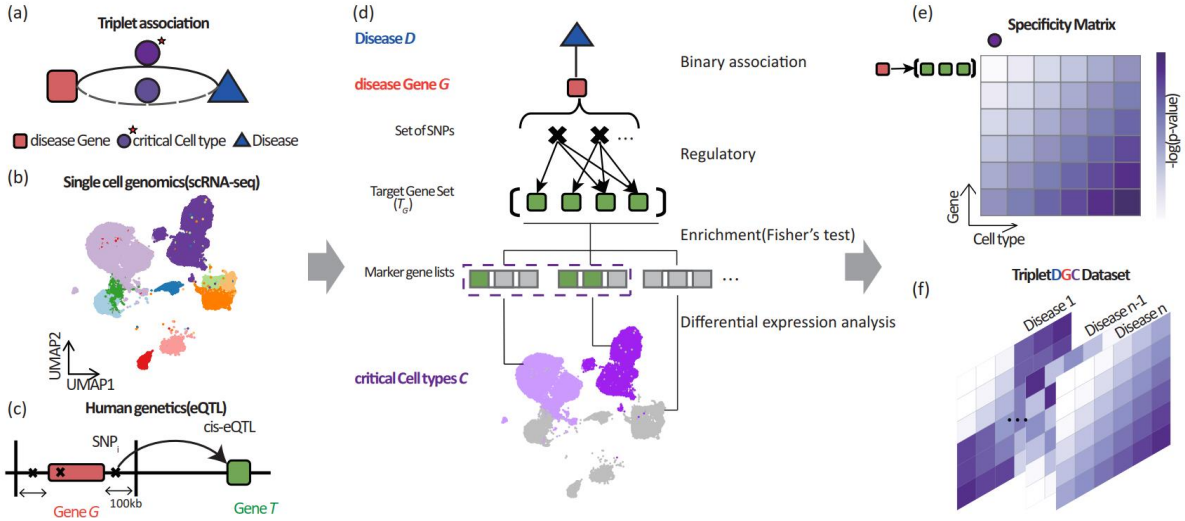
TripletDGC: Assessing critical Cell types of Disease Genes by integrating single-cell genomics and human genetics

Yaoqi SHOU, Bingbo WANG, Zitian YANG

Frontiers of Computer Science, DOI: 10.1007/s11704-025-41165-y

Problems & Ideas

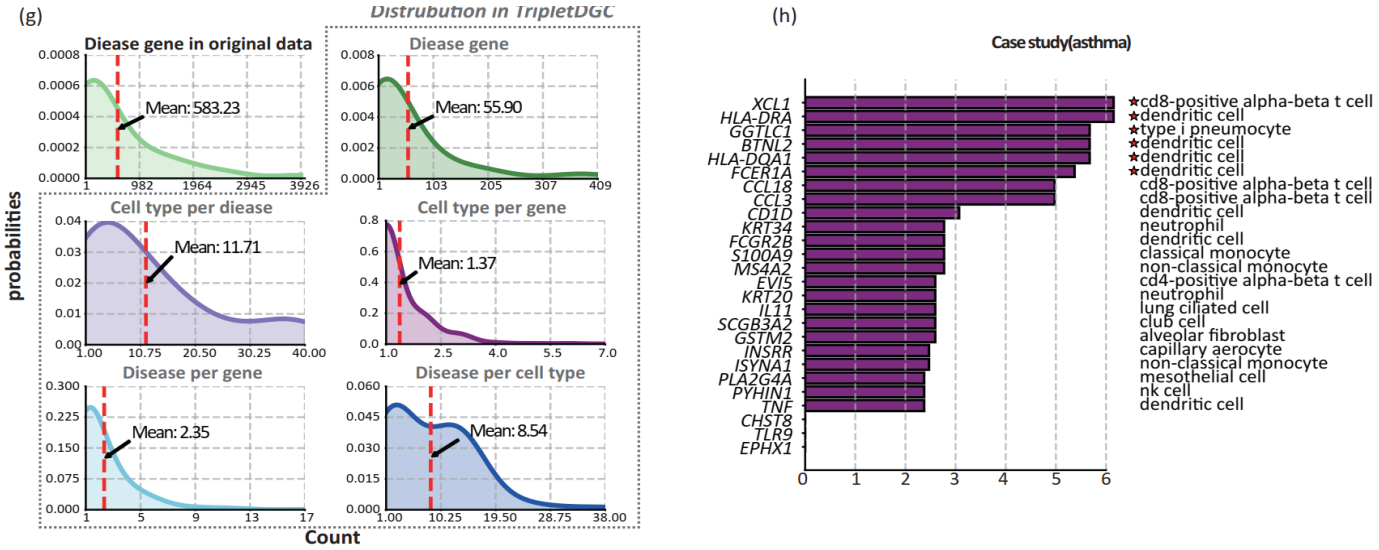
- Problems of the binary relation with gene and disease:
  - The binary relation fails to capture the complex interplay between genetic risk loci, disease genes, and specific cell states.
  - Binary relationships do not provide the granularity needed to associate disease occurrence with cell states.
- Ideas: identify critical cell types regulated by genetic risk loci associated with specific disease genes by integrating single-cell RNA sequencing data with human genetics data



(a) Triplet association between a disease, a disease gene and a cell type. (b) A toy UMAP visualization showing single-cell genomics (scRNA-seq) data. (c) An illustration of cis-eQTL showing human genetics(eQTL) data we used. (d) the framework of assessing critical cell types for the disease gene. (e) Specificity matrix, visualization of cell type specificity index(-log p-value) for each gene in the disease D. (f) Triplet DGC dataset structure.

Main Contributions

- Contributions:
  - Triplet Relational Dataset (TripletDGC): The construction of a triplet dataset linking disease, disease gene, and critical cell types, providing a high-resolution understanding of disease mechanisms;
  - TripletDGC delineates the association between disease genes-diseases-cell types, offering novel insights into disease mechanisms within the feature space defined by single-cell data;



(g) Statistical analysis of TripletDGC dataset and original data, the means of different counts are highlighted with red dashed lines. (h) A bar chart visualizing the critical cell types associated with selected asthma disease genes, based on their specificity index (-log(p-value)). Stars indicate triplets with a specificity index greater than 5.