**Two-sample Mendelian randomization using summary genetic data – MODEL ANSWER**

**ANALYSIS 1: Two-sample Mendelian randomization of one exposure on one outcome**

# . Do you think BMI causes CHD?

**YES**

# a. What is the odds ratio for coronary heart disease per unit increase in genetically elevated BMI?

The odds ratio for coronary heart disease per SD higher BMI due to genetic variation was 1.56 (logOR= 0.446) (95% confidence interval: 1.39 to 1.75) (estimated by the IVW method).

# b. Is there evidence for pleiotropy or violations of MR assumptions?

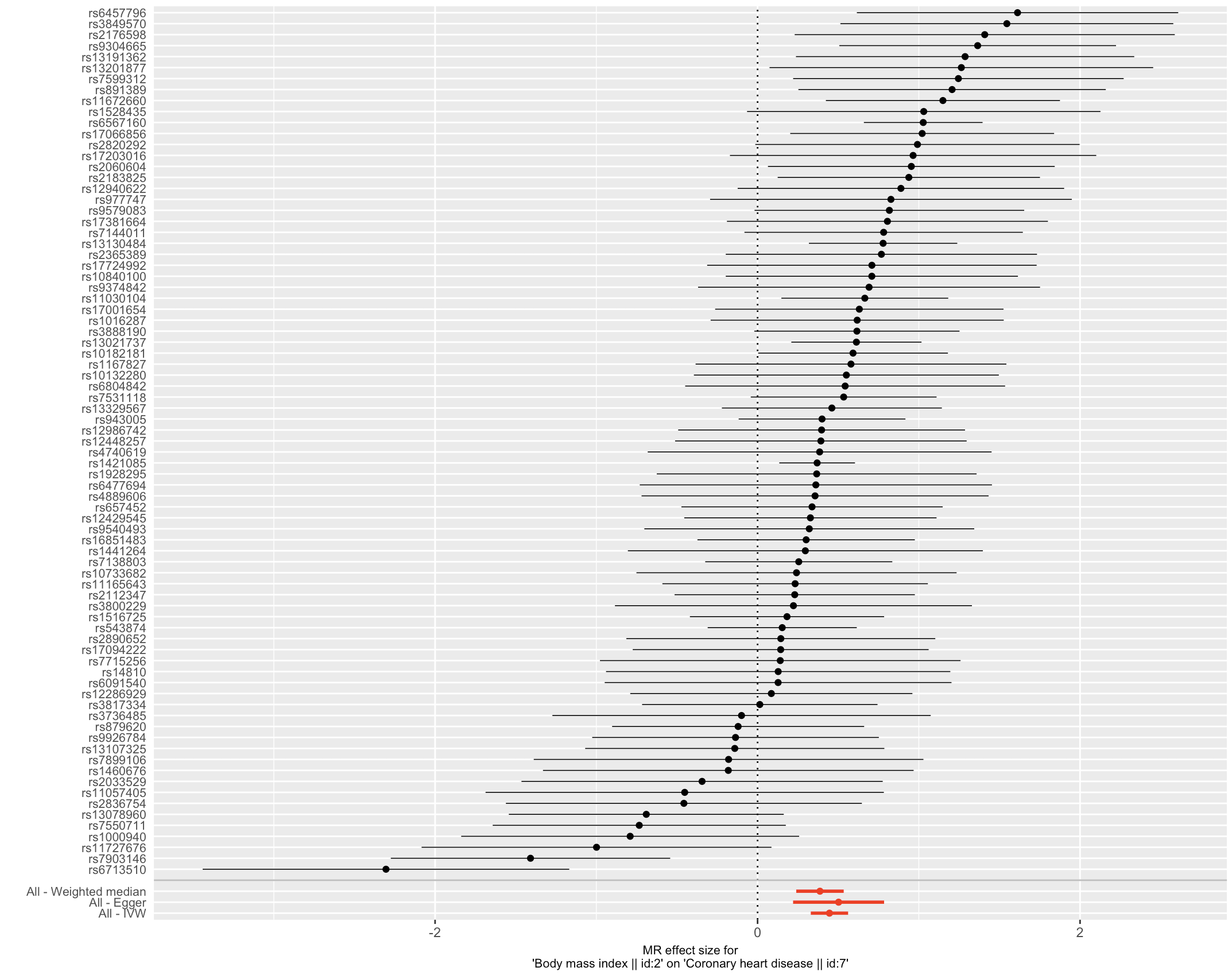
No strong evidence.

1. All MR analysis provide similar causal estimates.
2. The MR-Egger intercept test did not indicate strong evidence for unbalanced pleiotropy (P=0.66).

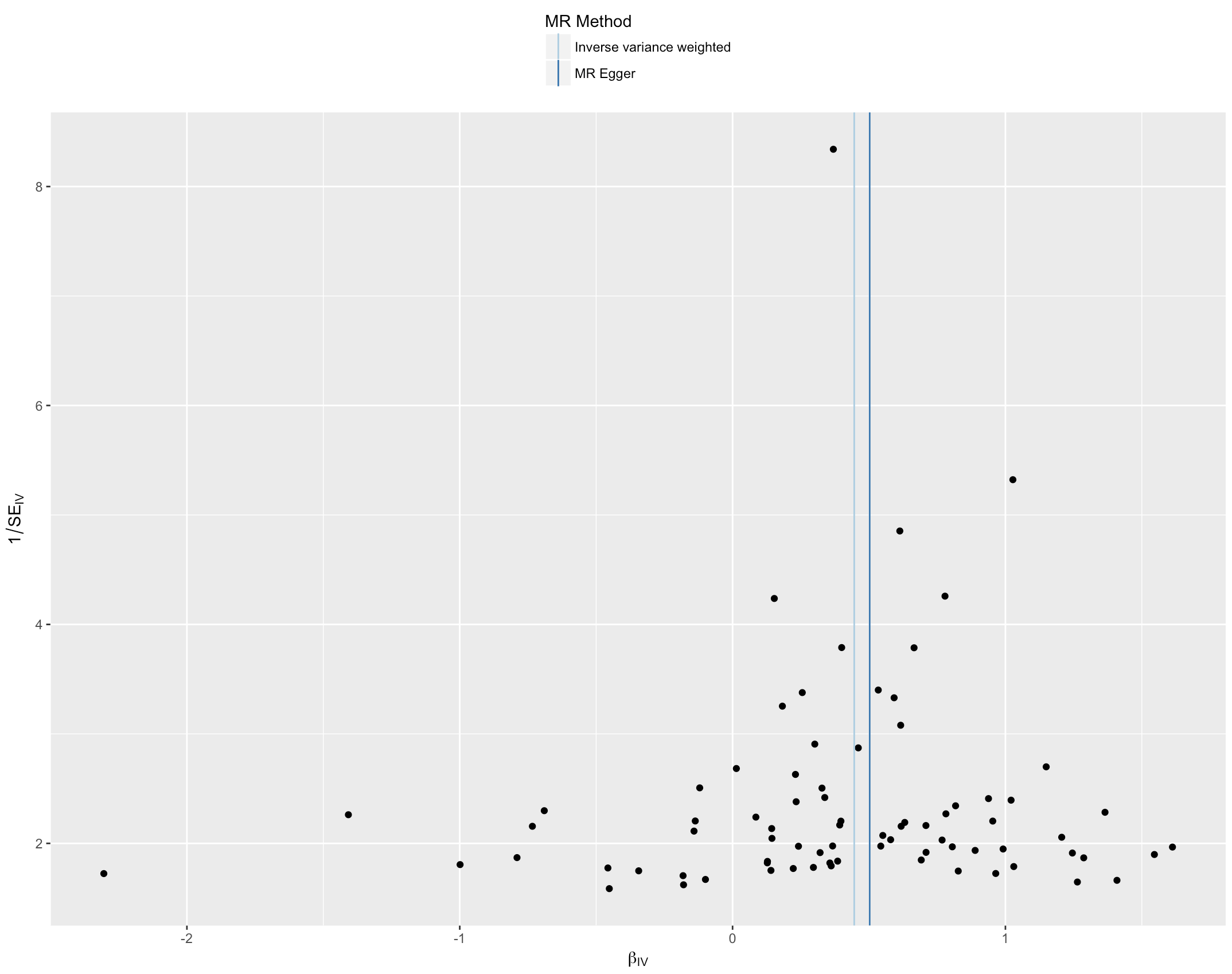
***Graphical tests***

It is also possible to visualise potential violations of MR assumptions graphically through forest plots, funnel plots and scatter plots (figures 1-3). The funnel plot looks reasonably symmetric, suggesting that any pleiotropy that is present is “balanced” and is not biasing the overall MR results consistently in the negative or positive direction. In other words, given their effects on BMI, there are as many SNPs with unusually strong protective effects on CHD as there are SNPs with unusually strong harmful effects on CHD.

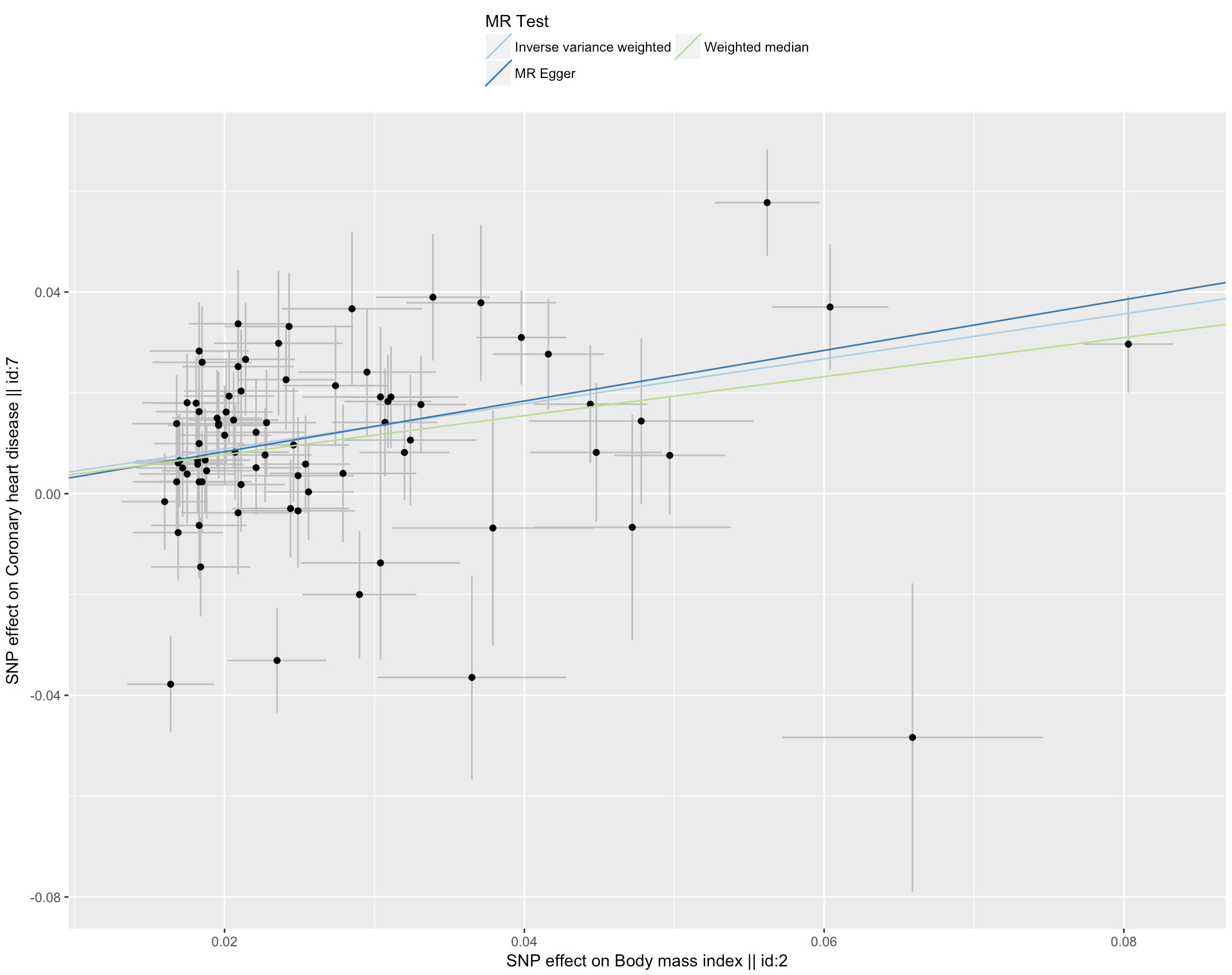
**Figure 1**. Forest plot of Mendelian randomization results. Plot shows effect of body mass index on coronary heart disease due to each SNP separately (estimated by ratio method) as well as combined across all SNPs into an overall effect (estimated by various methods). Heterogeneity could be indicative of violations of MR assumptions. Notice how some SNPs at the bottom and top of the plots have unusually strong protective or harmful effects, in comparison to the bulk of the SNPs. This could be indicative of alternative pathways between the SNPs and CHD not mediated by BMI (horizontal pleiotropy). Alternatively, these “outliers” could reflect other violations of assumptions, data handling errors or chance sampling variation (see explanatory text above).



**Figure 2**. Funnel plot of MR results. The plot shows ratio estimates of causal effect (x axis) plotted against the inverse of the standard error of the ratio estimates (Y axis). We generally expect the distribution to look symmetric, so that as estimates get less precise (going from top to bottom on the Y axis), the causal estimates “fan” out randomly on either side of the overall effect (indicated by the vertical lines). Asymmetry (i.e an imbalance of estimates on one side of the overall effect) is indicative of small study bias in a meta-analysis context. In Mendelian randomization, asymmetry is suggestive of unbalanced pleiotropy.



**Figure 3.** Scatter plot of results for BMI SNPs. The plot shows the SNP-CHD effects (Y axis) plotted against the SNP-BMI effects (X axis), where effect refers to the log odds for CHD or SD change in BMI per copy of the effect allele. If BMI causes CHD we would expect a linear dose response relationship, i.e. the SNP-CHD effects should increase as the SNP-BMI effects increase. Notice that there are a few “outlier” SNPs. Some BMI-raising alleles seem to be associated with negative log odds for CHD. Some BMI-raising alleles also seem to have unusually strong, risk raising, effects on CHD, in comparison to the other SNPs. “Outlier” SNPs in a scatter plot will generally correspond to the SNPs at the top or bottom of a forest plot of SNP ratio estimates (figure 1). Could these “outliers” be indicative of pleiotropic pathways to CHD? Possibly. However, there may be other explanations for outliers (see answer on heterogeneity above).



**ANALYSIS 2: Two-sample Mendelian randomization of many exposures on one outcome**

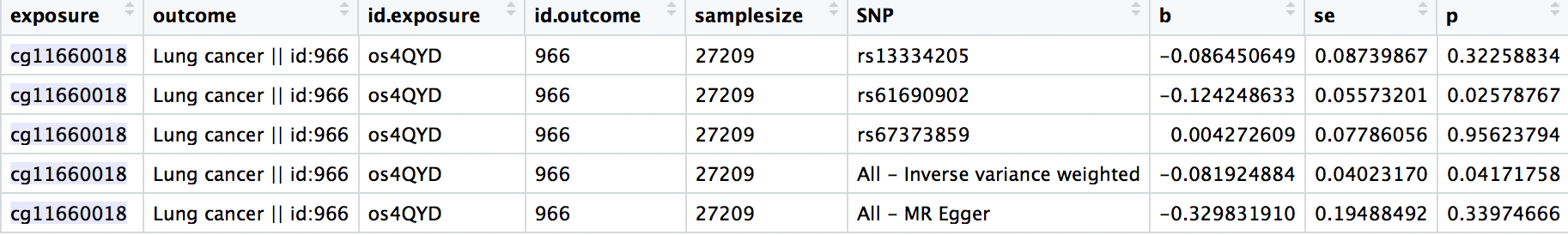
# . Do you think smoking CpGs causes lung cancer?

One CpG is strongly associated with lung cancer (cg23771366: one instrument), two CpG are potentially associated with lung cancer (cg11660018: three instruments and cg03636183: two instruments).

# a. What is the odds ratio for lung cancer per unit increase in genetically elevated cg11660018?

The odds ratio for coronary heart disease per SD higher BMI due to genetic variation was 0.921 (95% confidence interval: 0.851 to 0.996) (estimated by the IVW method).

# b. What is the Wald ratio of each SNP for this association?



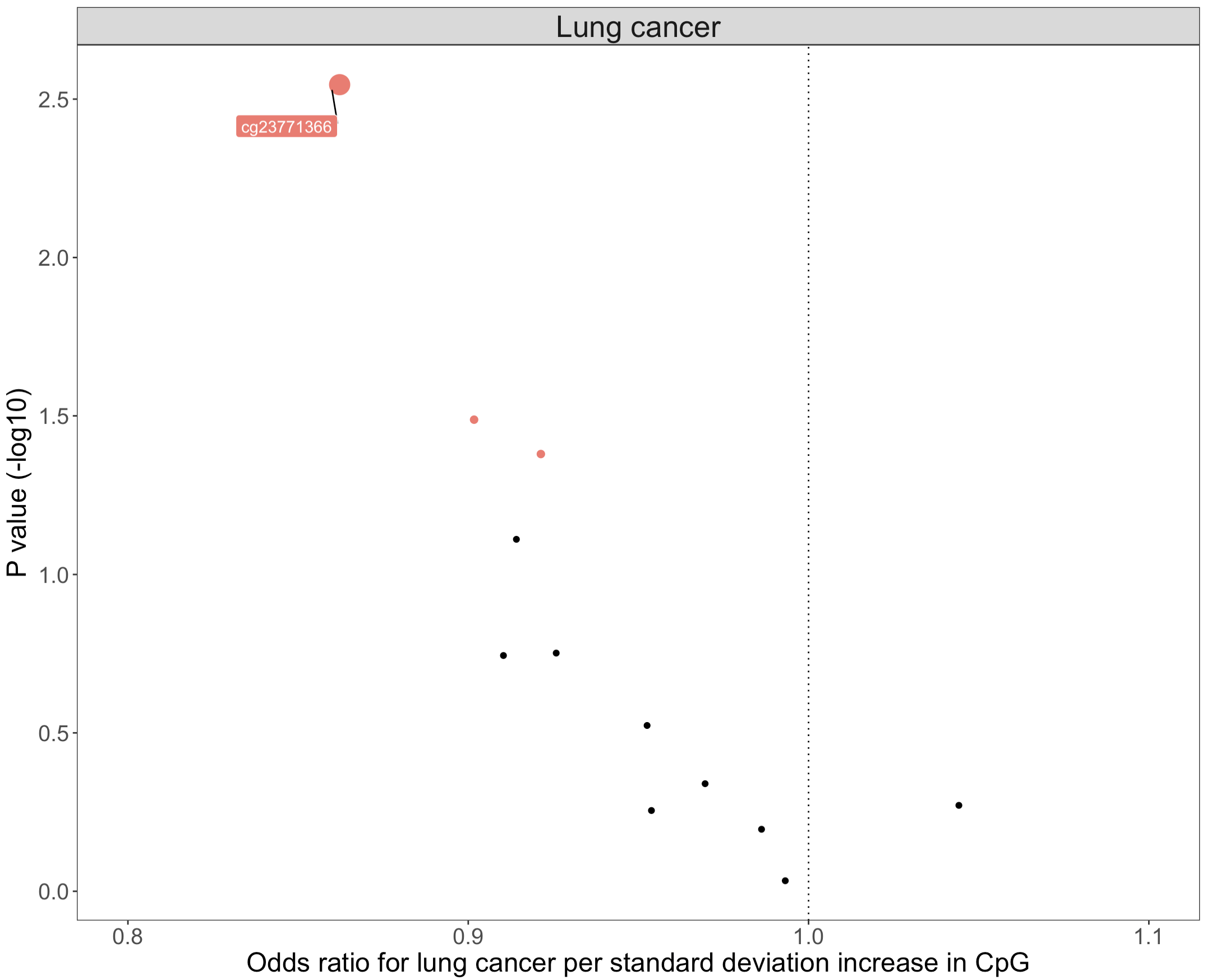
# c. Does the MR Egger / Weighted median results reliable?

Since we have limited number of SNPs as instruments, so MR Egger and Weighted median analysis can hardly provide reliable results.

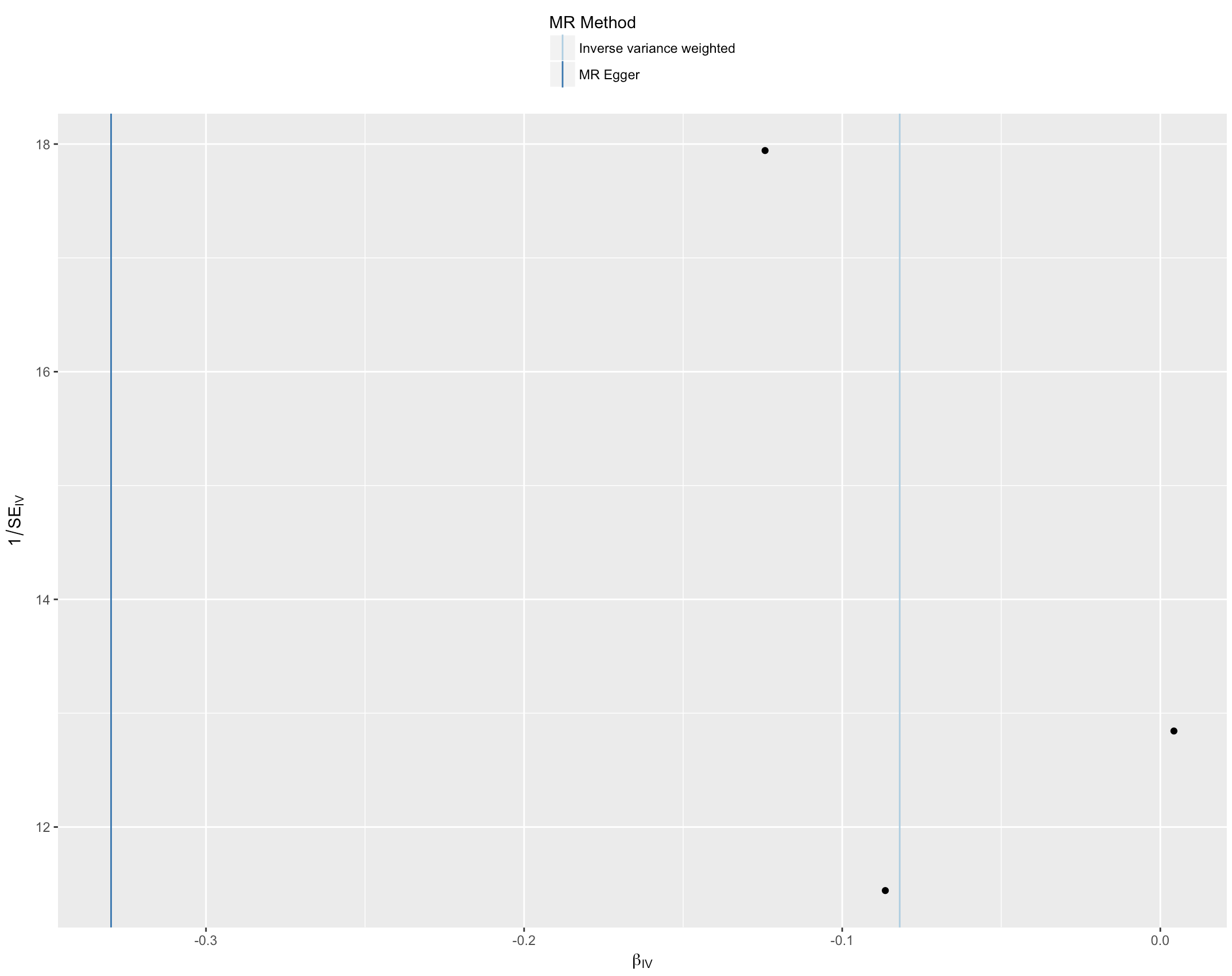
***Graphical tests***

It is also possible to visualise potential violations of MR assumptions graphically through forest plots, funnel plots and scatter plots (figures 1-3). The scatter and funnel plot only have three points, suggesting that the sensitivity analysis such as MR Egger and Weighted median is not reliable.

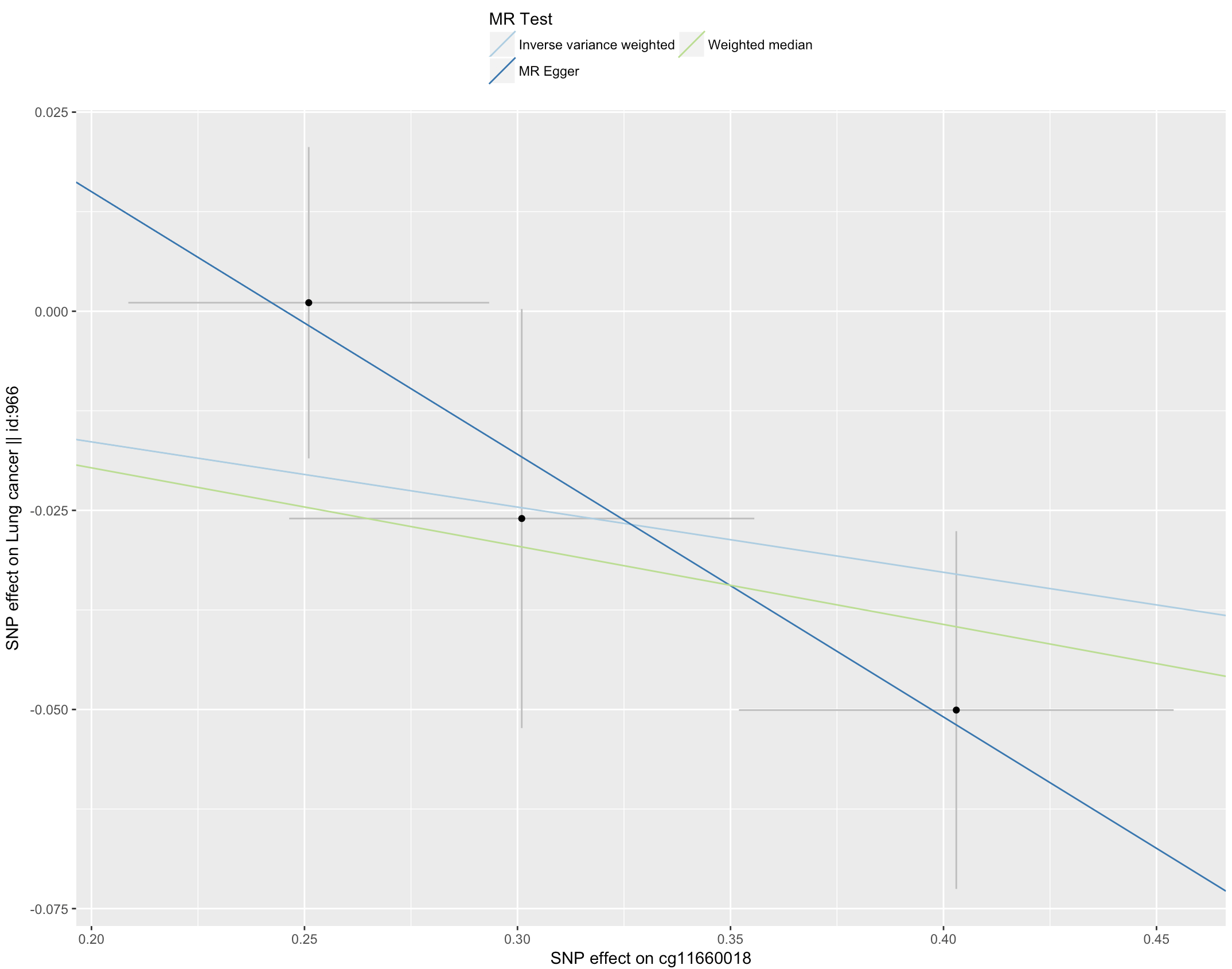
**Figure 1**. Volcano plot of Mendelian randomization results. The plot shows the CpG-lung cancer effects (X axis) plotted against the P value of the association (Y axis), where effect refers to the log odds for lung cancer per SD increase in methylation level of a CpG. Each point refers to one CpG site. The size of the point referring the level of significance.



**Figure 2**. Funnel plot of MR results for cg11660018 on lung cancer. The plot shows ratio estimates of causal effect (x axis) plotted against the inverse of the standard error of the ratio estimates (Y axis). We generally expect the distribution to look symmetric, so that as estimates get less precise (going from top to bottom on the Y axis), the causal estimates “fan” out randomly on either side of the overall effect (indicated by the vertical lines). But in this case, we only have three points, so it is hard to check the MR assumption using the funnel plot.



**Figure 3.** Scatter plot of the MR results for cg11660018 on lung cancer. The plot shows the SNP-lung cancer effects (Y axis) plotted against the SNP-CpG effects (X axis), where effect refers to the log odds for lung cancer or SD change in CpG per copy of the effect allele. The three points fit a linear line well but we are not able to claim a causality directly. We can further test whether these instruments share the same causal variant with lung cancer using colocalization analysis.



**Further reading**

Zheng J. et al. Recent development of Mendelian randomization. Curr Epidemiol Rep. 2017;4(4):330-345.

<https://www.ncbi.nlm.nih.gov/pubmed/29226067>

**Useful links for MR-Base**

1) Detailed documentation on how to use MR-Base R package

[**https://mrcieu.github.io/TwoSampleMR/**](https://mrcieu.github.io/TwoSampleMR/)

2) Github repo for R package (if you want to see on the TwoSampleMR R package works)

<https://github.com/MRCIEU/TwoSampleMR>

3) Web interface

www.mrbase.org/

4) MR-Base methods paper

https://elifesciences.org/articles/34408