Mendelian randomization using biological foundation models improves discovery power over additive QTLs and prioritizes cell-type specific therapeutic targets

Noah Lorincz-Comi[1†], Tina Keshavarzian[1], Varun Lodaya[1], Emma Bohn[1], Michael Iacocca[1], Gaganjot Kaur[1], Christopher E Hart, Brendan J. Frey[1], Christopher B. Cole[1†]

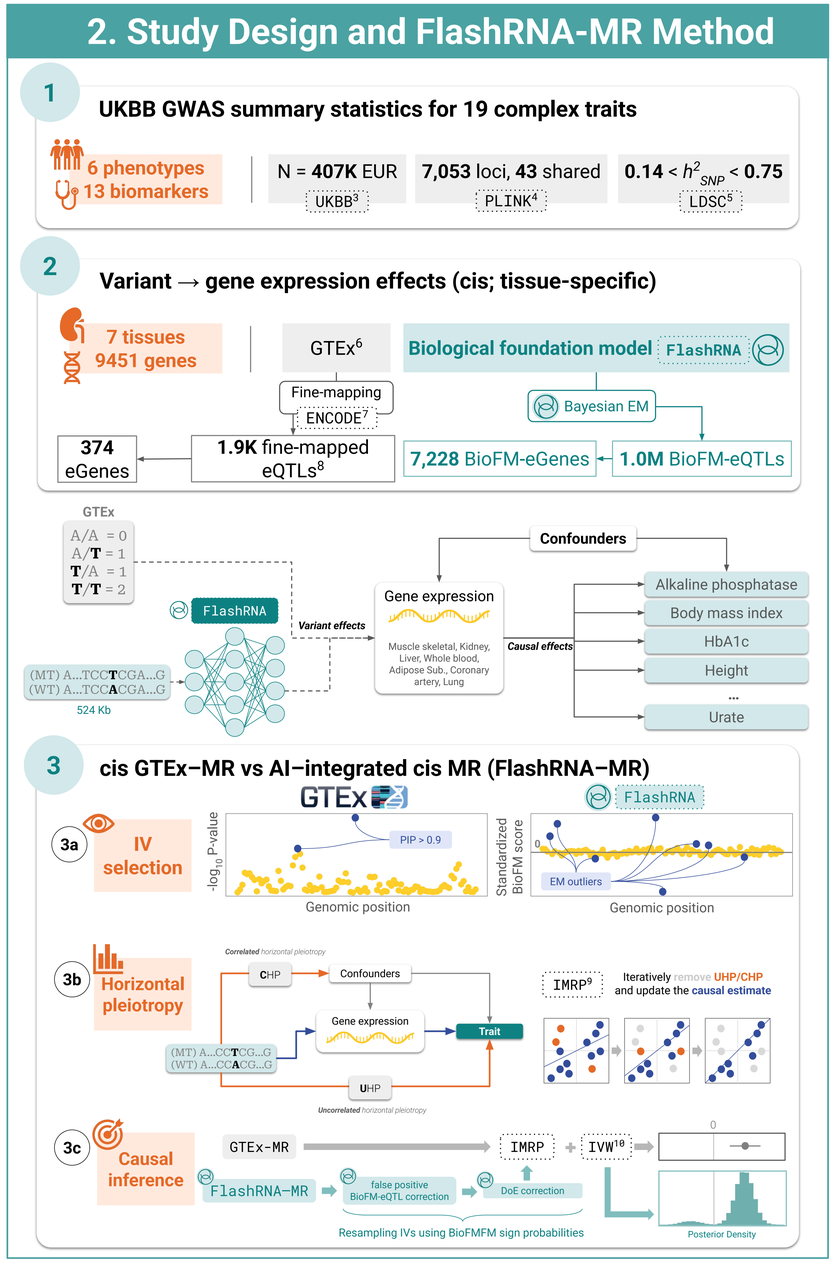
[1]Deep Genomics Inc., Toronto, ON
[†]Corresponding authors: noah.lorinczcomi@deepgenomics.com, christopher@deepgenomics.com

deep genomics

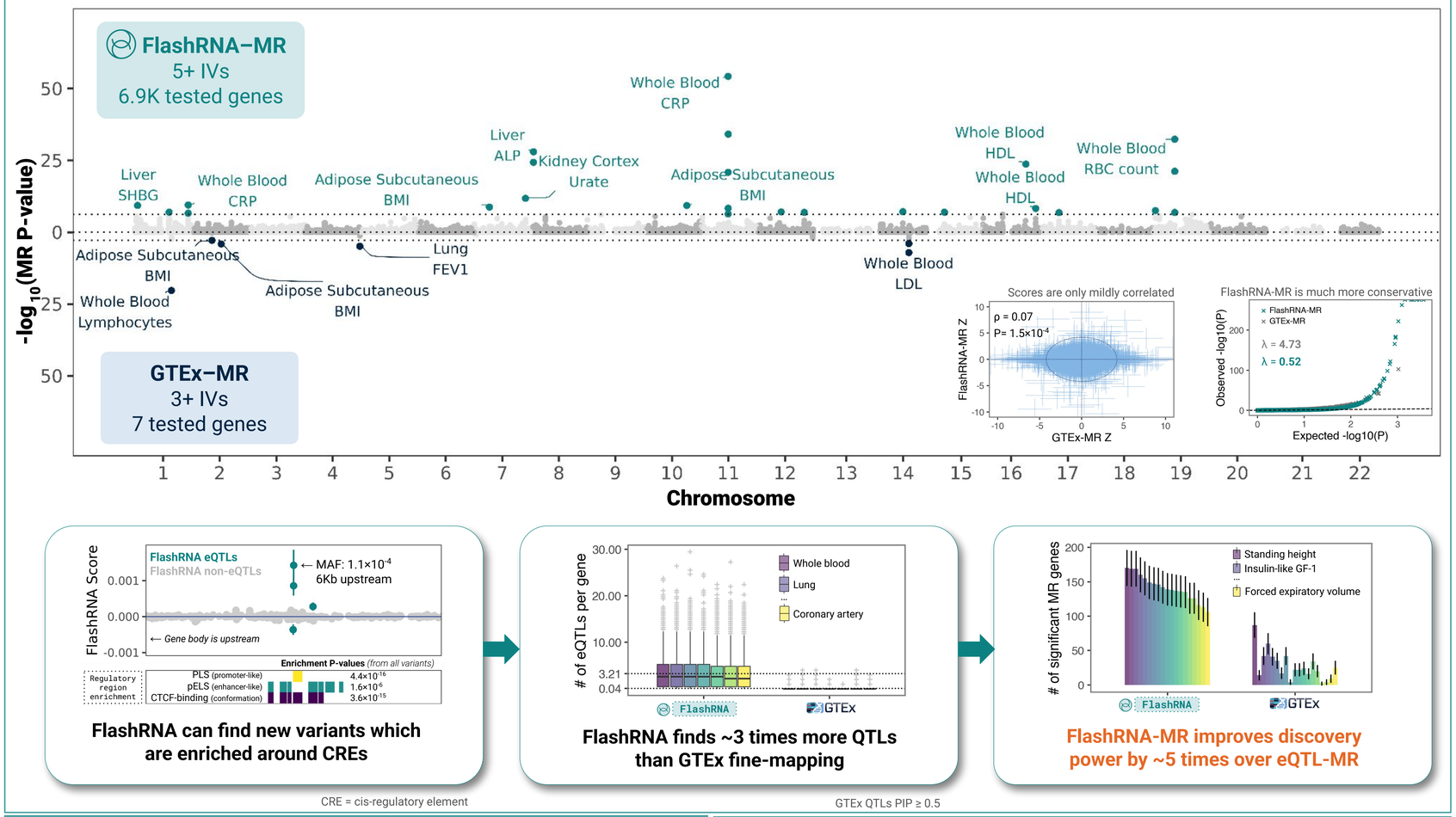
1. Background and Objectives

- **Biological foundation models** (BioFMs) such as FlashRNA[1] and AlphaGenome[2] can predict the regulatory effects of genetic variants across biological contexts, but frameworks for translating these predictions into **therapeutic target discoveries remain limited**.
- Mendelian Randomization (MR) is an established causal inference framework for target discovery that uses **genetically predicted molecular perturbations** to estimate a gene's effects on disease and other traits **using genetic variants as instrumental variables (IVs)**.
- **Quantitative trait loci (QTLs)** are genetic variants associated with variation in **molecular phenotypes**, such as gene expression (eQTLs).
- **Conventional MR** relies on experimentally measured QTLs, which may provide **few strong genetic instruments** due to low statistical power to detect them.
- We developed **AI-integrated Mendelian Randomization** (AIMR) to combine BioFM-predicted xQTLs with MR, enabling higher-powered, context-specific causal inference of **molecular mechanisms and therapeutic targets**.

2. Study Design and FlashRNA-MR Method



3. FlashRNA-MR finds 5x more genome-wide significant associations than eQTL-MR (n=141 vs 27)



FlashRNA can find new variants which are enriched around CREs

FlashRNA finds ~3 times more QTLs than GTEx fine-mapping

FlashRNA-MR improves discovery power by ~5 times over eQTL-MR

CRE = cis-regulatory element

GTEx QTLs PIP ≥ 0.5

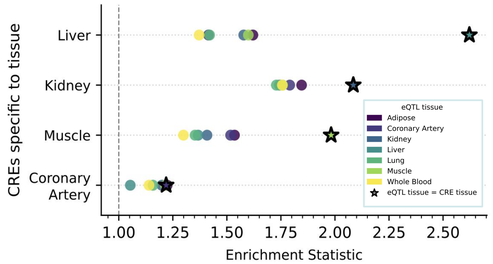
4. Next steps and future directions

Application

1. **Population-rare, high-impact variants may be better captured by the BioFM** than GTEx.
2. **Structural variants** characterized by the BioFM may have large functional effects and reveal key regulatory mechanisms.
3. **Cell type-specific BioFM predictions** from fine-tuned models may uncover novel cell & disease-specific therapeutic targets.

Methodological

1. Higher enrichment of **cis BioFM-eQTLs in tissue-specific CREs** suggests that active non-coding genomic regions could provide a **useful prior for BioFM-eQTL inference** across variant–gene pairs.
2. BioFM cis-variant effect estimates show **statistical correlation across genes and tissues**, highlighting an opportunity to further **improve inference by jointly modeling** or selecting among **correlated gene–tissue combinations**.



BioFM QTLs from FlashRNA are ~1.5x more enriched in CREs of their matched tissues than unmatched tissues

5. Study Conclusions

1. Here we introduce a method for **inferring QTLs** from BioFM models, finding at least one causal variant for **19x more genes** than cis-eQTLs from GTEx v8.
2. We show that these FlashRNA-QTLs can be used for discovery using **Mendelian Randomization (MR)** **increasing power by 5x** while remaining conservative.

This study demonstrates that BioFMs represent an important advance in causal target discovery when considered in robust analytical frameworks.

6. References

[1] Jung, A. J., Zhu, H., ... & Frey, B. J. (2025). FlashRNA: An Efficient Model for Regulatory Genomics. bioRxiv, 2025-10.
[2] Avsec Z., Latysheva, N., ... & Kohli, P. (2026). Advancing regulatory variant effect prediction with AlphaGenome. Nature, 649, 1206–1218
[3] Neale Lab. UK Biobank GWAS results. Available at: nealelab.is/uk-biobank. Accessed September 25, 2026.
[4] Chang, C. C., Chow, C. C., ... & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience, 4(1), s13742-015.
[5] Bulik-Sullivan, B. K., Loh, P. R., ... & Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature genetics, 47(3), 291-295.
[6] GTEx Consortium, Ardlie, K. G., ... & Dermitzakis, E. T. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science, 348(6235), 648-660.
[7] Sloan, C. A., Chan, E. T., ... & Cherry, J. M. (2016). ENCODE data at the ENCODE portal. Nucleic acids research, 44(D1), D726-D732.
[8] Kerimov, N., Tambets, ... & Alasoo, K. (2023). eQTL Catalogue 2023: New datasets, X chromosome QTLs, and improved detection and visualisation of transcript-level QTLs. PLoS Genetics, 19(9), e1010932.
[9] Zhu, X., Li, X., ... & Wang, T. (2021). An iterative approach to detect pleiotropy and perform Mendelian Randomization analysis using GWAS summary statistics. Bioinformatics, 37(10), 1390-1400.
[10] Burgess, S., Butterworth, A., & Thompson, S. G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. Genetic epidemiology, 37(7), 658-665.

Deep Genomics

Thank you for stopping by!

Interested to learn more?

Visit deepgenomics.com