scCapsNet: a deep learning classifier with the capability of interpretable feature extraction, applicable for single cell RNA data analysis

Lifei Wang[a,b], Rui Nie[a,b], Ruyue Xin[c], Jiang Zhang[c,*], Jun Cai[a,b,*]

[a]Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China

[b]University of Chinese Academy of Sciences, Beijing, 100049, China

[c]School of Systems Science, Beijing Normal University, Beijing, 100875, China

*Correspondence to: Jiang Zhang, E-mail: zhangjiang@bnu.edu.cn

Jun Cai, E-mail: juncai@big.ac.cn

Abstract

Recent advances in single cell RNA sequencing (scRNA-seq) call more computational analysis methods. As the data for non-characterized cells accumulates quickly, supervised learning model is an ideal tool to classify the non-characterized cells based on the previously well characterized cells. However, deep learning model is an appropriate tool to deal with vast and complex data such as RNA-seq data, but lacks of interpretability. Here for the first time, we present scCapsNet, a deep learning model adapted from CapsNet. The scCapsNet model retains the capsule parts of CapsNet and replaces the part of convolutional neural networks with several parallel fully connected neural networks. We apply scCapsNet to scRNA-seq data of mouse retinal bipolar cells and human peripheral blood mononuclear cells (PBMC). The results show that scCapsNet performs well as a classifier. Meanwhile, the results also demonstrate that the parallel fully connected neural networks function like feature detectors as we supposed. The scCapsNet model provides precise contribution of each extracted feature to the cell type recognition. Furthermore, we mix the RNA expression of two cells with different cell types and then use the scCapsNet model trained with non-mixed data to predict the cell types in the mixed data. Our scCapsNet model could predict cell types in a cell mixture with high accuracy.

Introduction

Single Cell RNA sequencing (scRNA-seq) could measure gene expression levels in individual cells. Using scRNA-seq data, it is possible to reveal heterogeneity in a cell population[1, 2], identify new cell types, computationally order cells along trajectories[3, 4], and infer the spatial coordinates of every individual cell in a population[5, 6].

As the scRNA-seq data accumulates quickly, it is important to retrieve similar cell types. For example, scMCA suggested a pipeline for cell type determination by comparing the input single-cell transcriptome with pre-calculated reference transcriptome to provide a match score based on gene expression correlation[7]. Since many cell types have already been well defined, supervised learning is an ideal tool to classify undefined cells. Besides the final goal of classification or similar cell type retrieval, the interpretability of the classification process is also important. By demonstrating which features are extracted for obtaining a specific decision and how these features contribute to the decision, the classifier could offer valuable information to the downstream operation such as biomarker discovery. Deep learning model is a proper tool to deal with vast and complex data such as RNA-seq data, but lacks of interpretability[8]. Therefore, there is a need for a model that utilizes the deep learning method with increased interpretability[9, 10].

Recent years, deep learning methods have been applied to process biological data[11-13]. Specifically, several deep learning models were used to analyze scRNA-seq data for various purposes. The neural networks which incorporate prior information were used to reduce the dimensions of the data[9] or discriminate tumor subtypes and their prognostic capability[10]. Variational inference (VI) method could achieve interpretable dimensionality reduction[14], single cell grouping[15], and approximating parameters which govern the distribution of expression value of each gene in each cell[16]. Generative adversarial network (GAN) was proved useful to simulate gene expression and predict perturbation in single cell[17].

Here for the first time, we applied the CapsNet to the single cell RNA-seq data. The CapsNet is a newly invented model that exhibits more interpretability property and could be used in the task of overlapping digit recognition[18].

Traditionally in object recognition, convolution neural networks (CNN) were used to gradually extract information from local to global contained in the image. Then the information would be uploaded to a logistic or softmax unit whose output indicate the probability that one entity exists

or the entity in the image belongs to one particular type[19]. The max-pooling between layers performs the function of a feature filter which selects the most active feature to feed to the next layer. Alternatively in the CapsNet model, all features detected by convolution neural networks (CNN) would be uploaded to the next layer through “routing-by-agreement” instead of max-pooling. To achieve this, a small group of neurons called “capsule” function as neurons in the traditional neuron networks. Between two layers of capsules, the information passes through “dynamic routing”, which makes the capsules in the next layer represent the information derived from a weighted sum of capsules in the previous layer. The output of the capsule is a vector and the length of the vector could represent the probability of the existence of the entity. The architecture of the CapsNet mainly contains two parts, a feature extraction part and a capsule network part. First, the convolution neural networks respond to extract local information which resides in several capsules called “Primary Capsule”. Then, the “dynamic routing” feeds the information stored in the Primary Capsule to the next layer of capsules which are called DigitCaps. Each DigitCap represents one digit type. After “dynamic routing”, the length of the DigitCaps indicates the probability of the object in the image belonging to that digit. The CapsNet performed well at recognizing overlapping digits[18]. The CapsNet has been used in protein structure classification and prediction

[20, 21] and could be potentially used in network biology and disease biology [11].

In our experiment, we substitute the feature extraction part which uses convolutional neural networks in CapsNet with several parallel fully connected neural networks that use Rectified Linear Unit (ReLU) as the activation function. We reckon that, the parallel neural networks would function as a feature extractor. We call this modified model as "scCapsNet" and apply it to scRNA-seq data of mouse retinal bipolar cells [22] and human peripheral blood mononuclear cells (PBMC)[23]. We show that after training, the feature extraction part of scCapsNet could capture features from scRNA-seq data. And the capsule network part would compute the precise contribution of those captured features for cell type recognition. We also mix the expression data of two cells with different cell types together and use scCapsNet trained with non-mixed data to predict the types of the mixed data. The scCapsNet performs well to correctly predict the two cell types.

Methods

Datasets and data preprocessing

The two scRNA-seq datasets used in this work are mouse retinal bipolar neurons profiled by the Drop-Seq technology [22] and human peripheral

blood mononuclear cells (PBMC) sequenced by 10X [23]. For data consistence, we directly adopt the data processing module from previous work[16]. After preprocessing, the mouse retinal dataset consists of 19829 cells with 13166 genes in each cell and the PBMC dataset consists of 11990 cells with 3346 genes in each cell. All expression data is converted into log-scale through log(x+1). We randomly divide the whole dataset into training set and validation set with ratio 9 to 1 by the method of train_test_split in the python package sklearn.model_selection.

scCapsNet model

We adapt our scCapsNet model from previous CapsNet model. The architecture of our model is shown in **Figure 1**. The model contains two parts, one of which is feature detection part and the other is capsule network part. In the original CapsNet, the convolutional kernels were used as a local feature detector for image input. Instead, we choose several parallel fully connected neural networks using Rectified Linear Unit (ReLU) activation function as the feature detector. The output of each fully connected neural network is a vector with equal length, and could be viewed as “primary capsule” in the original CapsNet. This part of our model converts the feature in RNA expression to the activities of local feature detectors. Next, the activity information would be delivered

through primary capsule to the capsule in the final layer by “dynamic routing”. The capsule in the final layer， which corresponds to cell type and called “type capsule”, is a sixteen-dimension vector. The dynamic routing operation with three iterations occurs between the primary capsule layer and the type capsule layer. In capsule model, the length of the final layer capsule represents the probability that an entity represented by the capsule exists. In our cell-type classification task, the length of the final layer type capsule represents the probability that one cell belongs to that cell type.

Feature extraction and its contribution to type recognition

The detailed description for dynamic routing could be found in the original CapsNet paper[18]. In our scCapsNet model, the primary capsules that store the extracted features are first multiplied by a weight matrix to produce “prediction vector”. The dynamic routing process would calculate the “coupling coefficients” between each primary capsule’s prediction vector and all type capsules. So the total number of coupling coefficients is the product of the number of primary capsules multiplied by the number of type capsules. The coupling coefficients of one primary capsule would sum up to one. After dynamic routing, the type capsule is actually a weighted sum of prediction vectors of all primary capsules. The weights are the coupling coefficients and the

magnitude of those coefficients which indicates the likelihood of the cell type containing the extracted feature stored in the primary capsules.

Neural networks for model comparison

In order to model comparison, we replace the capsule part in scCapsNet with fully connected neural networks. The concatenated feature extraction layer is directly connected to a fully connected neural network with final classification layer using sigmoid as activation function (**Figure S1**). The loss function is as the same as that of scCapsNet model.

Mixed dataset and two- type detection

We mix the RNA expression of two different type cells with ratios of 1:1, 3:2, 2:1, 5:2 and 3:1. For type prediction of mixed data, all models are trained with non-mixed data. In order to make the models output two cell types, the top two largest scores in the output of both scCapNet and comparison model are selected, and only the correct prediction of both types is viewed as positive result. Each scalar between 0 and 1 in the output of both scCapNet and comparison model represents the probability that one cell belongs to one particular cell type. A threshold could be set for the second largest score, the cell which outputs a score below this threshold would be discarded. The prediction accuracy among the non-discarded cells would be calculated. Then the percentage that

how many cells are not discarded is computed.

Results

To test our model, two single cell RNA-seq datasets are used. One consists of mouse retinal bipolar neurons profiled by the Drop-Seq technology and the other consists of human peripheral blood mononuclear cells (PBMC) generated by 10X. Both datasets are randomly divided into training set and validation set with a ratio of 9:1. The model is run several times to access the prediction accuracy. For mouse retinal bipolar neurons dataset consisting of 19829 cells of 15 types and 13166 genes, the training accuracy reaches nearly 100% and the average validation accuracy is around 98%. For human PBMC dataset consisting of 11990 cells of 9 types (we exclude cells with type “other”, so there were actually 8 types) and 3346 genes, the training accuracy reaches nearly 100% and the average validation accuracy is around 96%.

We want to find whether the primary capsule that output by feature extraction part could capture the properties in the single cell RNA-seq dataset. Furthermore, we want to explore how the extracted features contribute to the cell type recognition. The coupling coefficient could describe the relationship between the type capsule and the primary capsule (**method**). For our supervising learning task, we could group the

coupling coefficients of the cells with prior known same cell type and calculate the average coupling coefficients for cells with one specific type. So we first compute the average coupling coefficients of each cell type in the validation set of the PBMC dataset and draw the heatmaps (**Figure 2A**). From the heatmaps, we find that for input cells with one specific cell type, one or several primary capsules contribute to the type recognition for this specific cell type but not obvious for other type. For example, among all the averaged coupling coefficients, the coupling coefficients of the primary capsule nine-B cells type capsule and the primary capsule fifteen-B cells type capsule are relatively higher than rest coupling coefficients for B cells input (**Figure 2A, 0**). The coupling coefficient of the primary capsule zero-CD14+ monocyte type capsule is obviously higher than rest coupling coefficients (**Figure 2A, 1**).

Next, in order to explore how one primary capsule affects the cell type recognition, we combine the average coupling coefficients with type capsules corresponding to specific cell types together into an overall heatmap (**Figure 2B**). We also use TSNE and PCA to reduce the dimension of each primary capsule, and plot the 2D-TSNE results (**Figure 2C**) and 2D-PCA results (**Figure S2**). We find some relationship between the overall heatmap of average coupling coefficients and the 2D-TSNE plot of each primary capsule. The details would be described hereafter.

For example, the overall heatmap indicates the primary capsule zero contributes to the CD14+monocytes recognition, and the 2D-TSNE plot of primary capsule zero shows that the CD14+ monocytes are clustered together and locate on the top and cells of rest cell types are mixed together on the bottom (**Figure 2C, 0**). The overall heatmap indicates the primary capsule one and two contribute to the CD4 T cell recognition, and the 2D-TSNE plot of primary capsule one and two show that the CD4 T cells locate on the left bottom and are almost separate from the cells of other cell types (**Figure 2C, 1 and 2**).The overall heatmap indicates that the primary capsule three and four mainly contribute to the CD8 T cell recognition, and the 2D-TSNE plot of primary capsule three and four show that the CD8 T cells locate on the left and are separate from the cells of other cell types especially NK cells which normally could not be distinguished from CD8 T cells in most 2D-TSNE plot (**Figure 2C, 3 and 4**). The overall heatmap indicates that the primary capsule seven contributes to the dendritic cells recognition, and the 2D-TSNE plot of primary capsule seven shows that the Dendritic cells locate on the left and are almost separate from the cells of other cell types (**Figure 2C, 7**). The overall heatmap indicates that the primary capsule eight and nine contribute to the B cells recognition, and the 2D-TSNE plot of primary capsule eight and nine show that the B cells locate on the bottom and

are separate from the cells of rest cell types (**Figure 2C, 8 and 9**).The overall heatmap indicates that the primary capsule thirteen contributes to the FCGR3A+ monocytes recognition, and the 2D-TSNE plot of primary capsule thirteen shows that the FCGR3A+ monocytes locate on the bottom (**Figure 2C, 13**).The overall heatmap indicates that the primary capsule fifteen contributes to the NK cells recognition, and the 2D-TSNE plot of primary capsule fifteen shows that the NK cells locate on the bottom and are separate from cells of other cell types especially CD8+ T cells (**Figure 2C, 15**).

We also perform the same procedure on the mouse retina datasets. The heatmap of average coupling coefficients for input cells of prior known one specific cell type shows that, one or several primary capsules contribute to the type recognition for this specific cell type but not obvious for other type (**Figure 3A**). For example, among all the average coupling coefficients, the coupling coefficients of primary capsule fourteen-BC5A type capsule and primary capsule eleven-BC5A type capsule are obviously higher than other coupling coefficients for BC5A cells input (**Figure 3A, 2**). The coupling coefficients of primary capsule five-BC7 type capsule and primary capsule eleven-BC7 type capsule are obviously higher than other coupling coefficients for BC7 cells input (**Figure 3A, 3**). The coupling coefficient of primary capsule ten-BC3A type

capsule is obviously higher than other coupling coefficients for BC3A cells input (**Figure 3A, 11**). The coupling coefficient of primary capsule fourteen-BC5D type capsule is obviously higher than other coupling coefficients for BC5D cells input (**Figure 3A, 10**).

We also combine the average coupling coefficient with type capsules corresponding to specific cell type inputs together into an overall heatmap (**Figure 3B**) and draw the 2D-TSNE and 2D-PCA plot of outputs of every primary capsule (**Figure 3C and Figure S3**). The overall heatmap indicates that the primary capsule two contributes to the BC6 cell recognition, and the 2D-TSNE plot of primary capsule two shows that most BC6 cells are clustered together on the top and cells of other cell types especially cone cells are mixed together on the bottom (**Figure 3C, 2**). The overall heatmap indicates that the primary nine contributes to the BC5C cell recognition, and the 2D-TSNE plot of primary capsule nine shows that the BC5C cells are clustered on the top and cells of other cell types especially cone cells are mixed together on the bottom (**Figure 3C, 9**). The overall heatmap indicates that the primary four contributes to the BC3B and RBC cell recognition, and the 2D-TSNE plot of primary capsule four shows that the BC3B and RBC cells locate on the left and are almost separate from the rest cells (**Figure 3C, 4**). The overall heatmap indicates that the primary five contributes to the BC1A, BC1B and BC7

cell recognition, and the 2D-TSNE plot of primary capsule five shows that the BC1A, BC1B and BC7 cells located on the right and are almost separate from the rest cells (**Figure 3C, 5**). The overall heatmap indicates that the primary capsule ten contributes to the BC1B and BC3A cell recognition, and the 2D-TSNE plot of primary capsule ten shows that the BC1B and BC3A cell locate on the bottom and are almost separate from the rest cells (**Figure 3C, 10**). The overall heatmap indicates that the primary fourteen contributes to the BC5D, BC5B and BC5A cell recognition, and the 2D-TSNE plot of primary capsule fourteen shows that the BC5D, BC5B and BC5A cells locate on the left and are almost separate from the rest cells (**Figure 3C, 14**). The heatmap indicates that the primary capsule eleven contributes to the BC2, BC6, BC7, BC5A and RBC cell recognition, and the 2D-TSNE plot of primary capsule eleven shows that the BC2, BC6, BC7, BC5A and RBC cells are independently clustered for their own cell type and cells of other cell types are mixed together (**Figure 3C, 11**).

From the above observation, we find the association between heatmap of average coupling coefficients and 2D-TSNE plot of primary capsules. If the heatmap indicates one primary capsule contributes to certain cell type recognition, the cells belong to those cell types would independently cluster into separate groups, while cells labeled as other

cell types would mix together. This phenomenon reveals that the primary capsule actually captures some properties of scRNA-seq datasets, and contribution of those captured properties to the cell type recognition could be measured by coupling coefficients.

In hinton's paper, the dynamic routing made the CapsNet capable of recognizing multiple objects in an image even if the objects were overlapped, although the model was trained with the single non-overlapped objects in image[18]. We mimic the overlapped objects in an image by adding the gene expression of two cells with different ratio in single cell RNA-seq data. We want to test whether our model trained with non-mixed data could decouple the data which is a mixture of two different types of cells. The results show that our model could accomplish the cell type prediction of the mixture with high accuracy. As we expect, the more unequal the cell mixture is, the lower prediction accuracy the model could make (**Table 1, 2**).

Although the overall prediction accuracies of our scCapsNets model are slightly lower than those of the comparison model, when we set a threshold to the results, we could easily find that the comparison model tends to give a lower probability as compared to scCapsNet. Under the same threshold, the percentage that the cells which output a score

above this threshold by scCapsNet are obviously larger than the corresponding percentage of the comparison model (**Table 1, 2**).

Discussion

We demonstrate that the proposed scCapsNet model performs well in the cell type prediction. The parallel fully connected neural networks could function like a feature extractor as convolutional neural networks in the original CapsNet model. Furthermore, the capsule part of our scCapsNet model forces the feature extraction part to capture local feature even specific to just one cell type. The model could provide the precise contribution of each extracted feature to the cell type recognition.

To sum up, our scCapNet model could automatically extract features from data, and then use those extracted features to accomplish the cell type classification task and compute the exact contribution of each extracted feature to the type classification. Therefore, our scCapsNet model could potentially be used in the classification scenario where multiple information sources are available such as -omic datasets with data generated across different biological layers (e.g., transcriptomics, proteomics, metabolomics) [11]. By integrating the different information sources to classification, our scCapsNet model could provide the precise

contribution of each information source. And it is seem that our scCapsNet model is more suitable for biological data than original CapsNet model.

In view of such high prediction accuracy to recognize mixed data, our scCapsNet model and the comparison model could potentially apply to wide range of research areas related to scRNA-seq. For example, cells develop or differentiate from stem cells to terminal differentiated cells through a trajectory. Along the trajectory, there are several states that have already been well characterized before. Training with those well defined states which could be viewed as landmarks, these two models could infer the locations of the testing cells on the trajectory by reporting two landmark states. Then, the testing cells could be viewed as at an intermediate state between the two reported landmarks. Like reporting the intermediate state of a cell, these two models could also be used to infer the intermediate position of a cell given several well defined landmark cells.

Competing financial interests

The authors declare no competing financial interests.

Acknowledgments

This work was supported by grants from the National Key R&D Program of China [2018YFC0910402 to C.J.]; the National Natural Science Foundation of China [31571307 to C.J. and 61673070 to J.Z.]

Figures

Figure 1: Architecture of scCapsNet with two layers. First layers consist of sixteen parallel fully connected neural networks for feature extraction. Each output of the neural network is viewed as a primary capsule. The subsequent layer is directly adopted from CapsNet for classification. Information flows from primary capsules to final layer type capsules. Then the length of each type capsule represents the probability of input data belonging to the corresponding cell type.

Figure 2A: The heatmaps of coupling coefficients for PBMC dataset. Each heatmap represents the average coupling coefficients of inputs cells that belong to one specific cell type. In the heatmap, the row represents type capsules and column represents primary capsules. For example, the row zero represents B cells type capsule. The total number of the coupling coefficients is the product of the number of the type capsules multiplied by the number of the primary capsules. The order of the subplot is from left to right and top to bottom with index from zero to seven.

Figure 2B: The overall heatmap of average coupling coefficients with type capsules corresponding to specific cell type inputs for PBMC dataset.

The row represents type capsules and column represents primary capsules.

Figure 2C: The 2D-TSNE plots of primary capsules of input cells from validation set for PBMC dataset. Each subplot represents a primary capsule. The order of the subplot is from left to right and top to bottom with index from zero to fifteen.

Figure 3A: The heatmaps of coupling coefficients for mouse retina bipolar cells dataset. Each heatmap represents average coupling coefficients of inputs cells that belong to one specific cell type. In the heatmap, the row represents type capsules and column represents primary capsules. For example, the row zero represents the RBC type capsule. The total number of the coupling coefficients is the product of the number of the type capsules multiplied by the number of the primary capsules. The order of the subplot is from left to right and top to bottom with index from zero to fourteen.

Figure 3B: The overall heatmap of average coupling coefficients with type capsules corresponding to a specific cell type input for mouse retina bipolar cells dataset. The row represents type capsules and column represents primary capsules.

Figure 3C: The 2D-TSNE plots of primary capsules of input cells from validation set for mouse retina bipolar cells dataset. Each subplot represents a primary capsule. The order of the subplot is from left to

right and top to bottom with index from zero to fifteen.

Tables

Table 1: The performance of different models on the task of cell mixture types recognition for PBMC dataset. The threshold is set for the second largest score in the output of the model. The scCapsNet is our model, and NN is the comparison model which uses the fully connected neural networks to replace the capsule networks. The percentage is calculated by the number of the cells those are not filtered out by the threshold divided by the total number of cells. The accuracy is calculated among the cells that are not filtered out by the threshold.

Table 2: The performance of different models on the task of cell mixture types recognition for mouse retina bipolar cell dataset. The threshold is set for the second largest score in the output of the model. The scCapsNet is our model, and NN is the comparison model which uses the fully connected neural networks to replace the capsule networks. The percentage is calculated by the number of the cells those are not filtered out by the threshold divided by the total number of cells. The accuracy is calculated among the cells that were not filtered out by the threshold.

Supplementary Figures

Figure S1: The architecture of the comparison neural network model.

The comparison model retains the feature extraction layer of scCapsNet, but the capsule network part is substituted by a fully connected neural network with sigmoid activation function for classification.

Figure S2: The 2D-PCA plots of primary capsules of input cells from validation set for PBMC dataset. Each subplot represents a primary capsule. The order of the subplot is from left to right and top to bottom with index from zero to fifteen.

Figure S3: The 2D-PCA plots of primary capsules of input cells from validation set for mouse retina bipolar cells dataset. Each subplot represents a primary capsule. The order of the subplot is from left to right and top to bottom with index from zero to fifteen.

Reference

1. Zheng, C., et al., *Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing.* Cell, 2017. **169**(7): p. 1342-1356 e16.
2. Guo, X., et al., *Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing.* Nat Med, 2018. **24**(7): p. 978-985.
3. Setty, M., et al., *Wishbone identifies bifurcating developmental trajectories from single-cell data.* Nat Biotechnol, 2016. **34**(6): p. 637-45.
4. Qiu, X., et al., *Reversed graph embedding resolves complex single-cell trajectories.* Nat Methods, 2017. **14**(10): p. 979-982.
5. Halpern, K.B., et al., *Single-cell spatial reconstruction reveals global division of labour in the mammalian liver.* Nature, 2017. **542**(7641): p. 352-356.
6. Halpern, K.B., et al., *Paired-cell sequencing enables spatial gene expression mapping of liver endothelial cells.* Nat Biotechnol, 2018. **36**(10): p. 962-970.
7. Han, X., et al., *Mapping the Mouse Cell Atlas by Microwell-Seq.* Cell, 2018. **173**(5): p. 1307.
8. Almas Jabeen, N.A., Khalid Raza, *Machine Learning-Based State-Of-The-Art Methods For The Classification Of RNA-Seq Data.* bioRxiv doi: 10.1101/120592, 2017.
9. Lin, C., et al., *Using neural networks for reducing the dimensions of single-cell RNA-Seq data.* Nucleic Acids Res, 2017. **45**(17): p. e156.
10. Hung-I Harry Chen, Y.-C.C., Tinghe Zhang, Songyao Zhang, Yufei Huang, Yidong Chen, *GSAE: an autoencoder with embedded gene-set nodes for genomics functional characterization.*

arXiv:1805.07874, 2018.

11. Camacho, D.M., et al., *Next-Generation Machine Learning for Biological Networks.* Cell, 2018. **173**(7): p. 1581-1592.
12. Zhou, J., et al., *Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk.* Nat Genet, 2018. **50**(8): p. 1171-1179.
13. Alipanahi, B., et al., *Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning.* Nat Biotechnol, 2015. **33**(8): p. 831-8.
14. Ding, J., A. Condon, and S.P. Shah, *Interpretable dimensionality reduction of single cell transcriptome data with deep generative models.* Nat Commun, 2018. **9**(1): p. 2002.
15. Christopher Heje Grønbech, M.F.V., Pascal N Timshel, Casper Kaae Sønderby, Tune Hannes Pers, Ole Winther, *scVAE: Variational auto-encoders for single-cell gene expression data.* bioRxiv doi: 10.1101/318295, 2018.
16. Lopez, R., et al., *Deep generative modeling for single-cell transcriptomics.* Nature Methods, 2018. **15**(12): p. 1053-+.
17. Arsham Ghahramani, F.M.W., Nicholas M Luscombe, ed. *Generative adversarial networks simulate gene expression and predict perturbations in single cells*. bioRxiv doi: 10.1101/262501. 2018.
18. Sabour, S., N. Frosst, and G.E. Hinton, *Dynamic Routing Between Capsules.* Advances in Neural Information Processing Systems 30 (Nips 2017), 2017. **30**.
19. Lecun, Y., et al., *Gradient-based learning applied to document recognition.* Proceedings of the Ieee, 1998. **86**(11): p. 2278-2324.
20. Dan Rosa de Jesus, J.C., Wilson Rivera, Silvia Crivelli, *Capsule Networks for Protein Structure Classification and Prediction.* arXiv:1808.07475, 2018.
21. Fang, C., Y. Shang, and D. Xu, *Improving Protein Gamma-Turn Prediction Using Inception Capsule Networks.* Scientific Reports, 2018. **8**.
22. Shekhar, K., et al., *Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics.* Cell, 2016. **166**(5): p. 1308-1323 e30.
23. Zheng, G.X., et al., *Massively parallel digital transcriptional profiling of single cells.* Nat Commun, 2017. **8**: p. 14049.

Figure 1

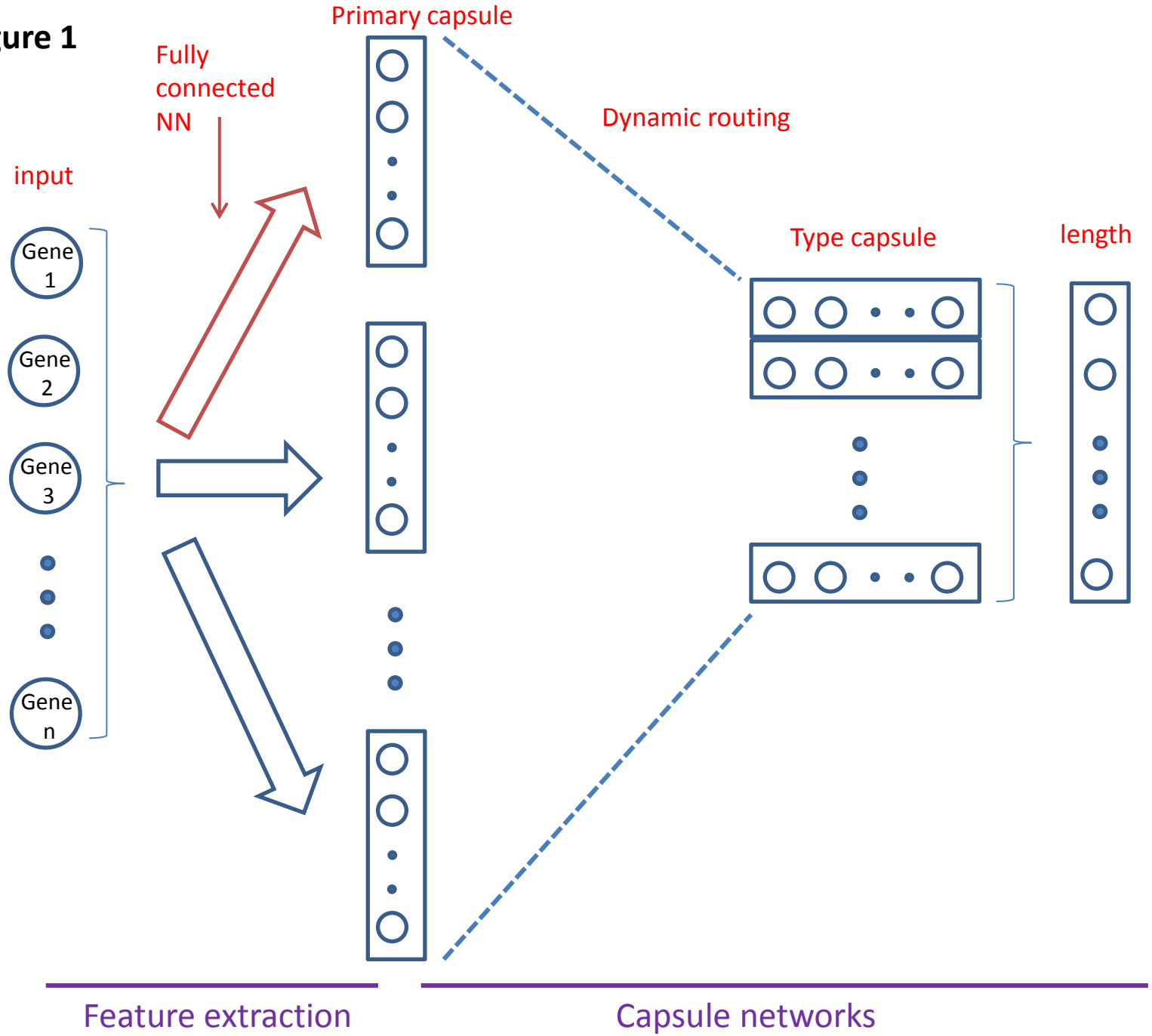

Primary capsule
Fully connected NN
Dynamic routing
input
Type capsule
length
Gene 1
Gene 2
Gene 3
Gene n
Feature extraction
Capsule networks

Figure 2A

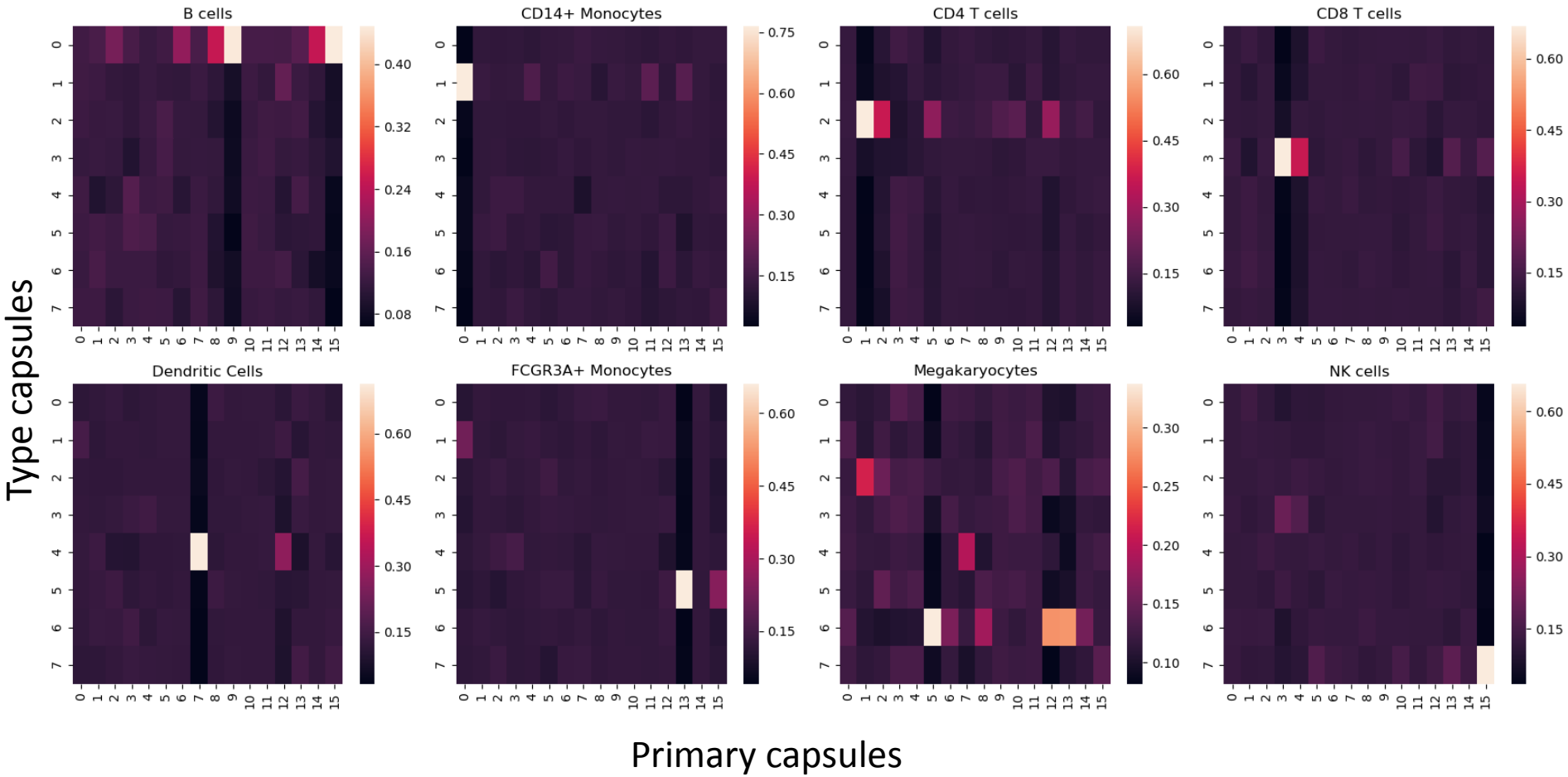

B cells
CD14+ Monocytes
CD4 T cells
CD8 T cells
Dendritic Cells
FCGR3A+ Monocytes
Megakaryocytes
NK cells
Type capsules
Primary capsules

Figure 2B

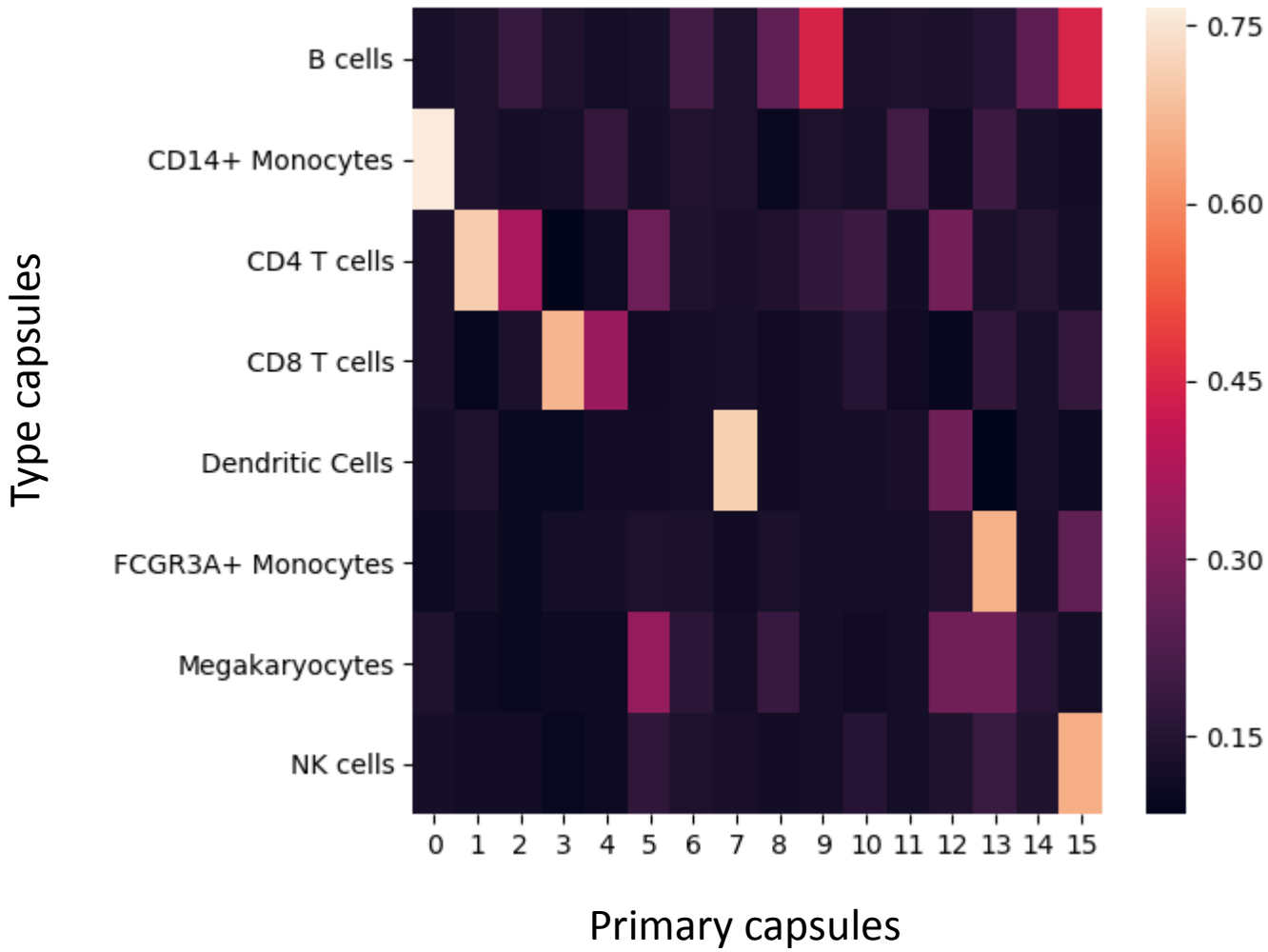

B cells
CD14+ Monocytes
CD4 T cells
CD8 T cells
Dendritic Cells
FCGR3A+ Monocytes
Megakaryocytes
NK cells
Type capsules
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Primary capsules
0.75
0.60
0.45
0.30
0.15

Figure 2C

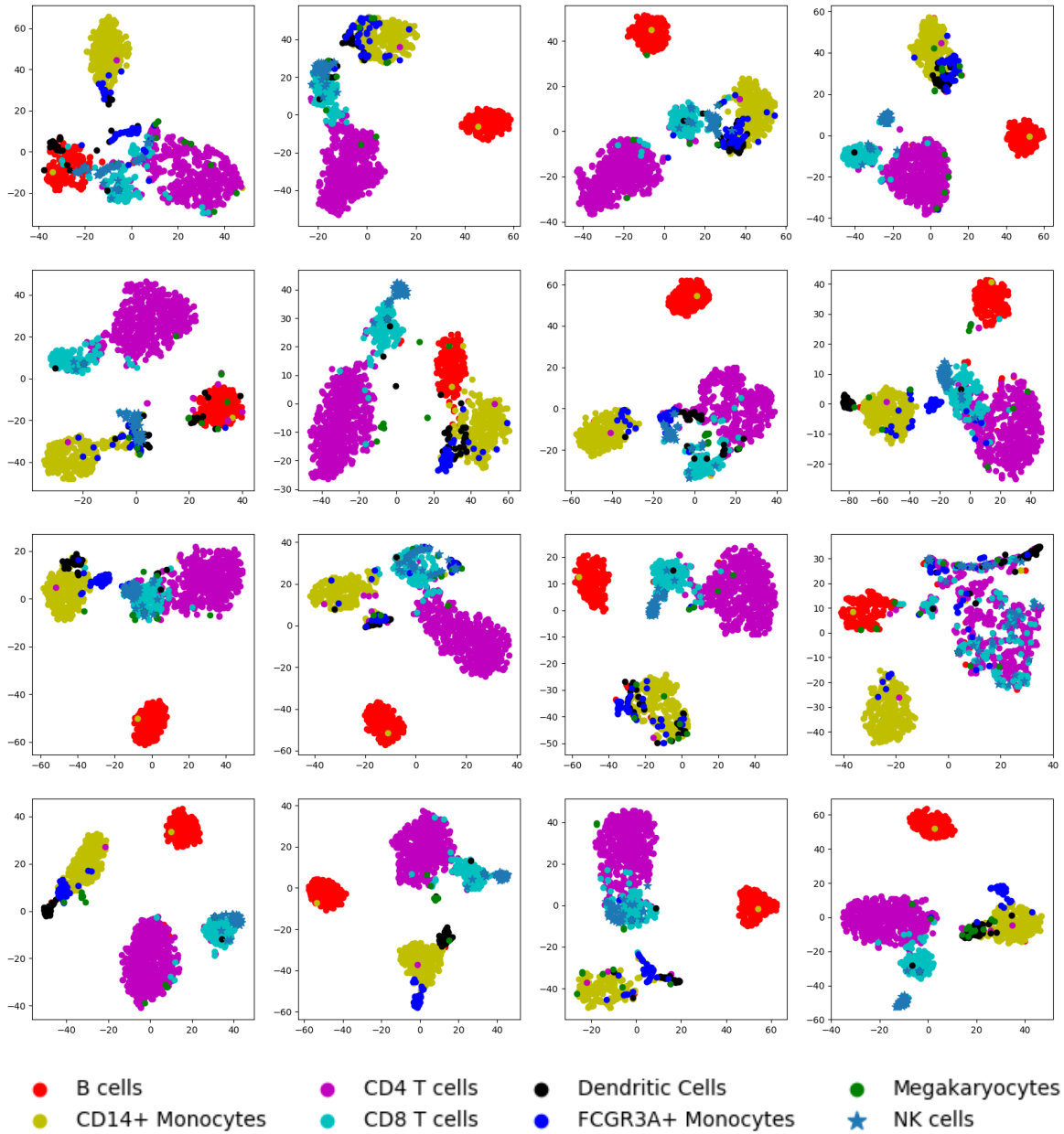

B cells
CD14+ Monocytes
CD4 T cells
CD8 T cells
Dendritic Cells
FCGR3A+ Monocytes
Megakaryocytes
NK cells

Figure 3A

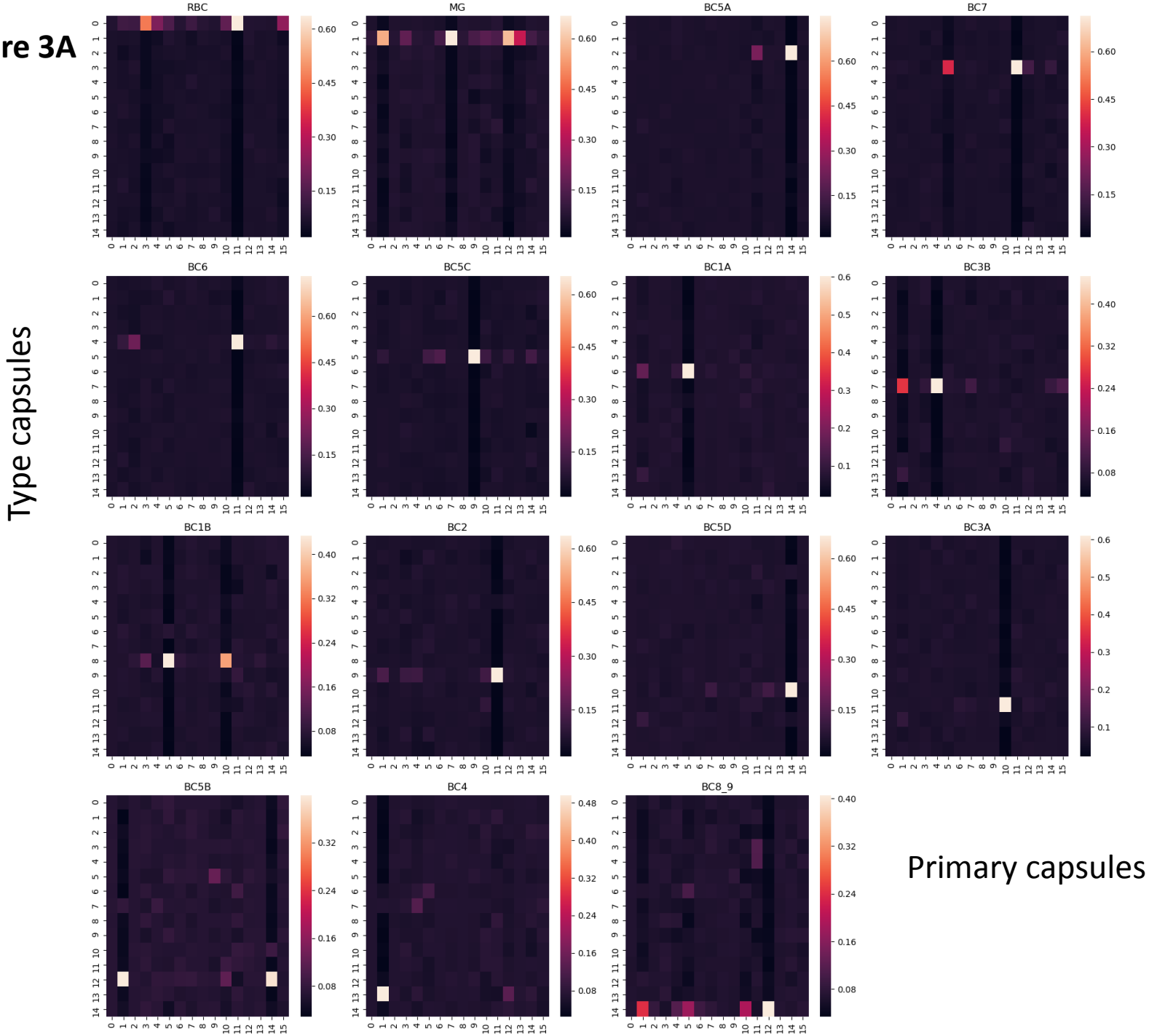

RBC
MG
BC5A
BC7
BC6
BC5C
BC1A
BC3B
BC1B
BC2
BC5D
BC3A
BC5B
BC4
BC8_9
Type capsules
Primary capsules

Figure 3B

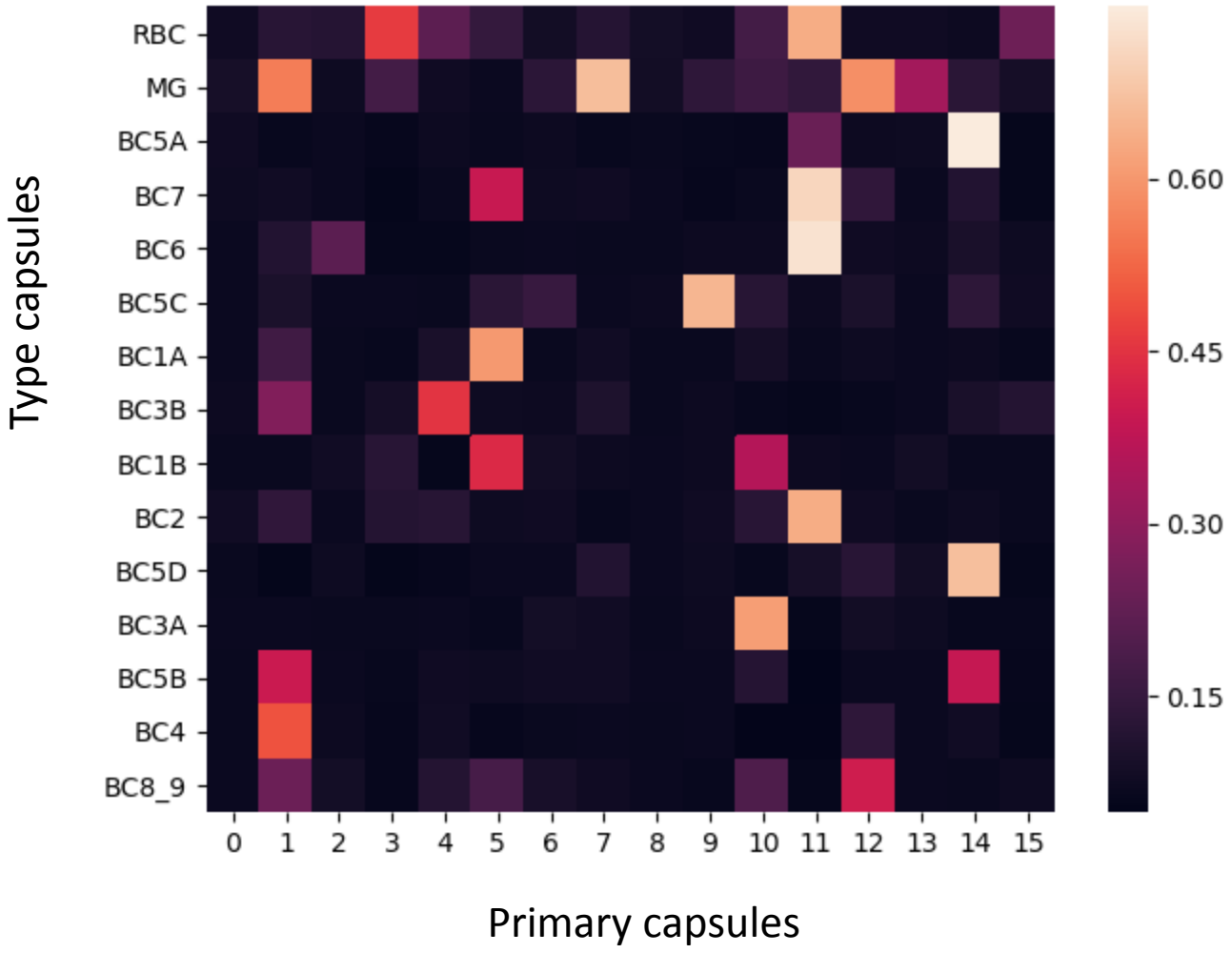

Type capsules
RBC
MG
BC5A
BC7
BC6
BC5C
BC1A
BC3B
BC1B
BC2
BC5D
BC3A
BC5B
BC4
BC8_9
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Primary capsules
0.60
0.45
0.30
0.15

Figure 3C

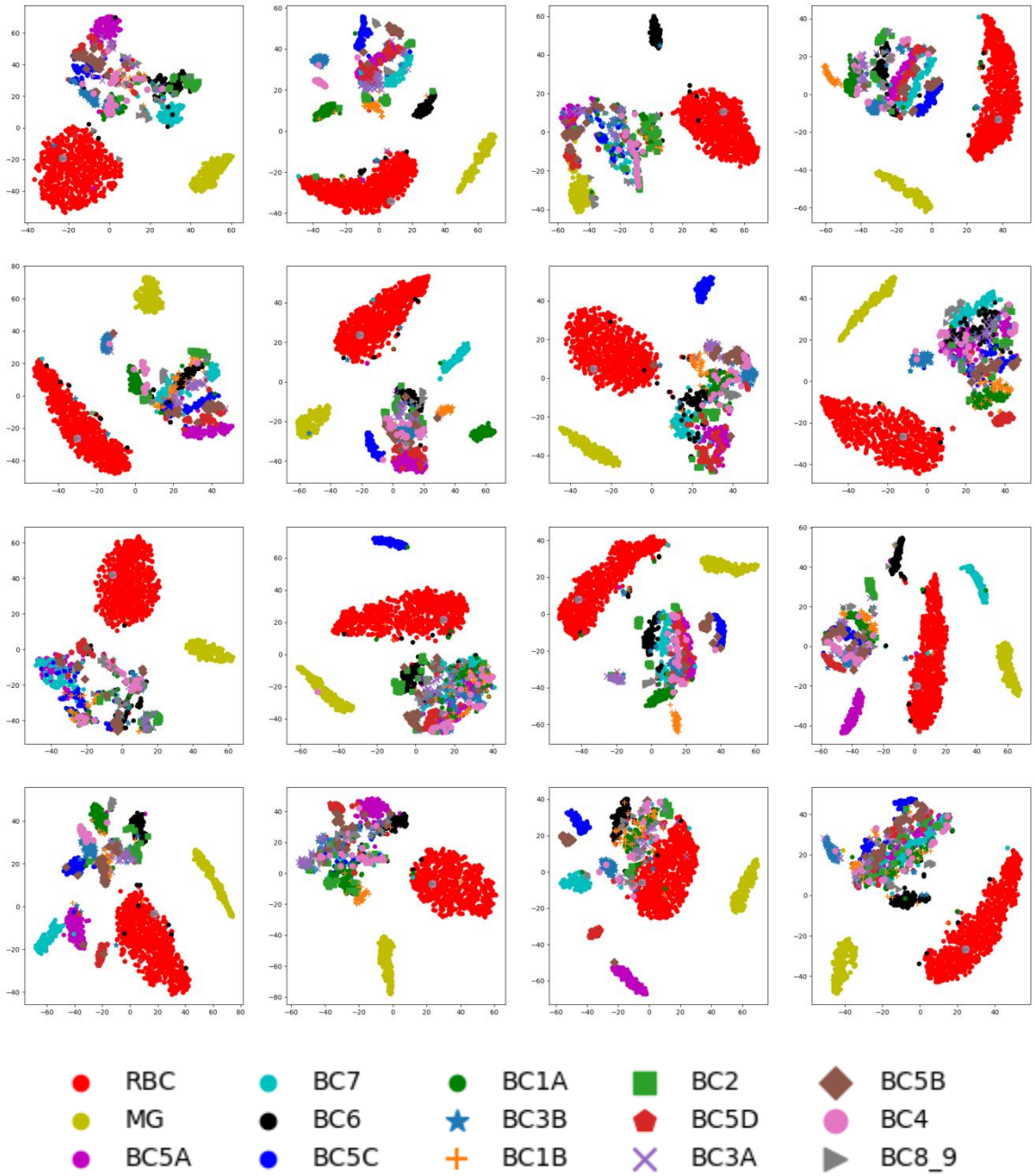

RBC
MG
BC5A
BC7
BC6
BC5C
BC1A
BC3B
BC1B
BC2
BC5D
BC3A
BC5B
BC4
BC8_9

Table 1

	threshold	0.5		0.4		0.3		0.2		0.1	
mixture	%	scCaps Net	NN	scCaps Net	NN	scCaps Net	NN	scCapsNet	NN	scCapsNet	NN
1:1	Accuracy	0.992	0.993	0.989	0.986	0.978	0.98	0.955	0.959	0.927	0.925
	percentage	0.157	0.073	0.3	0.195	0.469	0.364	0.658	0.593	0.852	0.829
3:2	Accuracy	0.983	0.974	0.975	0.982	0.962	0.968	0.942	0.949	0.905	0.908
	percentage	0.113	0.056	0.215	0.145	0.377	0.285	0.565	0.492	0.795	0.774
2:1	Accuracy	0.959	0.961	0.954	0.959	0.93	0.953	0.912	0.919	0.874	0.878
	percentage	0.058	0.028	0.139	0.091	0.254	0.196	0.434	0.371	0.687	0.675
5:2	Accuracy	0.907	0.912	0.907	0.932	0.888	0.919	0.87	0.877	0.835	0.842
	percentage	0.038	0.016	0.085	0.056	0.178	0.139	0.328	0.286	0.592	0.577
3:1	Accuracy	0.841	0.892	0.838	0.879	0.819	0.867	0.831	0.831	0.788	0.799
	percentage	0.027	0.013	0.06	0.039	0.127	0.1	0.256	0.233	0.524	0.509

Table 2

	threshold	0.5		0.4		0.3		0.2		0.1	
mixture	%	scCaps Net	NN	scCaps Net	NN	scCaps Net	NN	scCapsNet	NN	scCapsNet	NN
1:1	Accuracy	0.991	1.0	0.989	1.0	0.984	0.999	0.977	0.994	0.957	0.975
	percentage	0.153	0.009	0.256	0.047	0.403	0.149	0.584	0.344	0.809	0.638
3:2	Accuracy	0.988	1.0	0.981	1.0	0.976	0.996	0.966	0.99	0.943	0.964
	percentage	0.123	0.007	0.21	0.039	0.339	0.122	0.497	0.29	0.73	0.569
2:1	Accuracy	0.961	1.0	0.961	0.991	0.96	0.99	0.943	0.979	0.917	0.938
	percentage	0.081	0.005	0.146	0.027	0.242	0.086	0.388	0.221	0.618	0.475
5:2	Accuracy	0.919	1.0	0.93	0.987	0.924	0.981	0.911	0.952	0.876	0.894
	percentage	0.055	0.004	0.099	0.017	0.17	0.059	0.289	0.16	0.516	0.397
3:1	Accuracy	0.861	0.833	0.876	0.949	0.867	0.959	0.87	0.921	0.824	0.84
	percentage	0.037	0.001	0.068	0.01	0.121	0.039	0.219	0.12	0.433	0.342

Figure S1

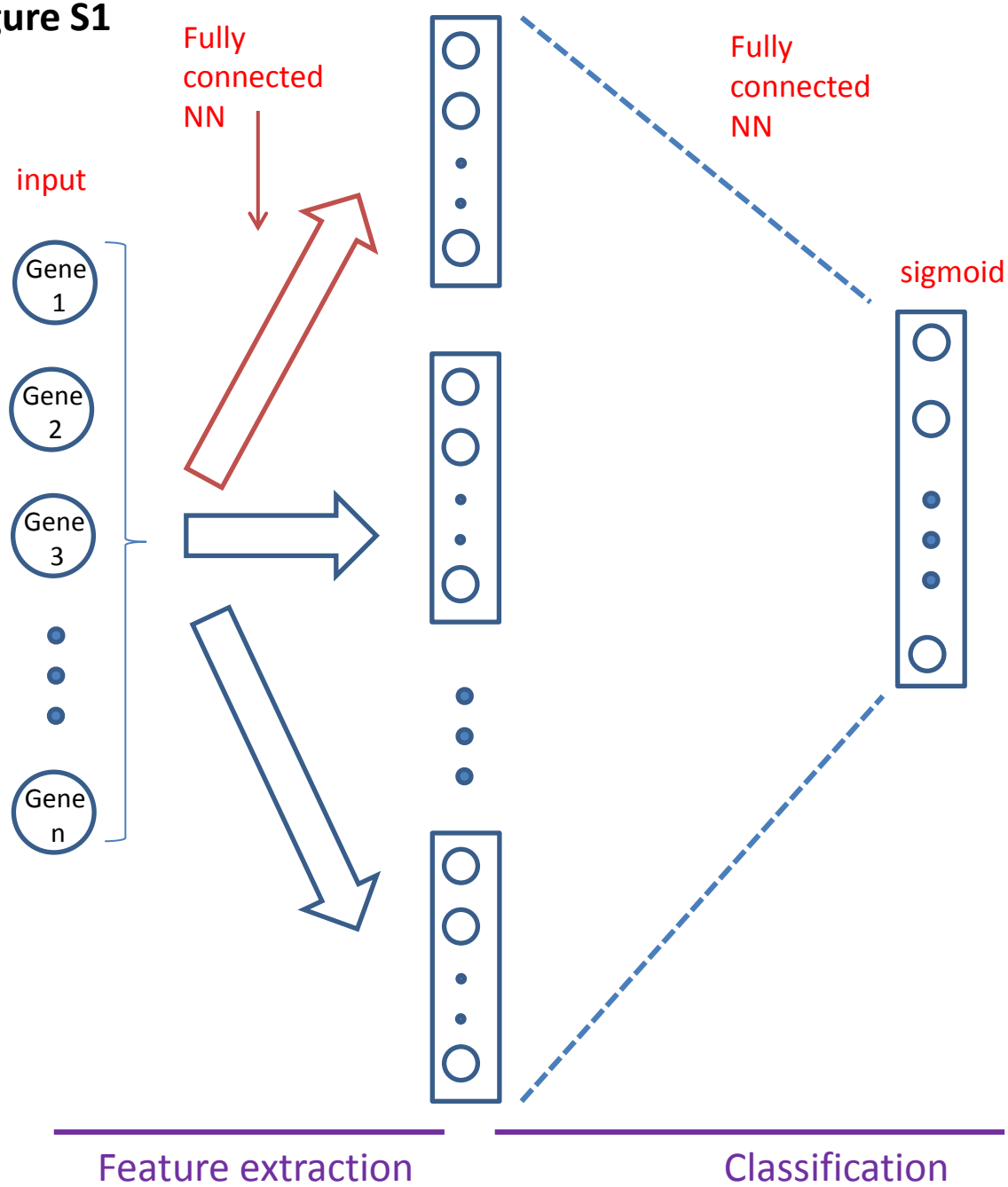

Fully connected NN
Fully connected NN
input
sigmoid
Gene 1
Gene 2
Gene 3
Gene n
Feature extraction
Classification

Figure S2

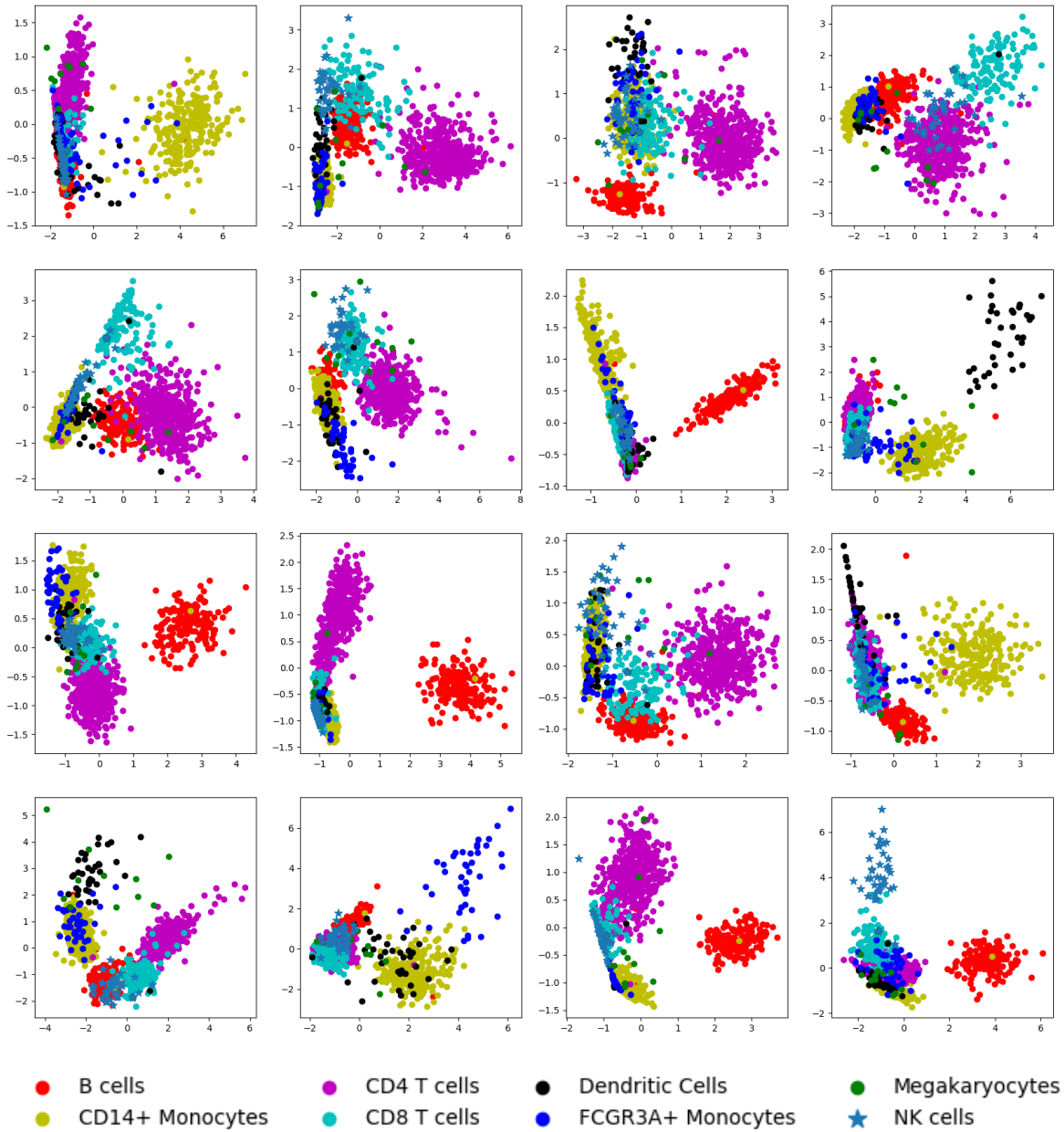
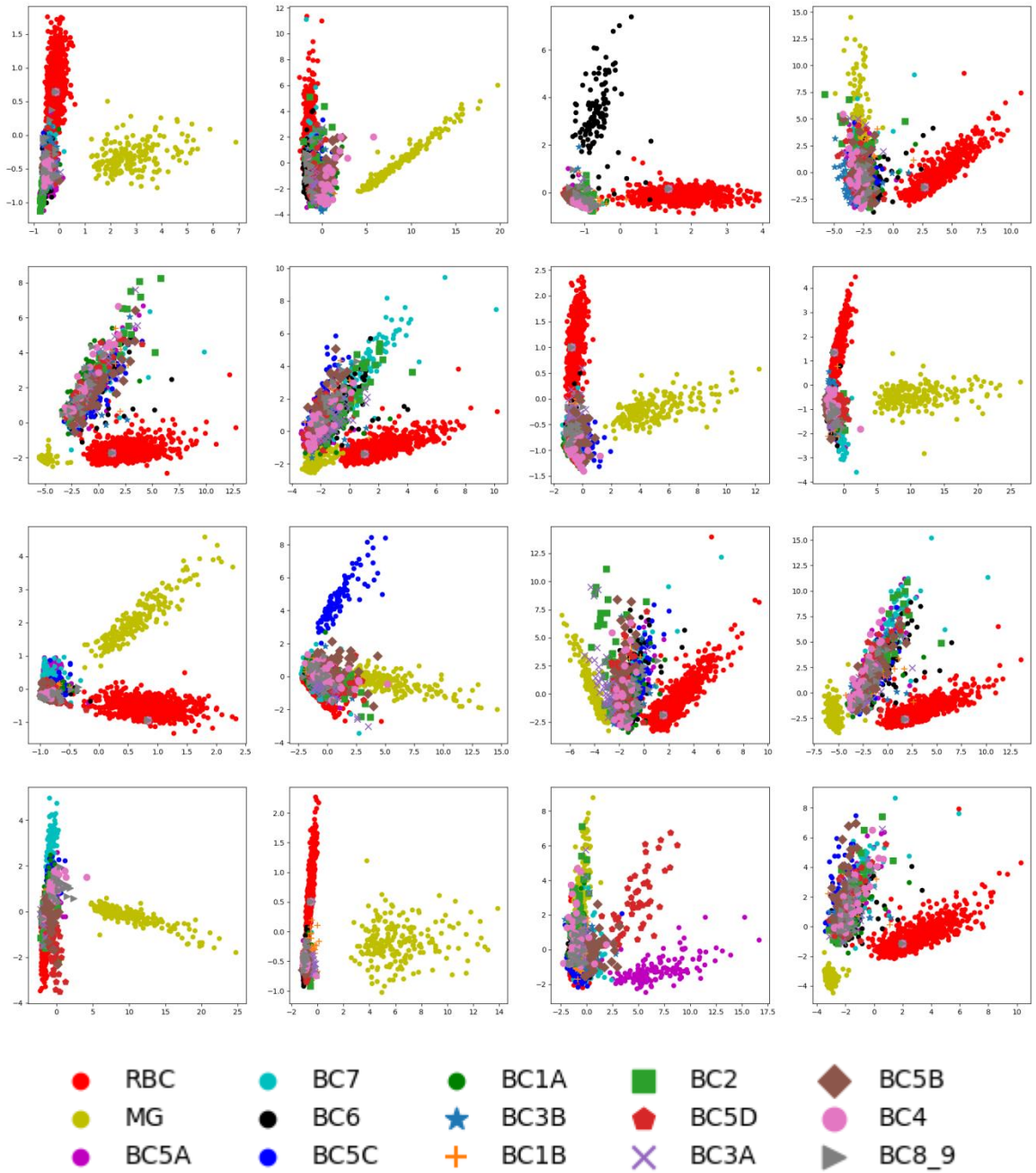

B cells
CD14+ Monocytes
CD4 T cells
CD8 T cells
Dendritic Cells
FCGR3A+ Monocytes
Megakaryocytes
NK cells

Figure S3


RBC
MG
BC5A
BC7
BC6
BC5C
BC1A
BC3B
BC1B
BC2
BC5D
BC3A
BC5B
BC4
BC8_9