

What is claimed is:

1. A method for improving data processing for gene-gene correlation, comprising:
processing gene expression data for normalization or imputation;
applying a noise regularization process to the normalized or imputed gene expression data; and
applying a gene-gene correlation calculation process to obtain correlated gene pairs.
2. The method of claim 1, wherein the gene expression data is single cell gene expression data.
3. The method of claim 1, wherein the noise regularization process comprises adding a random noise to an expression value of a gene in a cell in an expression matrix.
4. The method of claim 3, wherein the random noise is determined by an expression level of the gene.
5. The method of claim 3, wherein the random noise is determined by:
determining an expression distribution of the gene across all of the cells in the expression matrix;
taking from about 0.1 to about 20 percentile of an expression level of the gene as a maximal noise level;
generating a random number ranging from 0 to the maximal noise level under uniform distribution; and
adding the random number to the expression value of the gene in the cell in the expression matrix to obtain a noise regularized expression matrix.
6. The method of claim 3, wherein the random noise is determined by:
determining an expression distribution of the gene across all of the cells in the expression matrix;
taking one percentile of an expression level of the gene as a maximal noise level; x
generating a random number ranging from 0 to the maximal noise level under uniform distribution; and

adding the random number to the expression value of the gene in the cell in the expression matrix to obtain a noise regularized expression matrix.

7. The method of claim 1, wherein the gene-gene correlation calculation process is conducted within cell clusters.

8. The method of claim 1, further comprising enriching the gene expression data that is associated with the correlated gene pairs.

9. The method of claim 1 or claim 3 or claim 4 or claim 5 or claim 6, wherein Total Unique Molecular Identifier Normalization (NormUMI), Regularized Negative Binomial Regression (NBR), a deep count autoencoder network (DCA), Markov affinity-based graph imputation of cells (MAGIC), or single-cell analysis via expression recovery (SAVER) is used for processing gene expression data for normalization or imputation.

10. The method of claim 1 or claim 3 or claim 4 or claim 5 or claim 6, further comprising constructing a gene-gene correlation network based on the correlated gene pairs.

11. The method of claim 10, wherein the gene-gene correlation networks are cell type-specific.

12. The method of claim 10, further comprising using the gene-gene correlation networks for mapping molecular interactions, guiding experimental designs to investigate the biological events, discovering biomarkers, guiding comparative network analysis, guiding drug designs, identifying changes of gene-gene interactions by comparing healthy and disease states of cells, guiding drug development, predicting transcription regulation of genes, improving drug efficiency or identifying drug resistance factors.

13. A gene-gene correlation network, wherein the network is constructed based on correlated gene pairs, and wherein the correlated gene pairs are obtained using the method of claim 1.

14. A computer-implemented method for data processing for gene-gene correlation, comprising:

retrieving gene expression data;

processing the gene expression data for normalization or imputation;

applying a noise regularization process to the normalized or imputed gene expression data;

applying a gene-gene correlation calculation process to obtain correlated gene pairs, and constructing a gene-gene correlation network based on the correlated gene pairs.

15. The method of claim 14, wherein the gene expression data is single cell gene expression data.

16. The method of claim 14, wherein the noise regularization process comprises adding a random noise to an expression value of a gene in a cell in an expression matrix.

17. The method of claim 16, wherein the random noise is determined by an expression level of the gene.

18. The method of claim 16, wherein the random noise is determined by:

determining an expression distribution of the gene across all of the cells in the expression matrix;

taking from about 0.1 to about 20 percentile of an expression level of the gene as a maximal noise level;

generating a random number ranging from 0 to the maximal noise level under uniform distribution; and

adding the random number to the expression value of the gene in the cell in the expression matrix to obtain a noise regularized expression matrix.

19. The method of claim 16, wherein the random noise is determined by:

determining an expression distribution of the gene across all of the cells in the expression matrix;

taking one percentile of an expression level of the gene as a maximal noise level;

generating a random number ranging from 0 to the maximal noise level under uniform distribution; and

adding the random number to the expression value of the gene in the cell in the expression matrix to obtain a noise regularized expression matrix.

20. The method of claim 14, wherein the gene-gene correlation calculation process is conducted within cell clusters.

21. The method of claim 14, further comprising enriching the gene expression data that is associated with the correlated gene pairs.

22. The method of claim 14 or claim 16 or claim 17 or claim 18 or claim 19, wherein Total Unique Molecular Identifier Normalization (NormUMI), Regularized Negative Binomial Regression (NBR), a deep count autoencoder network (DCA), Markov affinity-based graph imputation of cells (MAGIC), or single-cell analysis via expression recovery (SAVER) is used for processing gene expression data for normalization or imputation.

23. The method of claim 14, wherein the gene-gene correlation networks are cell type-specific.

24. The method of claim 14 or claim 16 or claim 17 or claim 18 or claim 19, further comprising using the gene-gene correlation networks for mapping molecular interactions, guiding experimental designs to investigate the biological events, discovering biomarkers, guiding comparative network analysis, guiding drug designs, identifying changes of gene-gene interactions by comparing healthy and disease states of cells, guiding drug development, predicting transcription regulation of genes, improving drug efficiency or identifying drug resistance factors.

25. A system for generating a gene-gene network, comprising:

- a database configured to store gene expression data;
- a memory configured to store instructions;
- at least one processor coupled to the memory, wherein the at least one processor is configured to execute instructions for:
 - retrieving the gene expression data,
 - processing the gene expression data for normalization or imputation,
 - applying a noise regularization process to the normalized or imputed gene expression data,
 - applying a gene-gene correlation calculation process to obtain correlated gene pairs; and

constructing a gene-gene correlation network based on the correlated gene pairs; and
a user interface coupled to the processor and capable of receiving a query for gene-gene correlation and displaying the results of the correlated gene pairs and the constructed gene-gene correlation networks.

26. The system of claim 25, wherein the gene expression data is single cell gene expression data.

27. The system of claim 25, wherein the noise regularization process comprises adding a random noise to an expression value of a gene in a cell in an expression matrix.

28. The system of claim 27, wherein the random noise is determined by an expression level of the gene.

29. The system of claim 27, wherein the random noise is determined by:
determining an expression distribution of the gene across all of the cells in the expression matrix;

taking from about 0.1 to about 20 percentile of an expression level of the gene as a maximal noise level;

generating a random number ranging from 0 to the maximal noise level under uniform distribution; and

adding the random number to the expression value of the gene in the cell in the expression matrix to obtain a noise regularized expression matrix.

30. The system of claim 27, wherein the random noise is determined by:
determining an expression distribution of the gene across all of the cells in the expression matrix;

taking one percentile of an expression level of the gene as a maximal noise level;

generating a random number ranging from 0 to the maximal noise level under uniform distribution; and

adding the random number to the expression value of the gene in the cell in the expression matrix to obtain a noise regularized expression matrix.

31. The system of claim 25, wherein the gene-gene correlation calculation process is conducted with cell clusters.

32. The system of claim 25, wherein the at least one processor is further configured to enrich the gene expression data that is associated with the correlated gene pairs.

33. The system of claim 25 or claim 27 or claim 28 or claim 29 or claim 30, wherein Total Unique Molecular Identifier Normalization (NormUMI), Regularized Negative Binomial Regression (NBR), a deep count autoencoder network (DCA), Markov affinity-based graph imputation of cells (MAGIC), or single-cell analysis via expression recovery (SAVER) is used for processing gene expression data for normalization or imputation.

34. The system of claim 25, wherein the gene-gene correlation networks are cell type-specific.

35. The system of claim 25 or claim 27 or claim 28 or claim 29 or claim 30, wherein the at least one processor is further configured to utilize the gene-gene correlation networks for mapping molecular interactions, guiding experimental designs to investigate the biological events, discovering biomarkers, guiding comparative network analysis, guiding drug designs, identifying changes of gene-gene interactions by comparing healthy and disease states of cells, guiding drug development, predicting transcription regulation of genes, improving drug efficiency or identifying drug resistance factors.