Association and mechanism between $PM_{2.5}$ and COVID-19: AI, epidemiology, and genomics

Guoqing Feng

Tsinghua University, Beijing, China

2024.09.23

WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

Introduction

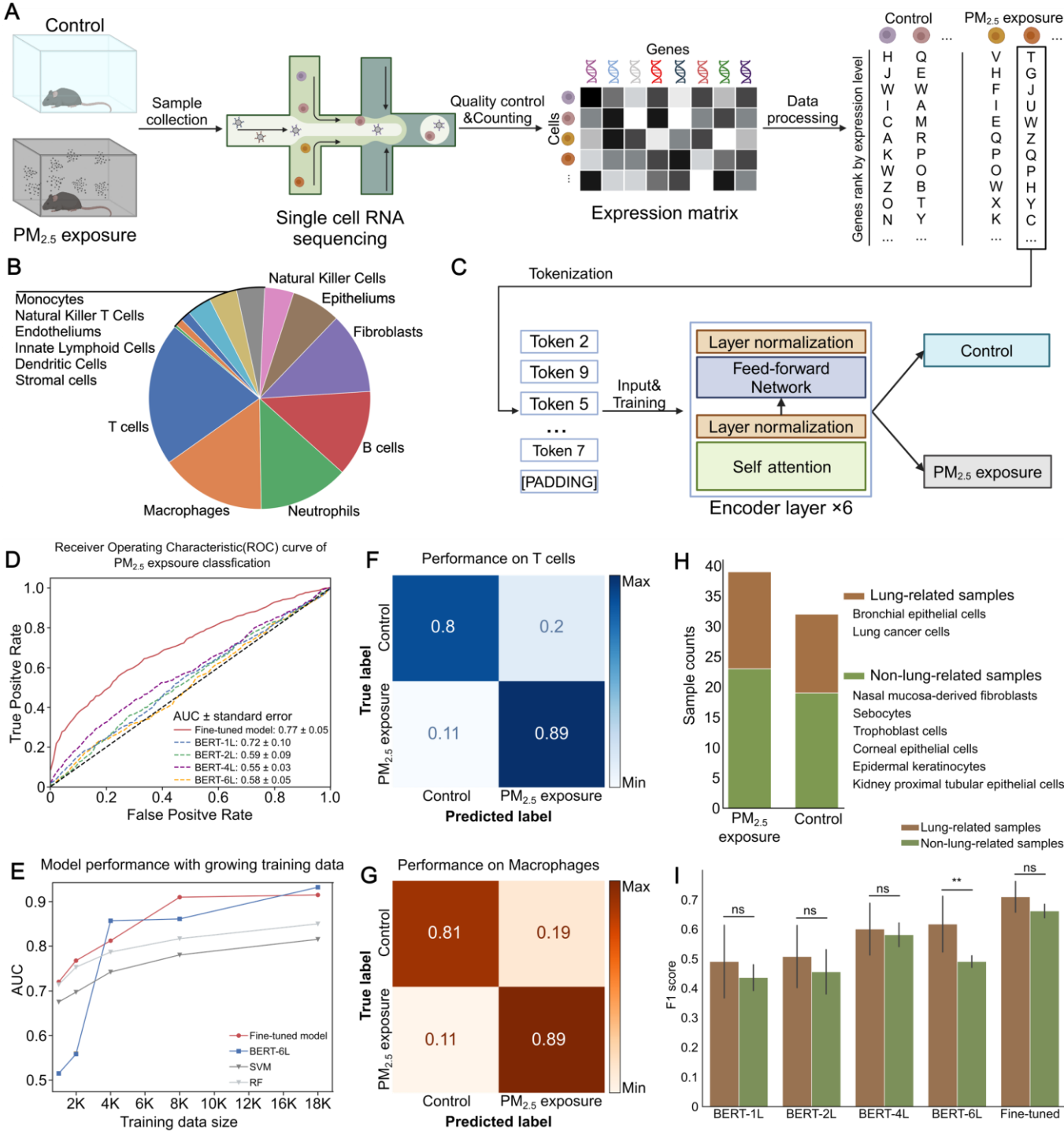
- From particulates to pandemic
  - Environmental factors could interfere viral infection
  - Association between $PM_{2.5}$ exposure and COVID-19 admission and mortality
  - However, the molecular mechanism is still elusive

- AI for science
  - AI empowered novel insights of scientific research
  - Considerable data size in biology (in transcriptomes)
  - Transfer learning ability of AI models (pretrain, fine-tune)

- Solving epidemiological problems with transcriptome AI
  - Obtain exposure/outcome-related gene network knowledge using limit data training
  - Discover potential exposure-outcome pairs with explicit pathway inference
  - Validate molecular mechanism by in silico perturbation

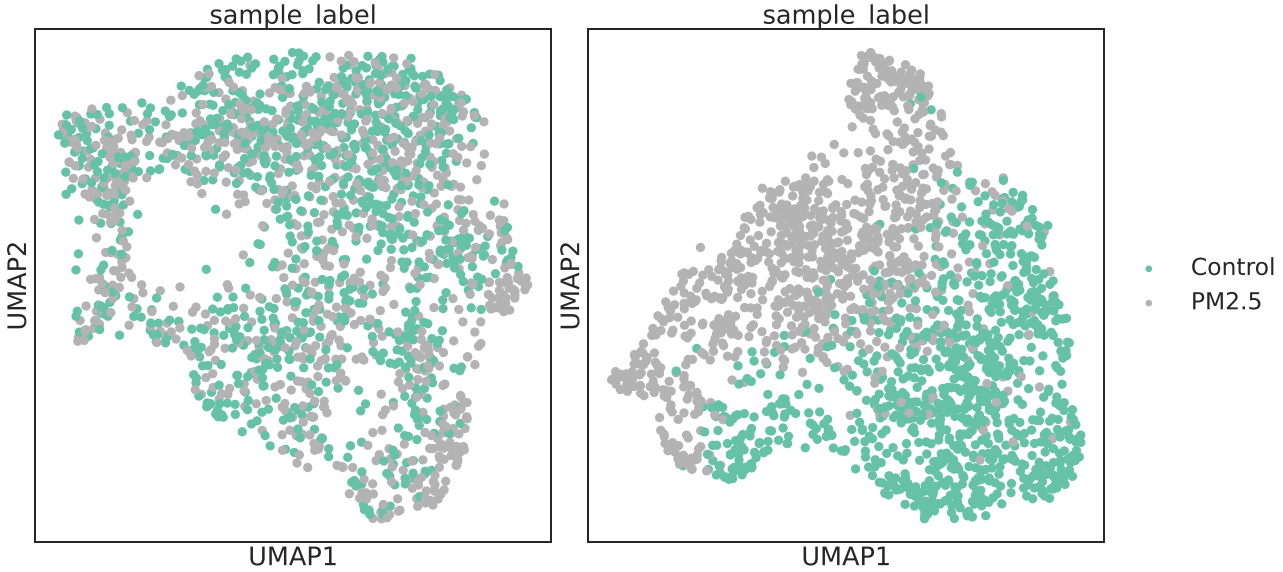
AI transcriptome model on $PM_{2.5}$ classification

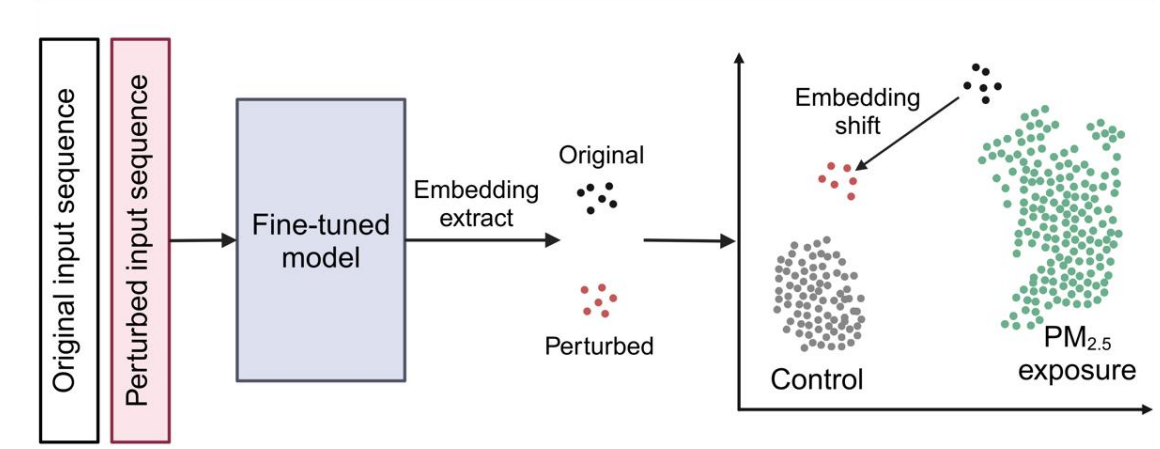
- Establishment of AI models
  - **Data source**: Lung tissue collected from mouse under $PM_{2.5}$ exposure
  - **Encoding**: rank-base input format
  - **Architecture**: 6 layer BERT
  - **Training**: fine-tune (Geneformer)/from scratch
- Performance
  - Outperform traditional ML
  - Generalization ability
- Implication
  - Good understanding on gene network under $PM_{2.5}$ exposure

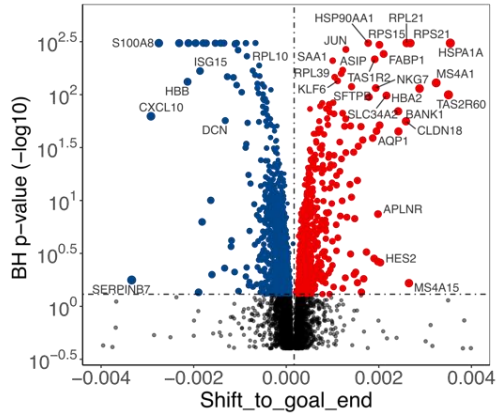


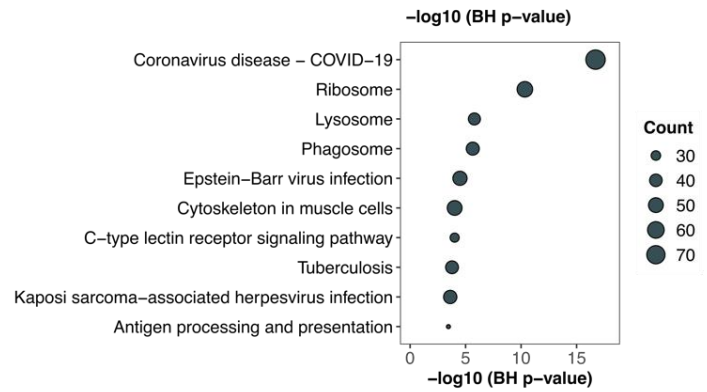
Positive association between $PM_{2.5}$ and COVID-19

- Revealed by in silico perturbation
  - Label clustering: models not trained (left)/ trained (right) with $PM_{2.5}$ transcriptomes
  - Gene input perturbation: e.g. ABC→A_C, large-scale screening
  - Embedding shifts obtained from 20k genes&3k cells—significant gene sets
  - Enrichment analysis: COVID-19, …
- Verified by epidemiological studies





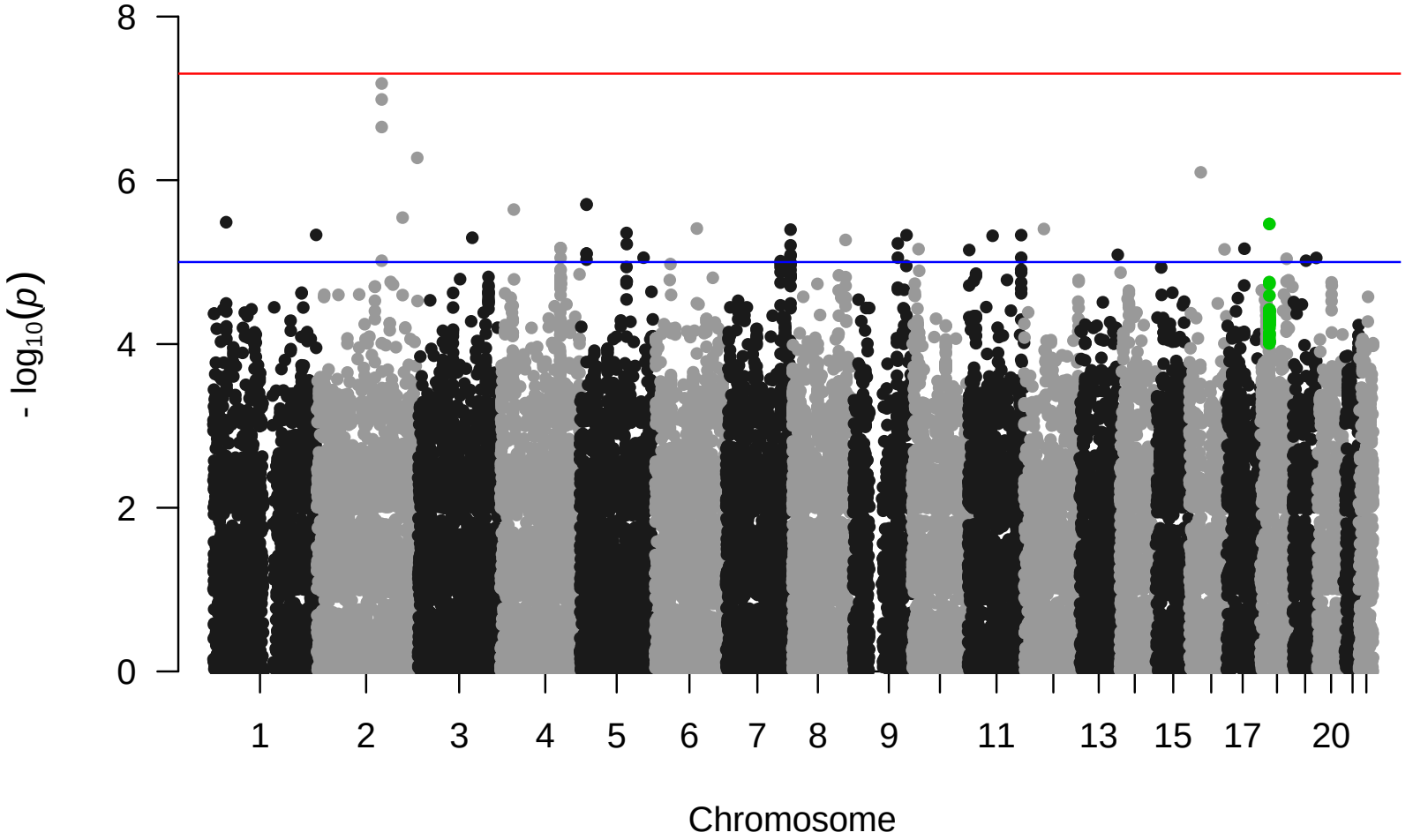




Identification of candidate variants

- Genome-wide association study
  - To identify genomic variants that are statistically associated with a risk for a disease
  - (Single nucleotide polymorphisms) SNP—COVID: consistent with previous studies
  - $PM_{2.5}$ x SNP – COVID: a set of highly significant variants on Chromosome 18
- significant ≠ functional !

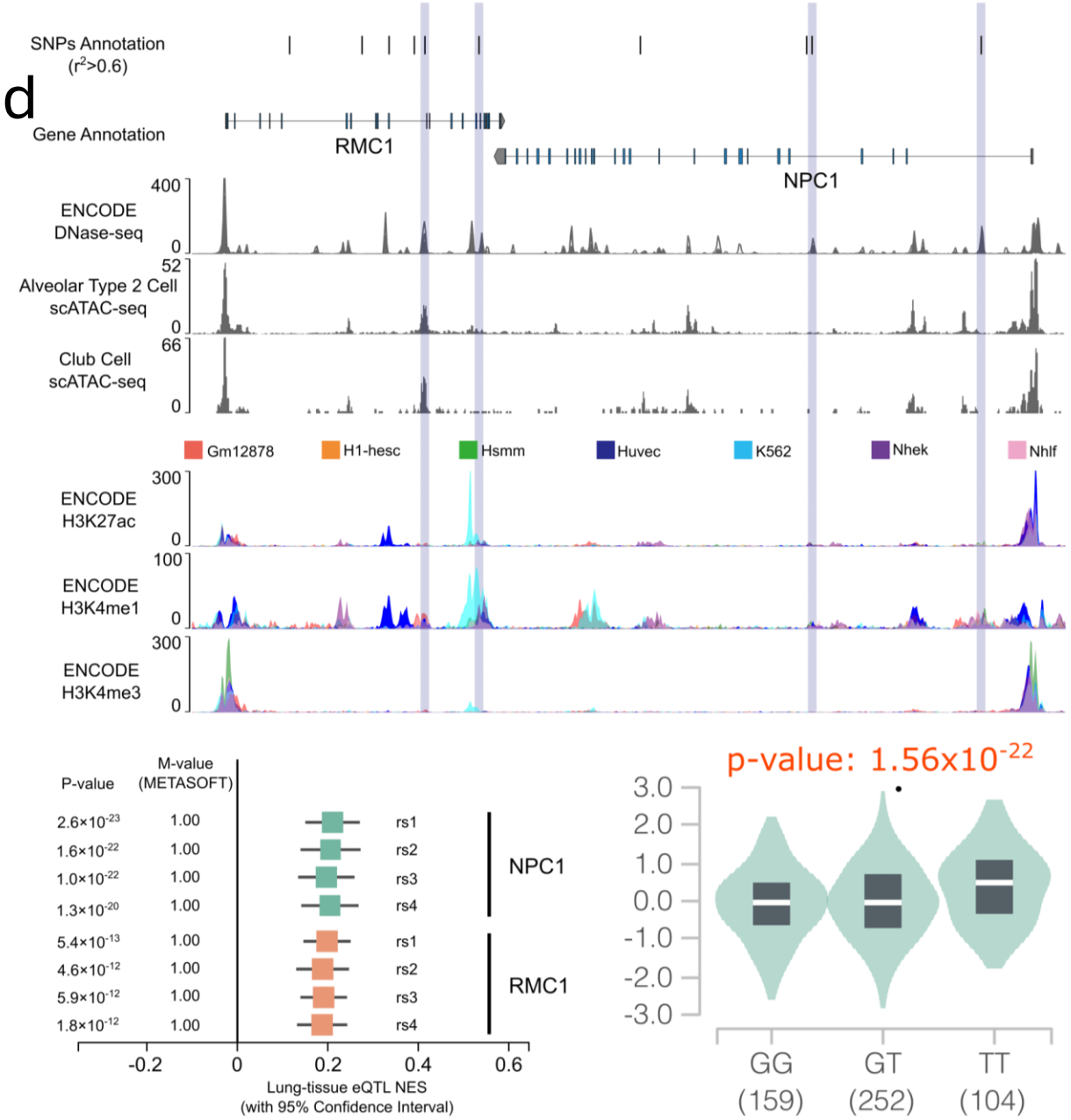
Genotype/SNP: GG, GC, CC (0, 1, 2)
Phenotype: infected, non-infected



NPC1 and *RMC1* as potential regulated genes

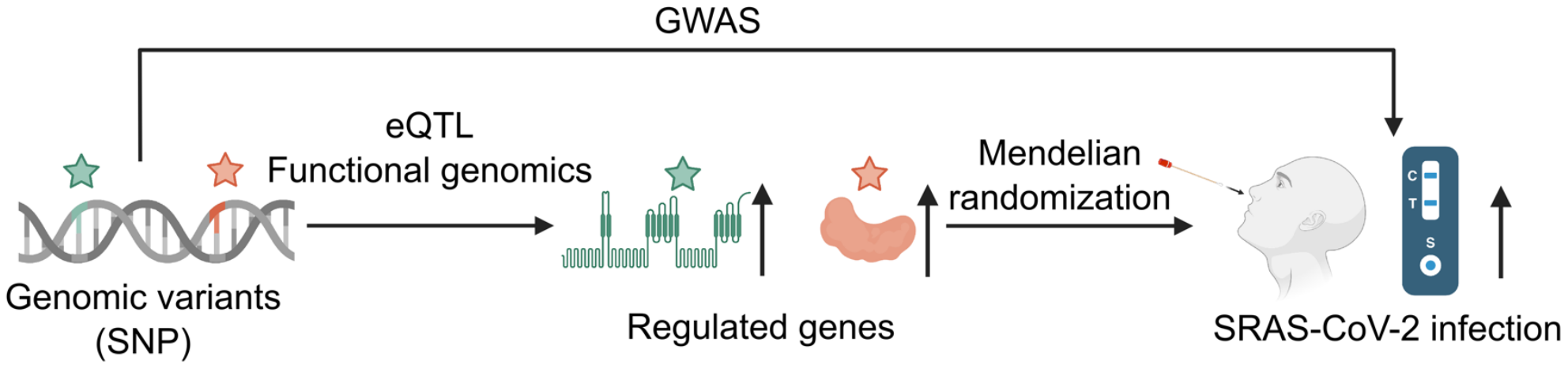
- Identifying functional loci is complicated
  - Genetic linkage
  - Tissue-specific gene expression pattern
- Potential mechanism inferred by:
  - Chromosome accessibility: DNase-seq, ATAC-seq, scATAC-seq, …
  - Histone modification: H3K27ac, H3K4me1, …
- *RMC1* and *NPC1* may be downstream regulated genes
  - Adjacent to the functional sites
  - Verified by eQTL data

  # Genomic loci that explain variation in expression levels of mRNAs.



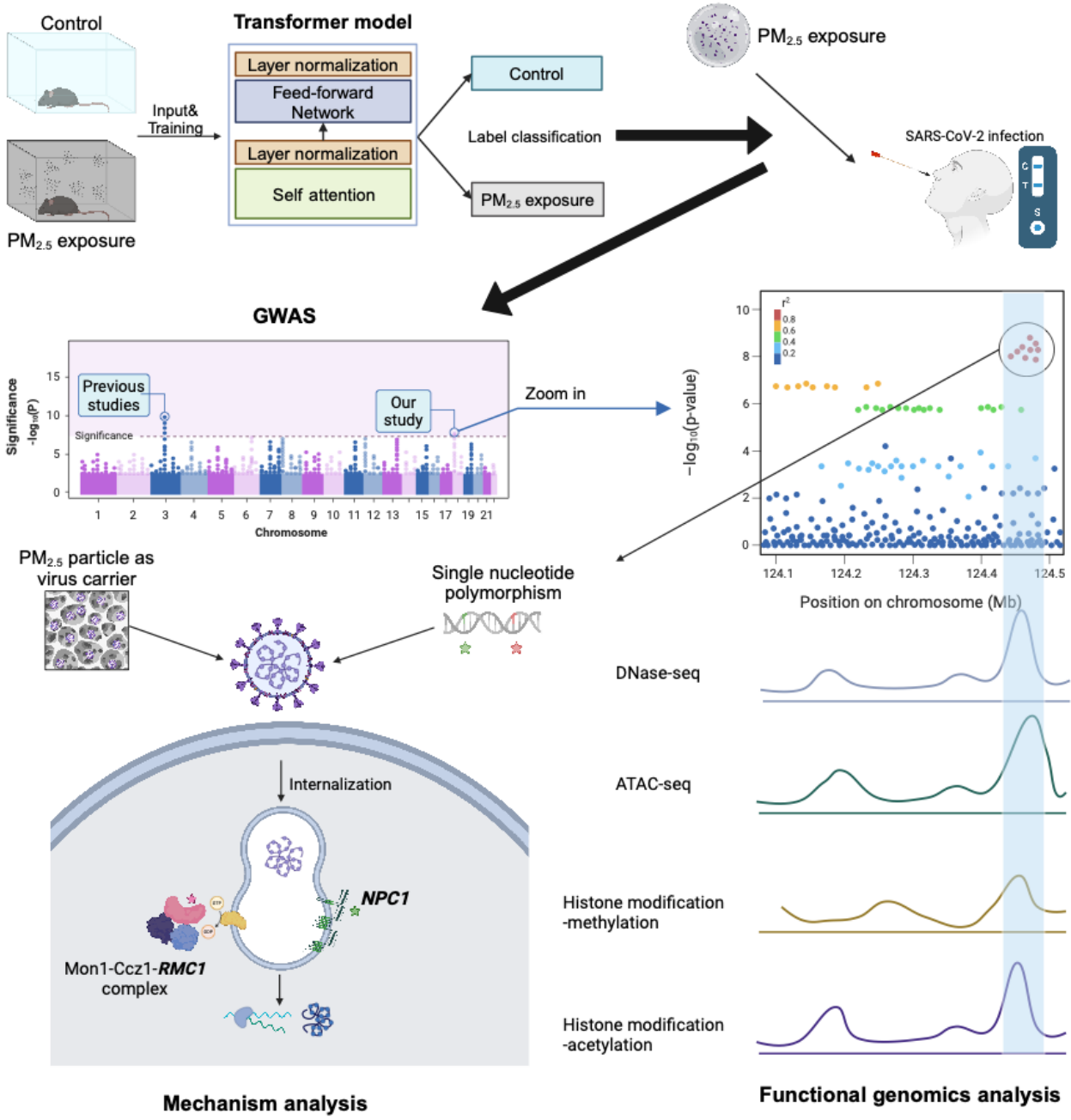
RMC1/NPC1 and SARS-CoV-2 infection

- Mendelian randomization
  - Based on GWAS and eQTL
- Previous studies
  - Biology of *RMC1* and *NPC1*
  - COVID-19 CRISPR screening
- More verification
  - *In silico*
  - *In vitro*
  - …



Overview of methods and key findings

- Transcriptome AI
  - Association between $PM_{2.5}$ and COVID-19
- GWAS
  - Variants interacting with $PM_{2.5}$ and increasing the risk of COVID-19
- Functional genomic analysis
  - *RMC1* and *NPC1* as candidate effector genes of SARS-CoV-2 infection under $PM_{2.5}$ exposure



Implication& Future work

- Implications:
  - $PM_{2.5}$ can assist viral infection, and the molecular mechanism is related to genotype – more stringent $PM_{2.5}$ control, identification of susceptible population
  - Associations and molecular mechanisms between various exposures and health outcomes have not been discovered, can be investigated by AI model

- Future work: more AI involvement in epidemiology
  - Data integration& AI development
    - Multi-modalites data
    - More prior knowledges
  - AI participation throughout the analysis:
    - AI-based mechanism prediction
  - We are building more versatile models/pipelines…
    - Significant related genes/pathways
    - Prediction of potential outcomes/ mechanism

Acknowledgements

Support from the National Natural Science Foundation of China

Qian Di lab, Tsinghua University

Key references

1. Theodoris, C.V., Xiao, L., Chopra, A. *et al.* Transfer learning enables predictions in network biology. *Nature* **618**, 616–624 (2023). https://doi.org/10.1038/s41586-023-06139-9

2. Downes, D.J., Cross, A.R., Hua, P. *et al.* Identification of *LZTFL1* as a candidate effector gene at a COVID-19 risk locus. *Nat Genet* **53**, 1606–1615 (2021). https://doi.org/10.1038/s41588-021-00955-3

3. Zheng Dong et al., Airborne fine particles drive H1N1 viruses deep into the lower respiratory tract and distant organs.Sci.Adv.9,eadf2165(2023).DOI:10.1126/sciadv.adf2165

Thank you!

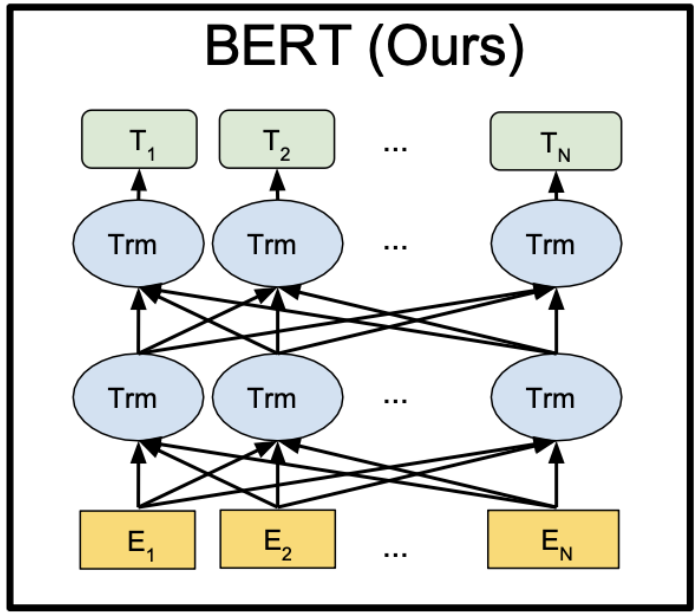
Guoqing Feng

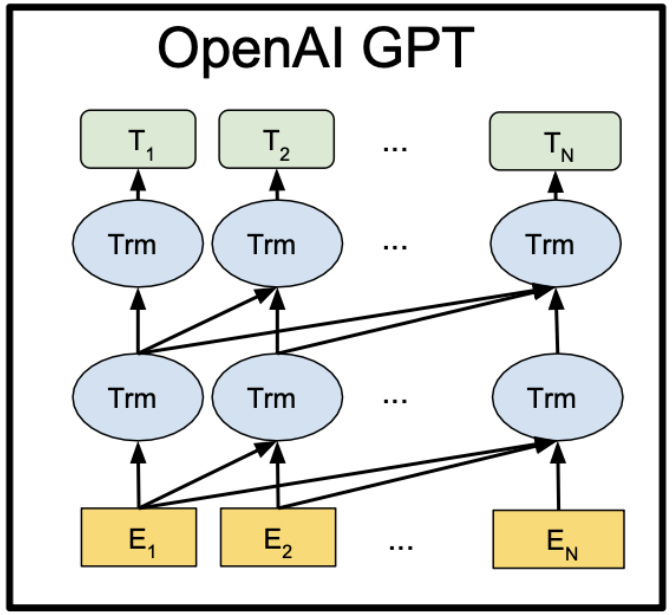
School of Biomedical Engineering& Vanke School of Public Health, Tsinghua University

fgq21@mails.tsinghua.edu.cn

WCE
WORLD CONGRESS OF EPIDEMIOLOGY 2024

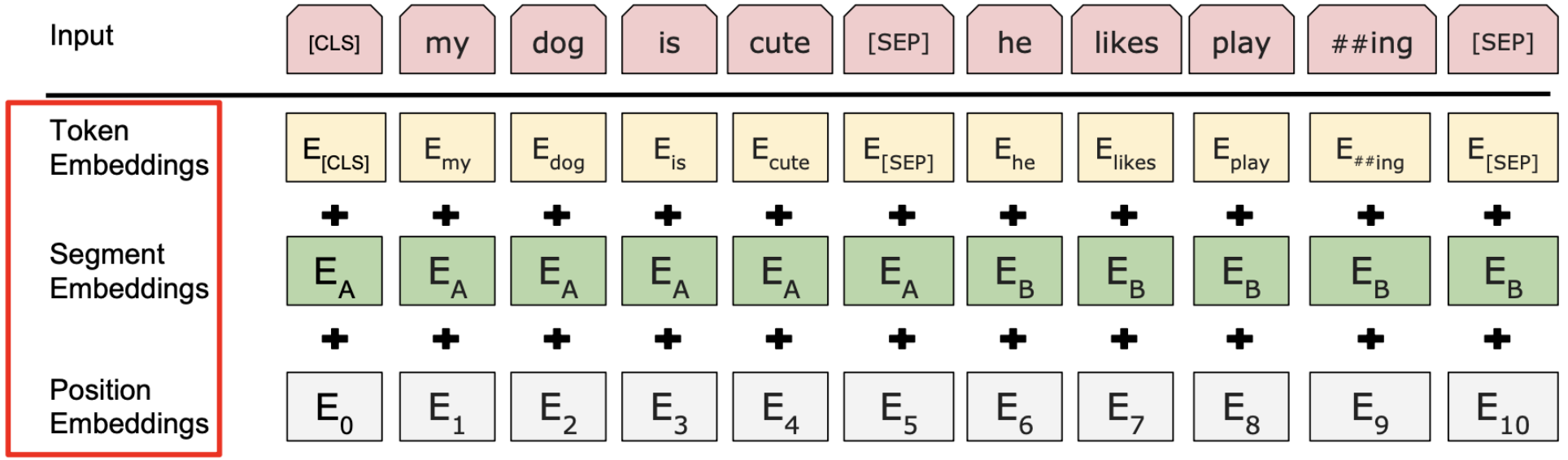
BERT structure





$\mathbf{BERT_{BASE}}$:
L =12, H=768, A =12,
Total Parameters=110M

$\mathbf{BERT_{LARGE}}$:
L =24, H =1024, A=16,
Total Parameters=340M



Study flowchart