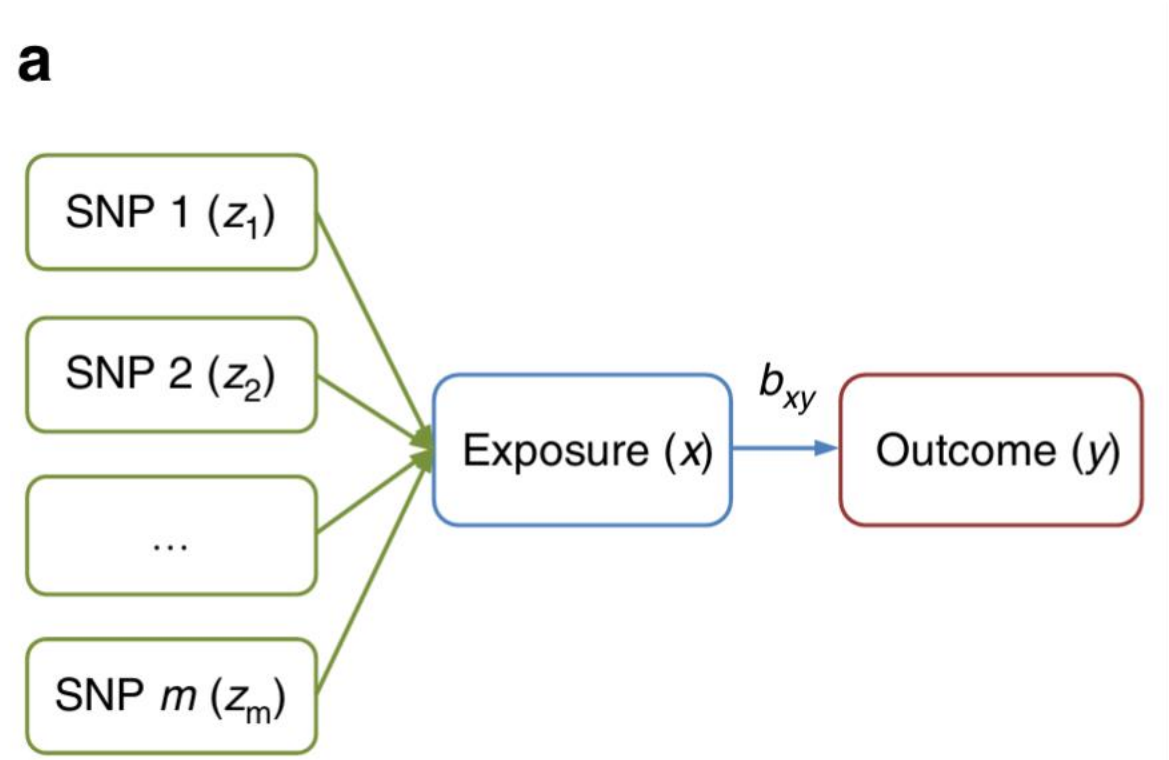
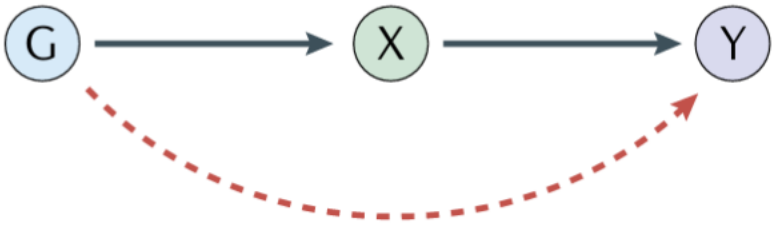
Visualization Inspiration

Isabel Elliott

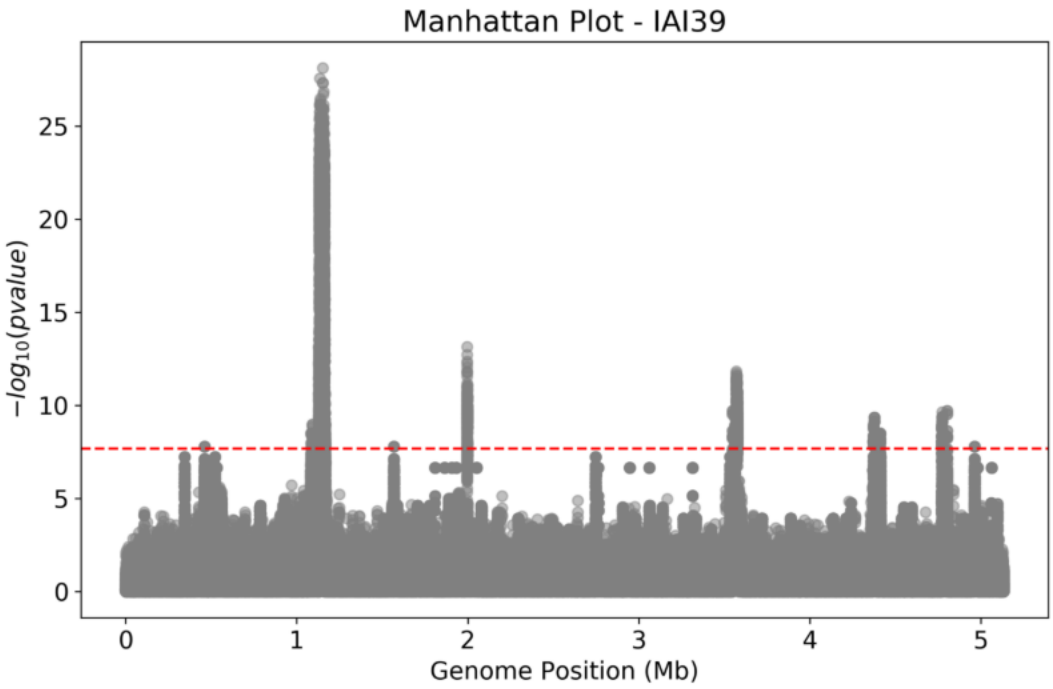
This document briefly covers and explains my thoughts regarding any graphs and/or visuals in the paper. It is messy and slightly confusing, so feel free to message me with any questions.

- Intro to MR diagram (example in ‘Causal associations between risk factors and common diseases inferred from GWAS summary data’) (simplified example in 'Mendelian Randomization primer’). This ensures readers will have a good understanding of the principles undlying MR before diving into GSMR.





- Manhattan plot for each rare CVD showcasing the most significant SNPs from GWAS results. (example from online)

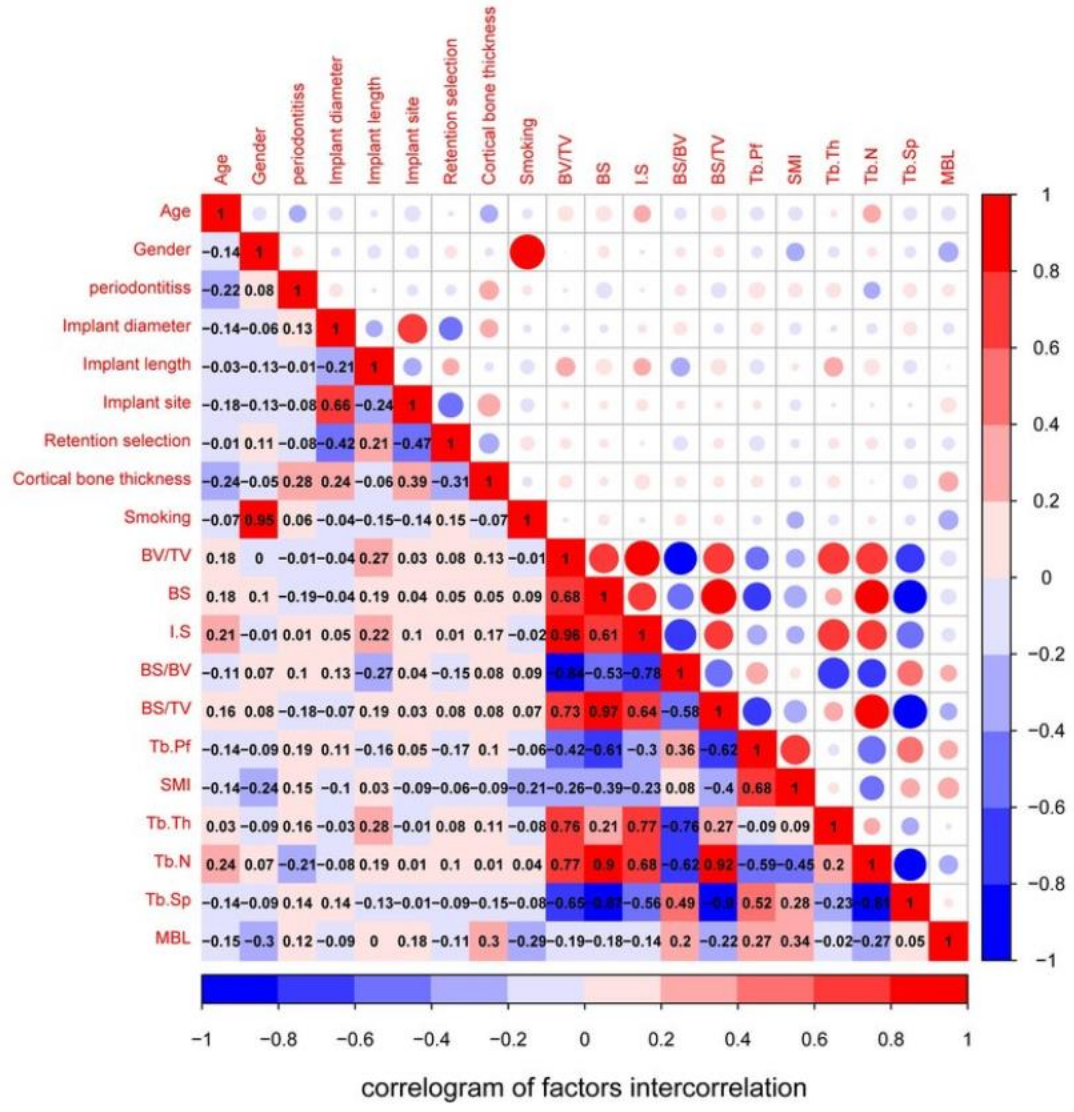


- Chart showcasing select proteins, GSMR effect sizes of said proteins, and their associated P-values. (example in 'Dual site proteomic analyses reveal potential drug targets for cardiovascular disease')

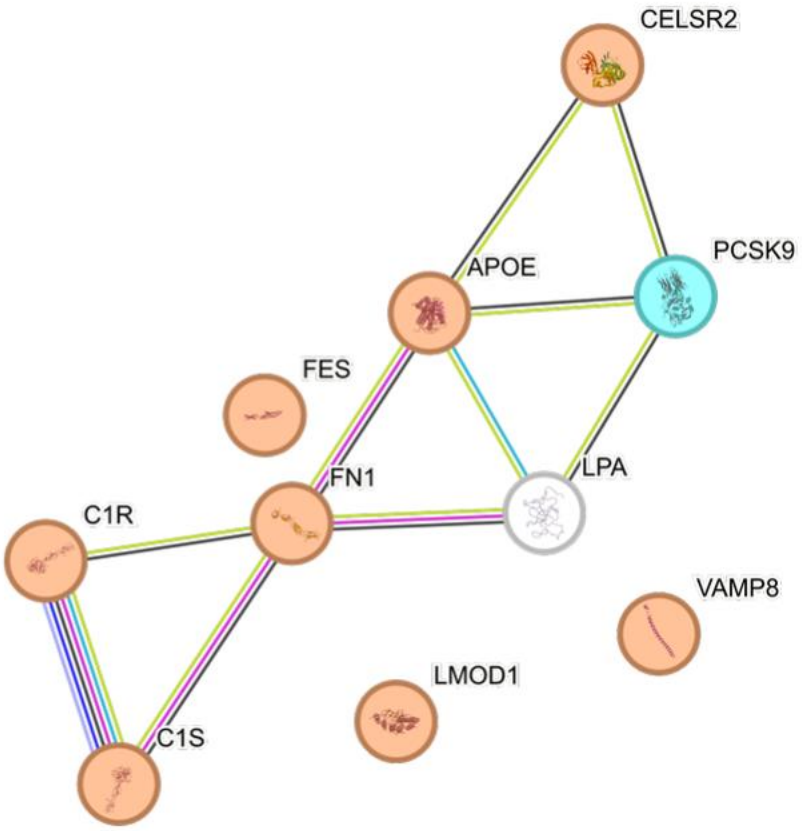
Protein Name	UniProtID(s)	GSMR effect size	GSMR p-value	H_3 PP	H_4 PP
Angina					
HLA-B	P01889	-0.090	8.035×10^{-5}	1.262×10^{-4}	9.690×10^{-1}
Type II Diabetes					
HLA-B	P01889	-0.104	9.553×10^{-9}	2.102×10^{-3}	9.979×10^{-1}
HLA-DRB1	P01911	-0.113	2.488×10^{-7}	9.663×10^{-1}	3.373×10^{-2}
FN3K	Q9H479	-0.096	2.723×10^{-4}	9.477×10^{-6}	9.343×10^{-4}
FN3KRP	Q9HA64	-0.112	2.596×10^{-3}	5.915×10^{-5}	4.751×10^{-1}
COMT	P21964, P21964-2	-0.112	7.732×10^{-4}	5.870×10^{-5}	7.288×10^{-1}
Essential Hypertension					
HLA-B	P01889	-0.045	1.520×10^{-5}	4.208×10^{-2}	9.475×10^{-1}
COMT	P21964, P21964-2	-0.094	1.092×10^{-5}	4.589×10^{-5}	9.961×10^{-1}

Table 1: PBMC measured isoform-specific protein group analysis. The table is divided into sections by the CVD/CVD risk-related trait, providing results that are significant in the GSMR analysis. Protein name along with the isoform-specific UniProtID(s) are provided in the first two columns. Additionally, the effect size and p-value from the GSMR procedure and the H_3 and H_4 posterior probabilities (PP) from the colocalisation are given. H_3 corresponds to the hypothesis that both the exposure and outcome have an associated SNP, but a different causal SNP in each case. Results are highlighted by H4 colocalisation posterior probability: the probability that both the exposure and outcome share a common causal SNP. Blue corresponds to mid-support, $0.5 < H4 < 0.8$, and orange corresponds to high-support, $H4 > 0.8$).

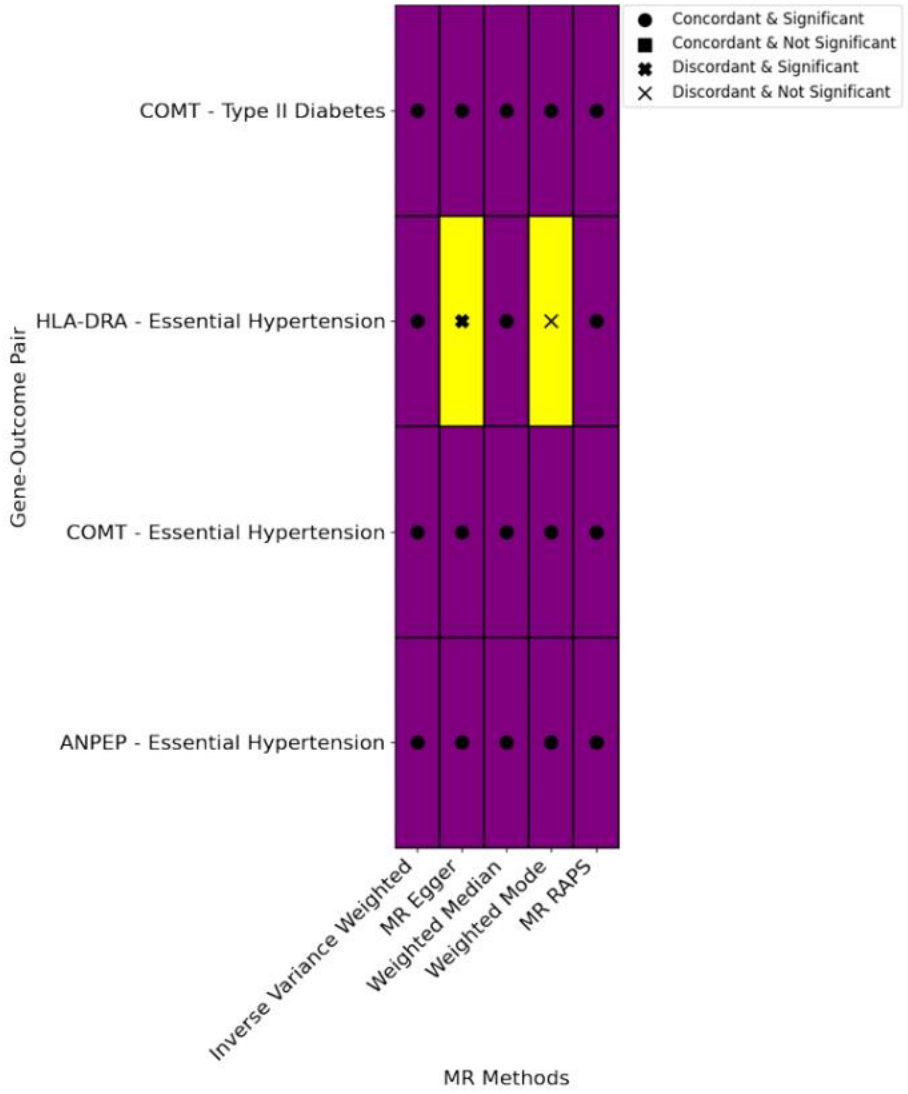
- A matrix comparing the results between the diseases. (a.k.a. number of shared significant protein associations or number of shared significant SNPs.)(Helps visualize pathophysiological pathways) (example from online)



- Any relevant String DB models.

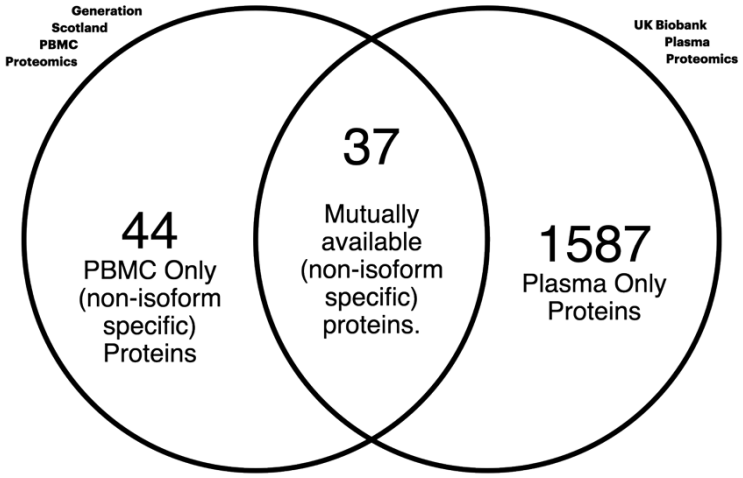


- Heat-map comparison in MRbase context. (Example in ‘Dual site proteomic analyses reveal potential drug targets for cardiovascular disease’)



- Graph showcasing the differences between plasma results and PBMC results. (Example in 'Dual site proteomic analyses reveal potential drug targets for cardiovascular disease')

(B)



- Quick question!!! Are we using both UK biobank and GScotland?? If so… a few graphs comparing any of those results are a must!


GSMR Result Comparison
(A)
Angina
Generation Scotland PBMC Proteomics
UK Biobank Plasma Proteomics
HLA-B
Concordant Direction of Effect
N/A
N/A
Opposing Direction of Effect
10 significant results [Protein only present in UK Biobank].
(B)
Type II Diabetes
HLA-C
HLA-DRB1
FN3K
FN3KRP
COMT
23 significant results [Protein only present in UK Biobank].
(C)
Heart Attack
No significant results.
4 significant results [Protein only present in UK Biobank].
(D)
High Cholesterol
19 significant results [Protein only present in UK Biobank].
(E)
Essential Hypertension
49 significant results [Protein only present in UK Biobank].
Tested in Generation Scotland but only significant in UK Biobank
ANPEP
HLA-DRA