Developing Robust Mendelian Randomization Methods to Account for Genetic Hitchhiking

Mendelian Randomization (MR) has emerged as a powerful tool in genetic epidemiology, enabling researchers to infer causal relationships between modifiable exposures and health outcomes using genetic variants as instrumental variables. This method leverages the random assortment of alleles during meiosis, akin to a natural randomized controlled trial, to mitigate confounding and reverse causation biases inherent in traditional observational studies (Davey Smith & Ebrahim, 2003). However, the accuracy of MR analyses can be compromised by genetic hitchhiking, a phenomenon where alleles at nearby loci are co-inherited due to linkage disequilibrium (LD) with a selected variant under positive selection.

Genetic hitchhiking, also known as selective sweeps, can introduce biases in MR studies by violating the core assumptions of instrumental variable analysis. Specifically, the presence of hitchhiking can lead to pleiotropy, where a genetic variant influences multiple traits, thereby confounding the causal inference (Lawlor et al., 2008). This research aims to enhance the robustness of MR by developing and integrating methods to detect and adjust for the effects of genetic hitchhiking.

The study will focus on three primary objectives:

1. **Identifying Selective Sweeps**: Utilizing genome-wide association study (GWAS) data, the research will develop algorithms to identify regions of the genome undergoing selective sweeps. This will involve detecting patterns of extended haplotype homozygosity and other signatures of selection (Burgess et al., 2015).

2. **Assessing Impact on MR Results**: The identified selective sweeps will be analyzed to understand their impact on MR results. This will include evaluating the extent to which genetic hitchhiking affects the validity of

instrumental variables and the magnitude of bias introduced in causal estimates (Davey Smith & Hemani, 2014).

3. **Implementing Statistical Techniques**: To mitigate potential biases, the study will implement advanced statistical techniques such as MR-Egger regression, weighted median, and contamination mixture methods. These methods will be adapted to account for the presence of genetic hitchhiking, ensuring more accurate causal inference (Hemani et al., 2018).

By addressing the challenges posed by genetic hitchhiking, this research aims to improve the reliability of MR analyses, thereby enhancing our understanding of the causal relationships between genetic variants and complex traits. The outcomes of this study will have significant implications for genetic epidemiology, providing more robust tools for causal inference and ultimately informing public health interventions and clinical practices.

Table of Contents

Detecting and Adjusting for Genetic Hitchhiking in Mendelian Randomization

Understanding Genetic Hitchhiking in Mendelian Randomization

Genetic hitchhiking, also known as genetic draft, occurs when an allele increases in frequency not because of its own advantage but due to its physical proximity to another allele under positive selection. This phenomenon

can introduce biases in Mendelian Randomization (MR) studies, which rely on genetic variants as instrumental variables (IVs) to infer causal relationships between exposures and outcomes. Detecting and adjusting for genetic hitchhiking is crucial to ensure the robustness of MR analyses.

Identifying Selective Sweeps

Selective sweeps are regions of the genome where a beneficial mutation has rapidly increased in frequency, carrying along nearby linked variants. These sweeps can be identified using various statistical methods:

1. **Extended Haplotype Homozygosity (EHH)**: This method measures the decay of haplotype homozygosity with distance from a core region. A slower decay indicates a selective sweep (NCBI).
2. **Integrated Haplotype Score (iHS)**: This score compares the EHH of derived and ancestral alleles. High absolute iHS values suggest recent positive selection (Nature).
3. **Composite of Multiple Signals (CMS)**: This method combines several statistics, including iHS, to detect selective sweeps with higher accuracy (Nature).

Assessing the Impact of Genetic Hitchhiking on MR Results

Genetic hitchhiking can confound MR results by violating the assumption that genetic variants used as IVs are not associated with confounders of the exposure-outcome relationship. To assess the impact of hitchhiking, researchers can:

1. **Examine Linkage Disequilibrium (LD)**: High LD between the IV and nearby loci under selection can indicate potential hitchhiking effects. Tools like LDlink can be used to explore LD patterns (NCBI).
2. **Perform Sensitivity Analyses**: Methods such as MR-Egger regression and the weighted median approach can help detect and adjust for pleiotropy, which may arise from hitchhiking (Nature).
3. **Use Multiple IVs**: Employing multiple genetic variants as IVs can mitigate the impact of any single variant being affected by hitchhiking.

The contamination mixture method can identify groups of variants with similar causal estimates, reducing bias (Nature).

Implementing Statistical Techniques to Mitigate Biases

To improve the accuracy of MR analyses, several statistical techniques can be implemented to detect and adjust for genetic hitchhiking:

1. **MR-PRESSO (Pleiotropy RESidual Sum and Outlier)**: This method detects and corrects for horizontal pleiotropy by identifying outlier variants that may be influenced by hitchhiking. It provides a corrected causal estimate after removing these outliers (Nature).
2. **MR-Clust**: This clustering approach groups genetic variants based on their causal estimates and identifies clusters that may be influenced by hitchhiking. It allows for the exclusion of biased clusters from the analysis (Nature).
3. **Penalized Regression Methods**: Techniques like LASSO (Least Absolute Shrinkage and Selection Operator) can be used to penalize the inclusion of variants with potential hitchhiking effects, thereby reducing their influence on the causal estimate (NCBI).
4. **Bayesian Methods**: Bayesian approaches can incorporate prior information about the likelihood of hitchhiking and adjust the causal estimates accordingly. These methods can provide more robust inferences in the presence of genetic hitchhiking (Nature).

Case Studies and Applications

Several studies have demonstrated the importance of detecting and adjusting for genetic hitchhiking in MR analyses:

1. **Cardiometabolic Traits**: A study investigating the causal relationship between lipid levels and coronary artery disease used MR-PRESSO to adjust for pleiotropy and hitchhiking, resulting in more accurate causal estimates (Nature).
2. **Mental Health Disorders**: Research on the causal effects of depression on various health outcomes employed MR-Clust to identify and exclude

clusters of variants affected by hitchhiking, improving the reliability of the findings (Nature).

3. **Cancer Risk**: An MR study on the relationship between body mass index (BMI) and cancer risk used penalized regression methods to account for genetic hitchhiking, leading to more robust conclusions (NCBI).

Future Directions

To further enhance the robustness of MR analyses, future research should focus on:

1. **Developing New Detection Methods**: Innovative statistical methods and machine learning approaches can improve the detection of genetic hitchhiking and selective sweeps.

2. **Integrating Multi-Omics Data**: Combining genetic data with other omics layers, such as epigenomics and transcriptomics, can provide a more comprehensive understanding of hitchhiking effects.

3. **Expanding Reference Panels**: Using diverse reference panels in LD analyses can help identify population-specific hitchhiking effects and improve the generalizability of MR findings.

4. **Collaborative Efforts**: Large-scale collaborations and consortia can facilitate the sharing of data and methods, accelerating the development of robust MR techniques.

By addressing the challenges posed by genetic hitchhiking, researchers can improve the accuracy and reliability of causal inferences in genetic epidemiology, ultimately advancing our understanding of the genetic basis of complex traits and diseases.

References

- https://augmentationlab.medium.com/writing-a-simulation-i-conways-game-of-life-86202a003040
- https://link.springer.com/chapter/10.1007/978-3-031-48465-0_5
- https://noweyr.github.io/
- https://pubmed.ncbi.nlm.nih.gov/35192729/

- https://www.biorxiv.org/content/10.1101/2022.08.30.505937v1
- https://www.cell.com/patterns/fulltext/S2666-3899(24)00096-5
- https://stackoverflow.com/questions/128478/should-import-statements-always-be-at-the-top-of-a-module
- https://conwaylife.com/forums/index.php
- https://conwaylife.com/?rle=9b2o$9bobo$4b2o6bo7b2o$2obo2bo2bo2bo7b2o$2o2b2o6bo$9bobo$9b2o!&name=(p30) queen bee shuttle
- https://www.sciencedirect.com/science/article/pii/S0927025620305401
- https://www.sciencedirect.com/science/article/pii/S0168952524000957
- https://www.nature.com/articles/s41467-019-14156-4
- https://academic.oup.com/biometrics/article/79/3/2771/7513843
- https://www.thelancet.com/journals/landia/article/PIIS2213-8587(23)00348-0/fulltext
- https://journals.plos.org/plosgenetics/article/file?id=10.1371/journal.pgen.1010596&type=printable
- https://www.bmj.com/content/362/bmj.k601
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7614635/
- https://learnpython.com/blog/define-function-python/
- https://andrewslabucsf.github.io/MR-tutorial/scripts/intro.html
- https://www.medrxiv.org/content/10.1101/2024.05.23.24307776v1.full.pdf
- https://www.sciencedirect.com/science/article/pii/S135964622300430X
- https://www.nature.com/articles/s41562-024-01879-8
- https://amymariemason.github.io/MR/
- https://pubmed.ncbi.nlm.nih.gov/25388149/
- https://biofish.medium.com/pyconway-6dc2479de409
- https://github.com/anpu9/UCB-CSC1C-proj1-Game-of-Life-
- https://www.sciencedirect.com/science/article/abs/pii/S0925838809019021
- https://longnow.org/ideas/conways-game-of-life-and-three-millennia-of-human-history/
- https://pubmed.ncbi.nlm.nih.gov/38706291/
- https://doaj.org/article/d8724b989ce84ef1b01d3ff655716c09
- https://www.freecodecamp.org/news/functions-in-python-a-beginners-guide/
- https://www.sciencedirect.com/science/article/pii/S0925838822008374
- https://pubmed.ncbi.nlm.nih.gov/32249995/

- https://stackoverflow.com/questions/139180/how-to-list-all-functions-in-a-module
- https://pubmed.ncbi.nlm.nih.gov/27625185/
- https://pypi.org/project/genal-python/
- https://python.land/introduction-to-python/functions
- https://www.nature.com/articles/s41408-024-01035-5
- https://academic.oup.com/mbe/article/40/6/msad139/7194669
- https://www.annualreviews.org/content/journals/10.1146/annurev-genom-090314-050016
- https://onlinelibrary.wiley.com/doi/10.1002/brb3.3551
- https://wellcomeopenresearch.org/articles/4-186
- https://research-information.bris.ac.uk/en/publications/strategies-to-investigate-and-mitigate-collider-bias-in-genetic-a
- https://academic.oup.com/ije/article/52/1/227/6651126
- https://www.nature.com/articles/srep13802
- https://towardsdev.com/python-functions-with-code-examples-91f3e73df655
- https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1009703
- https://www.sciencegate.app/keyword/693600
- https://www.nature.com/articles/s41380-024-02613-6
- https://academic.oup.com/ije/article/51/6/2031/6671803
- https://en.wikipedia.org/wiki/Conway'sGameof_Life
- https://www.sciencedirect.com/science/article/pii/S0921509321010285
- https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1011246
- https://www.mdpi.com/2075-4701/13/7/1174
- https://www.sciencedirect.com/science/article/pii/S1359645414008295
- https://www.nature.com/articles/s43856-024-00530-x
- https://bmjopen.bmj.com/content/bmjopen/13/9/e072087.full.pdf?with-ds=yes
- https://www.discovermagazine.com/the-sciences/mathematicians-prove-the-omniperiodicity-of-conways-game-of-life?ref=gorillasun.de
- https://www.nature.com/articles/s41380-021-01173-3
- https://onlinelibrary.wiley.com/doi/full/10.1002/gepi.22445
- https://github.com/matthijsknigge/mendelianrandomization
- https://conwaylife.com/wiki/Conway'sGameof_Life

- https://www.nature.com/articles/s41598-024-59326-7
- https://www.sciencedirect.com/science/article/pii/S0925838809019021
- https://www.youtube.com/watch?v=aabO1dwT2Fs
- https://github.com/peteryin21/py-merp
- https://kyushu-u.elsevierpure.com/en/publications/the-role-of-trace-element-segregation-in-the-eutectic-modificatio
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6969055/
- https://www.sciencedirect.com/science/article/pii/S1471531702000093
- https://www.nature.com/articles/s41467-019-09432-2
- https://www.publichealth.columbia.edu/academics/non-degree-special-programs/professional-non-degree-programs/skills-health-research-professionals-sharp-training/mendelian-randomization
- https://onlinelibrary.wiley.com/doi/pdf/10.1002/gepi.22544
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6424604/
- https://www.quantamagazine.org/maths-game-of-life-reveals-long-sought-repeating-patterns-20240118/
- https://www.nature.com/articles/s41598-024-62178-w
- https://www.semanticscholar.org/paper/Strategies-to-investigate-and-mitigate-collider-in-Mitchell-Hartley/dcd87514820bfb472e41b38dd7ace61a5890be98
- https://www.nature.com/articles/s43586-021-00092-5
- https://academic.oup.com/pmj/advance-article/doi/10.1093/postmj/qgae069/7688384
- https://www.nature.com/articles/s41431-022-01038-5
- https://pubmed.ncbi.nlm.nih.gov/36821633/
- https://medium.com/@alper.bulbul1/exploring-the-causal-pathways-with-python-implementing-mendelian-randomization-methods-in-research-adca766bcf2f
- https://www.nytimes.com/2020/12/28/science/math-conway-game-of-life.html
- https://www.cambridge.org/core/journals/psychological-medicine/article/mendelian-randomization-causal-inference-leveraging-genetic-data/333B418DF7B88B53D58032EA316D0834
- https://www.cell.com/ajhg/fulltext/S0002-9297(22)00551-1
- https://github.com/NCBI-Hackathons/MR_BACOn
- https://www.sciencedirect.com/science/article/pii/S0165032724004919
- https://conwaylife.com/forums/viewtopic.php?t=6489

- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8725623/
- https://www.nature.com/articles/s41598-019-41919-2
- https://www.frontiersin.org/articles/10.3389/fcimb.2016.00057/full
- https://www.sciencedirect.com/science/article/pii/S1359645416301963
- https://academic.oup.com/bioinformatics/article/31/6/957/215107
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7384151/
- https://github.com/OpenOmics/mr-seek
- https://github.com/topics/mendelianrandomization
- https://conwaylife.com/
- https://academic.oup.com/book/40819/chapter/348789403
- https://www.researchgate.net/publication/268231351MeRPahigh-throughputpipelineforMendelianRandomizationAnalysis
- https://www.mayoclinicproceedings.org/article/S0025-6196(23)00077-0/fulltext
- https://link.springer.com/article/10.1007/s11665-024-09511-6
- http://py-merp.github.io/intro.html
- https://www.medrxiv.org/content/10.1101/2024.05.23.24307776v1
- https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009238
- https://www.codewithc.com/an-in-depth-guide-to-python-functions-creation-usage-and-best-practices/
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9949638/
- https://math.stackexchange.com/questions/1536288/what-are-the-practical-uses-of-game-of-life-or-langtons-ant
- https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010596
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9123217/
- https://link.springer.com/article/10.1007/s10654-022-00879-0
- https://onlinelibrary.wiley.com/doi/full/10.1002/jbm4.10675
- https://academic.oup.com/ije/article/50/5/1639/6146681
- https://onlinelibrary.wiley.com/doi/10.1111/mec.13288
- https://onlinelibrary.wiley.com/doi/full/10.1111/mec.13288
- https://pubmed.ncbi.nlm.nih.gov/32965465/
- https://www.medrxiv.org/content/medrxiv/early/2024/04/18/2024.04.16.24305933.full.pdf
- https://www.nature.com/articles/s41398-024-02950-8
- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11161388/

- https://www.medrxiv.org/content/10.1101/2022.04.22.22274166v1.full.pdf
- https://ajp.psychiatryonline.org/doi/10.1176/appi.ajp.2020.19111158
- https://pubmed.ncbi.nlm.nih.gov/38263619/