this case we can think of FFQ as an independent trait which is caused by FC. In this case things get more complicated as the r2 of the SNP on FFQ (which is what we would be using for the comparison) depends on the correlation of FC and FFQ. This can be very variable but it rarely over 0.5[30] with the exception of alcohol and coffee consumption which have higher reliability. Thus, it can be realistic that the causal effect of FC on BMI is similar in size to the direct causal effect of FC on FFQ. In this case, the Steiger test may fail to detect the correct direction of effect. While our approach has similarities to Steiger's test, its aims and settings are quite different in many aspects. First, in our case the causal direction is clear because it is highly unlikely that FFQ items cause other traits, only the upstream FC can do so. Second, our underlying DAG is more complicated (with multiple exposures and underlying FC) than it is assumed by the Steiger test and our approach fully exploits the a priori knowledge of the DAG. On the other hand, the Steiger test could be used to select valid (direct) exposure (e.g. BMI) instruments in order to estimate the total (direct plus indirect) exposure->FFQ causal effect, which in turn could be used to derive direct and indirect SNP-FFQ effects.

Bibiography

1. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
2. Loh, P.-R. *et al.* Efficient Bayesian mixed model analysis increases association power in large cohorts. doi:10.1101/007799
3. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
4. Day, N. *et al.* EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *Br. J. Cancer* **80 Suppl 1**, 95–103 (1999).
5. Lotta, L. A. *et al.* Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nat. Genet.* **49**, 17–26 (2017).
6. Bingham, S. A. *et al.* Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records. *Br. J. Nutr.* **72**, 619–643 (1994).
7. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, (2018).
8. McDaid, A. F. *et al.* Bayesian association scan reveals loci associated with human lifespan and linked biomarkers. *Nat. Commun.* **8**, 15842 (2017).
9. VanderWeele, T. J. A three-way decomposition of a total effect into direct, indirect, and interactive effects. *Epidemiology* **24**, 224–232 (2013).
10. Rüeger, S., McDaid, A. & Kutalik, Z. Evaluation and application of summary statistic imputation to discover new height-associated loci. *PLoS Genet.* **14**, e1007371 (2018).
11. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).
12. Finucane, H. K. *et al.* Partitioning heritability by functional category using GWAS summary statistics. doi:10.1101/014241
13. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set

797 analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
798 14. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and
799 annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
800 15. Szklarczyk, D. *et al.* STRING v10: protein–protein interaction networks, integrated over the
801 tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
802 16. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing
803 biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
804 17. Bowden, J. *et al.* Improving the visualization, interpretation and analysis of two-sample
805 summary data Mendelian randomization via the Radial plot and Radial regression. *Int. J.*
806 *Epidemiol.* **47**, 2100 (2018).
807 18. Paul, D. R., Rhodes, D. G., Kramer, M., Baer, D. J. & Rumpler, W. V. Validation of a food
808 frequency questionnaire by direct measurement of habitual ad libitum food intake. *Am. J.*
809 *Epidemiol.* **162**, 806–814 (2005).
810 19. Dehghan, M. *et al.* Association of dairy intake with cardiovascular disease and mortality in 21
811 countries from five continents (PURE): a prospective cohort study. *Lancet* **392**, 2288–2297
812 (2018).
813 20. Adam, T. C. & Epel, E. S. Stress, eating and the reward system. *Physiology & Behavior* **91**,
814 449–458 (2007).
815 21. Consortium, G. & GTEx Consortium. Genetic effects on gene expression across human
816 tissues. *Nature* **550**, 204–213 (2017).
817 22. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies
818 disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
819 23. Bernstein, B. E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.*
820 **28**, 1045–1048 (2010).
821 24. Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased
822 coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic*
823 *Acids Res.* **47**, D607–D613 (2019).
824 25. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in

large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P10008 (2008).

26. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
27. Bowden, J. *et al.* Improving the visualisation, interpretation and analysis of two-sample summary data Mendelian randomization via the radial plot and radial regression. doi:10.1101/200378
28. Verbanck, M., Chen, C.-Y., Neale, B. & Do, R. Publisher Correction: Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 1196 (2018).
29. Hemani, G., Tilling, K. & Davey Smith, G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PLoS Genet.* **13**, e1007081 (2017).
30. Kristal, A. R., Peters, U. & Potter, J. D. Is it time to abandon the food frequency questionnaire? *Cancer Epidemiol. Biomarkers Prev.* **14**, 2826–2828 (2005).