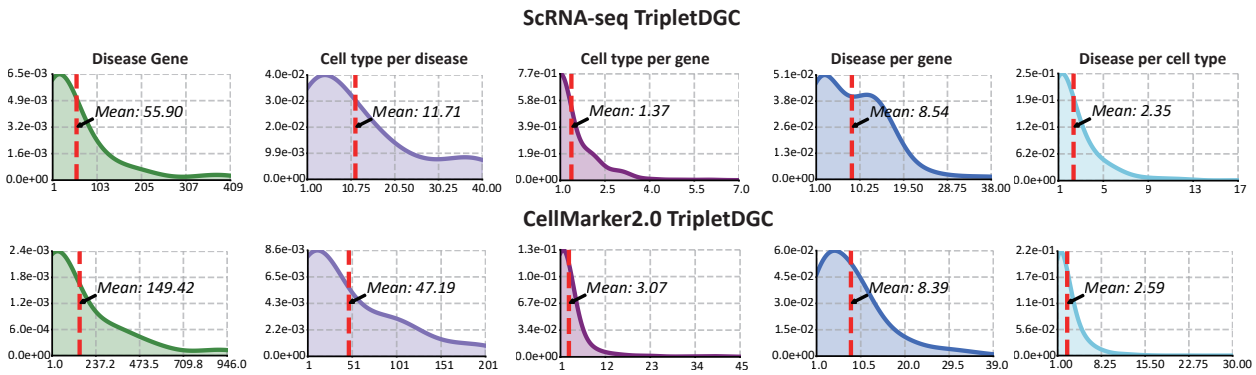
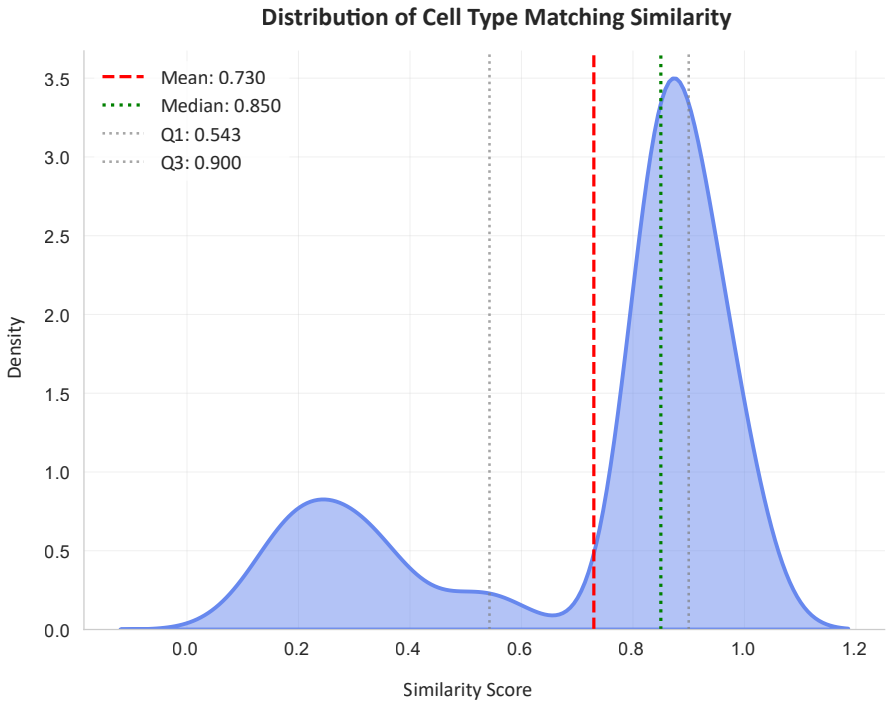
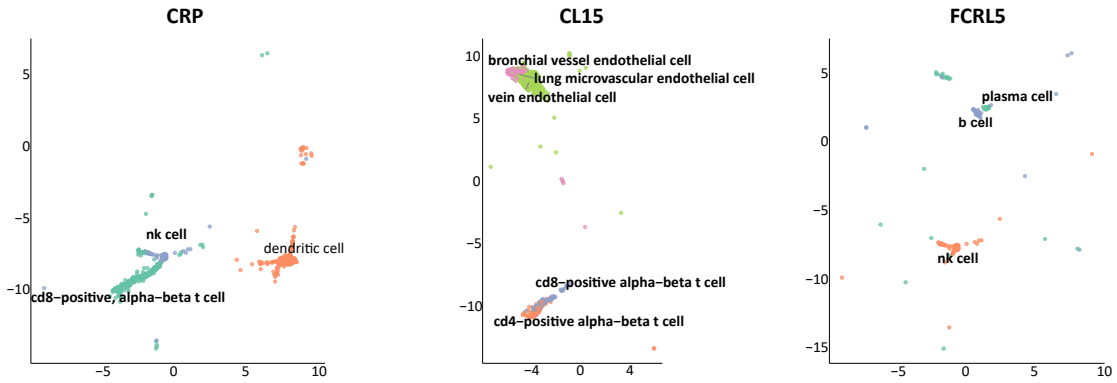
Supplementary Material



Supplementary Figure 1: **Statistical analysis of scRNA-seq and CellMarker2.0 TripletDGCs** We conducted a statistical analysis comparing the TripletDGC derived from single-cell data with that constructed from the CellMarker2.0 database. The results demonstrate that the mean values of the CellMarker2.0 TripletDGC are approximately three times higher than those of the scRNA-seq TripletDGC, both in terms of disease gene discovery with cell type specificity (green) and the identification of specific cell types associated with each disease gene and disease (medium purple, purple). This discrepancy may be attributed to the difference in the number of cell types between the two datasets, with the CellMarker2.0 database encompassing approximately 1,715 cell types, compared to only 475 cell types in the single-cell dataset. However, the number of diseases associated with each gene and cell type discovery is similar between the two TripletDGCs (blue, skyblue), underscoring the robustness of our findings regarding the specificity of individual genes.



Supplementary Figure 2: **Distribution analysis of cell type matching similarity scores.** The density plot illustrates the distribution of similarity scores between disease genes and cell type-specific patterns (n = 1758 disease-disease gene pairs). We conducted an analysis to evaluate the consistency of cell type specificity for disease genes identified by the CellMarker2.0 TripletDGC and the scRNA-seq TripletDGC. Due to the inherent discrepancies in cell type annotations between the two databases, we employed a string-matching algorithm to assign similarity scores. Specifically, for each disease gene shared by both TripletDGCs, we constructed a similarity matrix based on the sets of specific cell types identified for that gene. The similarity score for each disease gene was calculated by summing the scores of each row in the matrix and dividing by the size of the smaller cell type set. The resulting mean similarity score of 0.73 indicates a high degree of consistency in the specific cell types associated with disease genes across the two TripletDGCs, suggesting robust agreement in the identification of cell type specificity for most disease genes.



Supplementary Figure 3: **Single-cell data provides a feature space that captures cellular heterogeneity at high resolution.** The UMAP plot illustrates the distribution of three asthma-related disease genes within the feature space of single-cell sequencing data.When examining gene-level patterns, the single-cell dataset offers a cell-dimensional feature representation for each gene, enabling a comprehensive characterization of gene expression profiles across diverse cell types. This unique perspective not only enhances our understanding of gene-specific behaviors but also provides a valuable foundation for exploring intergenic relationships, such as co-expression patterns, regulatory networks, and functional interactions within specific cellular contexts.Furthermore, the single-cell dataset delineates cellular states with high precision, enabling the identification of specific cell states influenced by disease genes. This capability provides a robust foundation for further investigation into the mechanisms by which these genes contribute to disease pathology. By mapping gene effects to distinct cellular states, researchers can uncover critical insights into disease progression and identify potential therapeutic targets.